



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

FACULTAD DE CIENCIAS



**“CLASIFICACIÓN DE TERMOGRAMAS DE  
CÁNCER DE MAMA UTILIZANDO  
APRENDIZAJE POR TRANSFERENCIA”**

TESIS

PARA OBTENER EL GRADO DE

**DOCTOR EN CIENCIAS APLICADAS**

PRESENTA:

**M.C. JUAN CARLOS TORRES GALVÁN**

DIRECTOR DE TESIS:

**DR. FRANCISCO JAVIER GONZÁLEZ  
CONTRERAS**

SAN LUIS POTOSÍ, S.L.P.

18 DE JUNIO DEL 2021

# TESIS DE DOCTORADO

## CLASIFICACIÓN DE TERMOGRAMAS DE CÁNCER DE MAMA UTILIZANDO APRENDIZAJE POR TRANSFERENCIA

M.C. JUAN CARLOS TORRES GALVÁN

### COMITÉ QUE ACEPTA LA TESIS

---

DR. FRANCISCO JAVIER  
GONZÁLEZ CONTRERAS

DIRECTOR DE TESIS

---

DR. JORGE LUIS FLORES  
NÚÑEZ

SINODAL EXTERNO

---

DR. ELEAZAR SAMUEL  
KOLOSOVAS MACHUCA

SINODAL

---

DR. MIGUEL GHEBRE  
RAMÍREZ ELÍAS

SINODAL

JUNIO 2021

# Declaratoria de Autoría

Yo, **Juan Carlos Torres Galván**, estudiante del Posgrado en Ciencias Aplicadas de la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí, como autor de la tesis *“Clasificación de Termogramas De Cáncer de Seno Utilizando Aprendizaje por Transferencia”*, declaro que la tesis es una obra original, inédita, auténtica, personal, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales vigentes que protegen los derechos de autor y de propiedad intelectual e industrial. Las ideas, doctrinas, resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.

# Resumen

El cáncer de mama representa el 24.2% de los cánceres detectados en mujeres en el mundo con una mortalidad de 13%, detrás del cáncer de pulmón, mientras que en México tiene una incidencia del 39.5% en las mujeres y una mortalidad de 9.9%.

La mamografía es la técnica que se utiliza como estándar de oro en la cual su sensibilidad varía del 36% al 98% en función del tipo de tejido en la mama presentando una elevada tasa de falsos positivos, además de utilizar radiación, disminuir su efectividad cuando se aplica a menores de 50 años y ser un método incómodo en su adquisición.

La termografía es un método no invasivo que permite medir la radiación emitida y determinar la temperatura de un cuerpo apreciándose así anomalías térmicas en los senos que pueden traducirse en un signo de cáncer de mama. La técnica tiene una sensibilidad que oscila entre el 71% y el 94%, dependiendo en mucho de la experiencia y habilidad de la persona al interpretar los termogramas.

Una de las tareas para las que es empleado el aprendizaje profundo es la clasificación de imágenes mediante técnicas de transferencia de aprendizaje en una arquitectura previamente entrenada a la cual se le modifican las últimas capas de este para entrenarlo en el problema a estudiar.

Se utilizó la arquitectura *ResNet 101* que fue entrenada con un total de 311 termogramas provenientes de tres bases de datos adquiridos en diferentes centros, 226 de los cuales fueron utilizados para el entrenamiento y 85 para la validación. Dos enfoques fueron utilizados para la realización del mismo, una con clases balanceadas teniendo la misma probabilidad de ser clasificados como pacientes del grupo control o con anomalía térmica, y con clases desbalanceadas en los cuales se tiene un número mucho mayor de población sana.

Debido a que estos 311 datos suministrados al algoritmo es muy limitado tanto para su entrenamiento como su validación ya que usualmente se le alimenta con miles de imágenes para que pueda realizar una mejor generalización de las mismas, se aplica la técnica de aumento de datos en ambos enfoques. Se realizó un ajuste fino a la arquitectura al aplicarse en sendos estudios.

Los mejores resultados obtenidos en el enfoque balanceado una sensibilidad de 92.3% y una especificidad de 53.8% con un área bajo la curva de 0.722 y una exactitud de 73.1%, mientras que para el desbalanceado fue una sensibilidad de 84.6%, especificidad de 65.3%, área bajo la curva de 0.814 y exactitud de 74.9%.

Los resultados obtenidos son similares a los que obtendría un experto en el análisis de termogramas de mama al evaluarlos de forma manual utilizando las diferencias de temperatura en la escala de Ville Marie. Fue ejecutada la activación de mapeo de clases, la cual permite conocer las regiones que son más importantes para que el algoritmo clasifique a la imagen dentro de un grupo.

Una prueba para determinar la afectación de la sensibilidad térmica de las cámaras infrarrojas con las cuales fueron adquiridos los datos fue realizada. Los resultados arrojaron una variabilidad de

hasta el 9.6% en la clasificación de los termogramas al variar la sensibilidad térmica, muchos de estos cambios fueron cuando se tenía una posibilidad muy cercana al 50% de pertenecer a una u otra clase.

De los termogramas que tenían la información del BI-RADS, así como aquellos que fueron clasificados como controles, se realizó un análisis intentando catalogarlos como pacientes controles, con lesiones benignas o malignas según sea el caso manteniendo los enfoques previamente mencionados.

La información arrojada por esta prueba para clases balanceadas una sensibilidad de 77.7%, especificidad de 88.9% y una exactitud de 77.7%, mientras que para las clases desbalanceadas la sensibilidad es de 71.7%, especificidad de 90.4% y exactitud de 83.3%. Debido a que solamente se contaron con 33 termogramas para la clase balanceada y 84 para la desbalanceada, se sugiere incrementar el tamaño de la base de datos para contar con un resultado más robusto de generalización al emplear este enfoque.

***Palabras Clave:*** *Transferencia de aprendizaje automático, termografía, cáncer de mama*

# Abstract

Breast cancer represents the 24.2% of all the women's cancers detected in the world with mortality of 13%, just behind of the lung cancer, while in Mexico it has an incidence of 39.5% in women with a mortality of 9.9%.

Mammography is the gold standard with a sensitivity that range between 36% to 98%, depending on the tissue of the breast presenting a high rate of false positives, in addition to use radiation, the decrease of accuracy when it's applied in persons of less to 50 years and the patient's discomfort in his adquisition.

Thermography is a non-invasive technique that can measure the emitted radiation and determine the temperature of the body showing the thermal abnormalities in the breast that could be a sign of breast cancer. This technique has a sensitivity between 71% and 94%, depending on the experience and ability of the person to interpret the thermograms.

One of the tasks of the application of deep learning is image classification using transfer learning techniques in a network previously trained, changing the last three layers for make a training in the prople to study.

*ResNet 101* was trained with a total of 311 thermograms acquired in three different centers, 226 of them used for training and 85 for validation of the architecture. Two approaches were employed, one with balanced class, having the same probability to be classified as controls or abnormal, and unbalanced classes with a higher number of healthy people, expected scenario in a clinic probe.

The 311 images fed it to the network are short for evaluate their performance in comparison with the thousands of data normally used to have a better generalization of them, data augmented and fine tuning is applied in both aproaches.

The best results in balanced approach were a sensitivity of 92.3% and specificity of 53.8% with an are under ROC curve of 0.722 and a 73.1% of accuracy, while in the unbalanced probe a sensitivity of 84.6%, specificity of 65.3%, area under ROC curve of 0.814 and 74.9% of accuracy.

Results obtained for the network are similar to obtained for an expert in thermogram's analysis using the themperature differences in Ville-Marie scale. Class activation mapping was employed for know the most important regions for the algorithm to classify an image in each category.

It was possible to determine if the thermal sensitivity of the camera used for acquire the data affects the results. A variability of 9.6% in thermograms classification was observed, many of this when the image was close to the 50% to be classified in one or another category.

Thermotrams with information of the BI-RADS classification, and those that are part of the control group were employed for another test trying to classify it in benign lesions, malignant lesion and controls, using both approaches previously described.

Results for this test using balanced classes, was a sensitivity of 77.7%, specificity of 88.9% and 77.7% in accuracy, while for unbalanced classes, the sensitivity obtained was 71.7%, specificity of 90.4% and a 83.3% in accuracy. On account that we only had 33 thermograms for the balanced approach and 84 for unbalanced, it is suggested to increase the size of the database for have robust results that allow the generalization of thermograms using this approach.

**Keywords:** *Transfer learning, thermography, breast cancer.*

# Agradecimientos

Quiero agradecer a Dios por permitirme lograr un grado académico que en un principio lucía solamente un sueño obtenerlo, y la oportunidad de hacer algo que me agrada, con todas las complicaciones y dificultades, pero sobre todo, las satisfacciones, alegrías y el crecimiento que me ha dado. Quiero agradecer a mi familia que ha estado junto a mí en cada momento de mi vida tanto personal como profesional. A mis amigos que tanto en aguas tranquilas como en momentos de apremio han estado para compartir las diferentes etapas de la vida.

Agradezco al Dr. Javier González por haber sido mi asesor tanto en maestría como en doctorado y ayudarme en este camino totalmente nuevo para mí. De igual forma agradecer al proyecto 278291 (SRE-CONACYT) y al proyecto P105 de CEMIE-Solar (Secretaría de Energía-CONACYT).

Agradecer también a la Coordinación para la Innovación y Aplicación de la Ciencia y la Tecnología (CIACYT) por las facilidades y apoyo brindado, así como al Laboratorio Nacional de Ciencia y Tecnología de Terahertz (LANCYTT), a los diferentes investigadores por sus conocimientos y consejos compartidos.

A los compañeros de laboratorio por su apoyo en el transcurso del posgrado, así como al SPIE quienes me brindaron la oportunidad de conocer otras formas de hacer ciencia, ampliar los horizontes, conocer otros entornos, culturas y a creer que se puede estar a la altura de otras personas en países de primer mundo.

A su vez quiero agradecer al Dr. Edgar Guevara por su apoyo e impulso a superarme en cada momento, así como por los consejos brindados en las diferentes etapas del posgrado con el objetivo de estar mejor preparado para los retos que se enfrenta un investigador en la vida diaria. Adicionalmente, el Dr. Edgar Guevara aprovecha para agradecer al proyecto 528 de “*Cátedras CONACYT*”, siendo este uno de los productos del apoyo financiero brindado.

Al Dr. Neville Hogan, quien en el 2010 fue una inspiración para mí el conocer proyectos de ciencia e ingeniería enfocados en mejorar la calidad de vida de las personas enfocados en la parte médica con su plática dada en la Facultad de Ingeniería, el tiempo dedicado años después en otra visita al platicar sobre posibilidades de cursar un posgrado y finalmente por permitirme cumplir un sueño de conocer su laboratorio en el Massachusetts Institute of Technology (MIT) en Cambridge, Massachusetts y darme cuenta con uno de sus alumnos que enfrentábamos problemas similares en el transcurso del posgrado.

Un especial agradecimiento al CONACYT que a través del programa de “*Becas Nacionales CONACYT*” con el número de beca 336988, así como por la “*Beca de Movilidad 2019 Nacional*”, número 291277, así como al COPOCYT por el apoyo especial en la modalidad B de la *Convocatoria para Promover actividades de Investigación, Desarrollo Tecnológico e Innovación (IDTI) del Consejo Potosino de Ciencia y Tecnología*, que junto con el apoyo económico de parte del Posgrado en Ciencias Aplicadas hicieron posible la estancia en la Universidad de Guanajuato, Campus León con la Dra. Claudia Luévano en el Departamento de Ciencias Médicas.

También agradezco el apoyo del COPOCYT al brindar el apoyo especial en la modalidad A de la *Convocatoria 2018 de Apoyos Especiales para Promover la Investigación, el Desarrollo Tecnológico y la Innovación (IDTI) del Consejo Potosino de Ciencia y Tecnología* para asistir a la Conferencia *Photonics North 2018* en Montreal, Québec, Canadá, siendo posible junto con el proyecto 32 del “*Centro Mexicano de Innovación en Energía Solar*” otorgado por el LANCYTT.

*Para mi mamá, mi abuela y mi familia, quienes  
siempre me han apoyado*

# Índice general

	Página
<b>Declaratoria de Autoría</b>	II
<b>Resumen</b>	III
<b>Abstract</b>	V
<b>Agradecimientos</b>	VII
<b>Dedicatoria</b>	VIII
<b>Índice General</b>	XII
<b>Índice de Figuras</b>	XIII
<b>Índice de Tablas</b>	XV
<b>1. Introducción</b>	1
1.1. Objetivos	2
1.2. Justificación	2
1.3. Hipótesis	3
<b>2. Cáncer de Mama</b>	4
2.1. Factores de Riesgo	6
2.2. Técnicas de Detección de Cáncer de Mama	7
2.3. Tumores de Cáncer de Mama	8
2.3.1. Tipos de Cáncer de Seno	10
2.3.2. Etapas del Cáncer de Seno	13
	X

<b>3. Termografía Infrarroja</b>	<b>16</b>
3.1. Termografía Infrarroja en Cáncer de Mama	20
3.2. Protocolo De Adquisición de Termogramas	22
3.3. Análisis Manual de Termogramas	23
<b>4. Aprendizaje Automático</b>	<b>26</b>
4.1. Matemática Aplicada Al Aprendizaje Automático	27
4.2. Aprendizaje Profundo	32
4.3. Redes Neuronales Convolucionales	41
4.3.1. Capas de Una Red Neuronal Convolutcional	43
4.3.2. AlexNet	45
4.3.3. GoogLeNet	46
4.3.4. Inception V3	47
4.3.5. VGG	48
4.3.6. ResNet	49
4.4. Transferencia de Aprendizaje	51
4.5. Aumento de Datos	53
4.6. Métricas de Desempeño	55
<b>5. Transferencia de Aprendizaje Automático en Termografía</b>	<b>59</b>
5.1. Primera Comparativa De Arquitecturas Pre-Entrenadas	60
5.2. Base de Datos de Termogramas	61
5.2.1. Análisis Estadístico de Normalidad	64
5.2.2. Generación de Datos Sintéticos	65
5.3. Selección de Mejor Arquitectura Pre-Entrenada	66
5.4. Búsqueda de Parámetros Óptimos para la Red	67
5.5. Comparativa Resolución de Imágenes	73
5.6. Comparativa Cámaras	76
5.7. Imágenes Ajustadas	76
5.8. Clasificación Categoría de la Lesión en la Mama	77
5.9. Análisis de Individual de las Bases de Datos	80

<b>6. Conclusiones</b>	<b>82</b>
<b>6.1. Trabajo a Futuro</b> . . . . .	83
<b>6.2. Publicaciones</b> . . . . .	85
<b>6.3. Participaciones en Congresos y Simposiums</b> . . . . .	85
<b>Apéndice A. Códigos Utilizados</b>	<b>86</b>
<b>Bibliografía</b>	<b>87</b>

# Índice de figuras

	Página
2.1. Tasas de Morbilidad y Mortalidad del Cáncer de Mama en México . . . . .	5
2.2. División de Mama en Cuadrantes . . . . .	5
3.1. Espectro Electromagnético . . . . .	16
3.2. Esquema de funcionamiento de una cámara infrarroja . . . . .	20
3.3. Capturas de pantalla de <i>Flir Tools</i> . . . . .	24
4.1. Cálculos de entrada y salida de una Neurona . . . . .	34
4.2. Representación de una Red Neuronal . . . . .	35
4.3. Representación Simple de Cálculos Realizados en una Red Neuronal . . . . .	36
4.4. Funciones de Activación . . . . .	39
4.5. Representación de Capas de una Red Neuronal Convolutacional . . . . .	44
4.6. Esquema de la arquitectura <i>AlexNet</i> . . . . .	46
4.7. Representación de las capas de la arquitectura GoogLeNet. . . . .	47
4.8. Representación de las capas de la arquitectura Inception V3. . . . .	48
4.9. Representación de las arquitecturas VGG. . . . .	49
4.10. Comparativa de un bloque regular y un bloque residual. . . . .	50
4.11. Representación de las capas de la arquitectura ResNet. . . . .	50
4.12. Representación de Transferencia de Aprendizaje . . . . .	52
4.13. Representación de Aumento de Datos de una Imagen . . . . .	54
4.14. Matriz de Confusión . . . . .	56
5.1. Diagrama de Flujo Análisis de Termogramas . . . . .	63

5.2. Bases de Datos Termogramas . . . . .	64
5.3. Grid Search Balanceado . . . . .	68
5.4. Grid Search Desbalanceado . . . . .	69
5.5. Curva ROC y Matrices de Confusión . . . . .	71
5.6. Clasificación Termogramas VS Sensibilidad Térmica . . . . .	74

# Índice de tablas

	Página
2.1. Descripción mamográfica de nódulos en las mamografías . . . . .	8
2.2. Manejo de Lesiones Mamarias Según BI-RADS . . . . .	9
4.1. Ventajas y desventajas de las redes profundas . . . . .	33
5.1. Resultados Primer Comparativa de Redes Pre-Entrenadas . . . . .	61
5.2. Distribución de pacientes por base de datos . . . . .	62
5.3. Datos de las Cámaras FLIR Utilizadas . . . . .	63
5.4. Métricas de Desempeño de Diferentes Arquitecturas con Clases Desbalanceadas . . . . .	67
5.5. Métricas de Desempeño de Diferentes Arquitecturas con Clases Balanceadas . . . . .	67
5.6. Resultados Parámetros Óptimos . . . . .	70
5.7. Comparativa de Berg <i>et al</i> mamografía con senos densos . . . . .	73
5.8. Redondeo Matrices Temperatura Red Desbalanceada . . . . .	74
5.9. Redondeo Matrices Temperatura Red Balanceada . . . . .	75
5.10. Prueba Clasificación por Cámaras . . . . .	76
5.11. Resultados Termogramas Ajustados . . . . .	77
5.12. Base de Datos Multiclases . . . . .	78
5.13. Clasificación Categoría de la Lesión en la Mama . . . . .	79
5.14. Resultados por centro de adquisición . . . . .	80
5.15. Integración de Grupos por enfoque. . . . .	80
5.16. Métricas de Desempeño por Grupos. . . . .	81

# Capítulo 1

## Introducción

El cáncer de mama es una de las enfermedades que más afecta a las mujeres tanto en el país como en el mundo, pero que en caso de ser descubierta de forma temprana tiene una alta tasa de supervivencia.

El método usado para tamizaje es la mamografía, sin embargo dentro de sus desventajas se encuentran una alta tasa de falsos positivos, la exposición a radiación y la incomodidad de las pacientes debido a la presión ejercida en su adquisición, así como el no ser recomendado cuando se tiene una edad inferior a los 50 años.

La termografía es un método que mide la radiación emitida por una superficie para determinar la temperatura de un cuerpo, revelando así anomalías en la temperatura del mama, siendo una forma en la que se manifiesta una enfermedad como el cáncer de mama. Es así que su uso puede ser ya sea de un pretamizaje que por su movilidad se utilice en comunidades donde el acceso a mastógrafos sea limitado, así como una herramienta adjunta que permita una mejor toma de decisiones para los expertos en el área de la salud.

De manera clásica, los termogramas son evaluados de manera manual, dependiendo de un puntaje en función de la vascularización del tejido sumado a la diferencia de temperatura que presente con su parte contralateral, sin embargo, es común que el criterio para otorgar el puntaje de vascularización que corresponderá al termograma sea diferente según la persona que lo analice, así como tener dificultades al momento de realizar la búsqueda de los puntos con mayor o menor temperatura en el mama y el poder acertar al hacer que coincida con el punto contralateral. Estas diferencias se acentúan cuando al realizar este protocolo se carezca de experiencia en la evaluación de las imágenes, presentándose así una **ausencia de estandarización** para la valoración de las imágenes.

El uso de técnicas de aprendizaje automático para la clasificación de imágenes es una de las aplicaciones que se le está tratando de dar para poder ser utilizadas por la parte clínica como herramientas que le permitan tener uniformidad en la decisión y que no sea necesario tener a un experto en la patología que se está estudiando para que un médico pueda ayudarse de éste y realizar un diagnóstico acertado.

Es necesario aclarar el hecho de que en ningún momento se trata de reemplazar a un especialista para dar un diagnóstico o tomar una decisión sobre el paciente, sino que estos métodos puedan ayudar a un médico de forma tal que la percepción visual de la persona y la ilusión óptica en la cual puede caer fácilmente una persona, no sean factores que afecten el desempeño de un experto, y que también permitan a un médico que no sea especialista en el tema a tomar una decisión con respecto a un paciente.

Contar con un **método automático** que proporcione **uniformidad en el juicio** de los mismos ayudará a que con esta técnica puedan ser detectados los pacientes que tienen una anomalía térmica en las mamas sin importar la experiencia de la persona que evalúe las imágenes o la carga de trabajo que tenga, y enfocarse en las áreas que sean más relevantes para su análisis.

## 1.1. Objetivos

El **objetivo general** del trabajo es el poder contar con un algoritmo de aprendizaje automático a través de la transferencia de aprendizaje de arquitecturas previamente entrenadas donde se pueda tener una clasificación de una base de datos de termogramas que arroje valores de sensibilidad y especificidad equivalentes a los que obtenga un experto en termografía.

Para lograrlo se tienen los siguientes **objetivos específicos**:

- ➔ Investigar la forma de realizar aprendizaje automático supervisado de imágenes.
- ➔ Comparar el desempeño de las arquitecturas al realizar la transferencia de aprendizaje a la clasificación de termografías.
- ➔ Identificar los elementos más relevantes de la red para que los termogramas sean clasificados en cada categoría.
- ➔ Conocer si el cambio de cámara infrarroja influye en la clasificación de termogramas o el algoritmo es insensible a estos cambios.

## 1.2. Justificación

El uso de la termografía puede ser útil para poder detectar de manera temprana la aparición de anomalías que pueden traducirse en cáncer de mama en pacientes donde se tiene un tejido denso y la mamografía no puede dar un resultado concluyente, esto puede quedar de manifiesto en las pacientes que tienen menos de 50 años de edad.

Existen varios estudios que intentan hacer una clasificación automática de los termogramas, pero para efectuarlo realizan una segmentación manual de varias características y una vez que eligen ciertos rasgos es que lo meten a un clasificador.

En el presente estudio se propone un **método de clasificación para imágenes ya existentes** que ayude a **mejorar la estandarización** de esta labor. En dicho método, no se realizará ningún procesamiento previo a la imagen, de esta forma, el algoritmo, en función de los rasgos que aprenda, determinará cuáles termogramas son los que pueden ser clasificados como anómalos y aquellos que no presenten una anomalía térmica mediante el uso de transferencia de aprendizaje de arquitecturas de redes neuronales convolucionales, las cuales han sido entrenadas previamente con más de un millón de imágenes que son capaces de clasificarlas de forma correcta en mil diferentes categorías.

La determinación de cuáles imágenes pertenecen a cada categoría está dada por un diagnóstico hecho mediante la biopsia, o mediante la clasificación BI-RADS de mamografía según lo establecido en las diferentes bases de datos empleadas en el estudio.

### 1.3. Hipótesis

Con lo anterior, las hipótesis empleadas en el presente trabajo son:

- $H_0$ : No es posible realizar una clasificación de termogramas de forma automática empleando redes neuronales convolucionales logrando valores de sensibilidad y especificidad similares a las reportadas en la técnica por expertos en el campo.
- $H_1$ : Es posible realizar una clasificación de termogramas de forma automática empleando redes neuronales convolucionales logrando valores de sensibilidad y especificidad similares a las reportadas en la técnica por expertos en el campo.

## Capítulo 2

# Cáncer de Mama

Las mamas femeninas tienen tamaño, forma y simetría variables [1]. El pezón está rodeado de un área pigmentada, ligeramente sobreelevada y circular llamado areola [1]. Normalmente el término ‘*seno*’ y ‘*glándulas mamarias*’ son utilizados como equivalentes, aunque de manera estricta no son sinónimos, debido a que las mamas contienen tejidos (grasa, vasos, nervios, etc.) que son distintos a los glandulares [2].

Según lo comentado en el libro de Moore *et al* [1], las mamas son divididas en cuatro cuadrantes para la localización anatómica y descripción de quistes y tumores como se muestra en la Figura 2.2. En el mismo texto se indican a los inferiores como las regiones menos vascularizadas cuando se requiere una incisión quirúrgica [1]. La piel de la mama no difiere estructuralmente de la pared torácica vecina, pero la piel del pezón y la areola es altamente especializada, habiendo variaciones raciales e individuales en el tamaño, forma y color de la areola y pezón [2].

Los dermatólogos consideran que los **enrojecimientos** o **edemas** (hinchazones causadas por acumulación de líquido en los tejidos) de la piel de la mama pueden ser debido a cáncer subyacente por lo que las mamas deben ser examinadas cuidadosamente en todos los pacientes con enfermedades de la piel que puedan ser asociados con malignidad sistemática [2].

El **cáncer** describe desórdenes asociados con células que tienen un crecimiento irregular y conducen a la formación de un tumor, invasión en tejidos circundantes y la difusión a otras partes del cuerpo, teniendo algunos factores que incrementan las probabilidades de las personas de ser afectadas por esta enfermedad, tales como la predisposición genética, edad, estilo de vida de la persona, así como la exposición a agentes cancerígenos [3].

Los datos para México [4] no varían mucho de lo presentado a nivel mundial, ya que representan el cáncer que más presencia tiene de todos los cánceres diagnosticados en el año con 14.3% de toda la población, siendo así el 16% del total de los casos de cáncer en mujeres, con una mortalidad de 9.9%, muy cercano al de próstata que tiene un 10%.

Omitiendo los datos del cáncer de piel no melanoma, el cáncer de mama es a nivel mundial el más frecuente en las mujeres de todas las edades [4, 6, 8], siendo el 24.2% de los nuevos cánceres detectados al año, presentando la tasa de incidencia más alta (46.3%) y el segundo de mayor mortalidad (13%), solo por detrás del cáncer de pulmón.

En el cáncer de mama se da un crecimiento de las células de manera no controlada en las cuales se da un crecimiento que invade el tejido circundante y se expande de manera sistemática [9], el cual puede causar también aplastamiento, ampliación o inversión del pezón [2].



Figura 2.1: El mapa realizado por Castrezana Campos [5] muestra la tasa de morbilidad y mortalidad del cáncer de mama en México

Existen ciertos signos y síntomas del cáncer de mama; para diferenciar, los **signos** son objetivos, es decir, no dependen de la percepción del paciente, sino que son totalmente observables o pueden ser medidos, por su parte, los **síntomas** solamente pueden ser sentidos por el paciente, del cual el más común es una nueva masa o protuberancia; cuando esta no es dolorosa, dura y con bordes irregulares es más probable que pueda ser cáncer, aunque esto no excluye que aquellos que sean sensibles a la palpación, blandos, redondeados y causen dolor no lo sean. [10].



Figura 2.2: División de Mama en Cuadrantes según lo indicado por Moore *et al* [1].

Algunos otros síntomas que pueden presentarse son:

- Hinchazón de parte o de toda la mama

- Irritación o hendiduras en la piel
- Dolor en la mama o pezón
- Retracción (contracción) de los pezones
- Enrojecimiento, descamación o engrosamiento de la piel de la mama o del pezón
- Secreción de la mama que no sea leche materna

Existen diferentes técnicas de tamizaje empleadas para detectar esta enfermedad, tales como la mamografía, resonancia magnética, termografía y ultrasonido [9], en las cuales entre **mayor** sea el valor de la **sensibilidad**, **mayor** será la **probabilidad** de poder **detectar un tumor** en un paciente; entre **mayor** sea la **especificidad**, se tendrá una **mayor probabilidad** de que un **diagnóstico positivo** sea **verdadero** [3].

## 2.1. Factores de Riesgo

Algunos de los factores de riesgo para que aparezca el cáncer de mama son el tener una menstruación previa a los 12 años, el inicio de la menopausia posterior a los 55, edad a la cual tuvieron el primer parto, lactancia a los hijos, evitar el uso de terapia hormonal, así como antecedentes familiares de la enfermedad; Shawon *et al* [11] realizó un estudio en el que asocia de manera inversa al tamaño corporal en la niñez y la juventud con el riesgo de presentar cáncer de mama en la etapa adulta al indicar que con tamaños corporales más grandes a los 7 y 18 años se llega a tener un riesgo reducido de cáncer de mama tanto previos como posteriores a la menopausia.

En el mismo estudio se menciona que al tener un tamaño corporal mayor a los 18 años ha sido asociado con pequeños tumores, mientras que a los 7 años se tiene una reducción del 27% del riesgo de tener cáncer postmenopáusico, en comparación con los tamaños corporales chicos, y que al reducir su tamaño corporal de los 7 a los 18 años se tienen probabilidades más bajas de tener cáncer de mama contra los que se mantuvieron en una talla corporal chica.

Aunque el cáncer de mama también puede aparecer en hombres, este es relativamente raro, siendo cerca del 0.2% de todos los cánceres de hombres en los Estados Unidos, y menos del 1% de todos los cánceres de mama, siendo algunos factores de riesgo asociadas a una función testicular reducida, exceso de exposición a estrógeno, síndrome de Klinefelter y a una historia familiar de carcinoma de mama, aunque estudios de control indican que la enfermedad es más común en hombres de raza negra así como en aquellos que nunca se casaron [2].

Existe un riesgo atribuible a la población, el cual se refiere a la proporción de casos de enfermedad que no hubieran ocurrido si el factor de riesgo fuera eliminado de la población [12], siendo una asociación entre el factor de riesgo y la enfermedad, así como la prevalencia de la enfermedad en la población, con lo que se analiza la disposición que tiene para contraer la enfermedad.

Si bien es cierto que el cáncer no se puede prevenir, el estudio comandado por [13] sugiere que si el índice de masa corporal es inferior a  $25 \frac{kg}{m^2}$  puede reducir cerca del 16.2% los casos de cáncer de mama, indicando que es uno de los factores de riesgo que es fácilmente medible y modificable.

De esta forma, se puede observar que ciertos hábitos puedan ayudar a reducir los factores de riesgo, tales como mantener un peso saludable, mantenerse físicamente activo, tener una buena alimentación, no fumar y limitar el consumo de alcohol.

## 2.2. Técnicas de Detección de Cáncer de Mama

Una de las técnicas que más promoción se hace en medios de comunicación masivos para detectar el cáncer de mama es la autoexploración, en donde el principal fin es familiarizarse con su apariencia, textura y anatomía [14, 15], sin embargo esta técnica tiene una baja sensibilidad que ronda entre el 12 % y 14 % [16], Godavarty *et al* [17] indica que en ensayos de control aleatorio no lleva a un efecto positivo y lleva en realidad a falsos positivos.

La **mamografía** se utiliza para evaluar la estructura anatómica de la mama permitiendo que las anormalidades sean identificadas [18], es la única técnica de diagnóstico aceptado como tamizaje para el cáncer de mama [19, 20]; la presencia del tumor se detecta utilizando rayos X para mostrar la anatomía interna de una mama al detectar masas y calcificaciones con lo cual se ha reducido la mortalidad de esta enfermedad cuando se realizan exámenes de manera regular [3, 21].

Tiene una **sensibilidad** que es altamente variable, ya que para mujeres con pecho graso llega a alcanzar 98 %, mientras que para mujeres con tejido denso es de 36 % [22, 23], aunque en general se le considera de 84 % [3]; por su parte, la **especificidad** del método es de 92 % [3] siendo por ello el método que llega a ser más recomendado para tamizaje así como por su costo moderado. Kerlikowske *et al* [24] indica que la mamografía tiene una tasa de sensibilidad mayor para mujeres de 50 años o más debido al incremento en el tejido graso contenido.

Sin embargo, el método tiene una alta tasa de falsos positivos, siendo uno de los factores más importantes cuando la paciente tiene un **tejido mamario denso** disminuyendo la sensibilidad cuando la paciente es menor de 50 años, ya que además de tener un tejido más denso, se tienen cánceres biológicamente más agresivos y a una prevalencia reducida de la enfermedad; de igual forma, la capacitación y habilidad del radiólogo para el análisis, presentando una tasa excesiva de rellamados provocando que las pacientes no asistan a la segunda prueba de tamizaje, se tengan más dosis de radiación, que se realice un alto número de biopsias innecesarias, y ansiedad del paciente [3, 8, 9, 24-26].

El estudio llega a ser una experiencia incómoda para la paciente en la cual llega a sentir dolor por la compresión de la mama con lo que se pueden presentar riesgos de ruptura alrededor del tumor canceroso [18] desde las 22 libras de presión, en los equipos de mamografía se utilizan 42 libras, con el riesgo potencial de enviar las células malignas a la sangre; además que en algunos casos, la radiación a la que está expuesta la paciente por lo que algunos estudios sugieren que pudiera ser un factor que incremente la posibilidad de incidencia de la enfermedad [27].

El **ultrasonido** es otra técnica utilizada, que permite mostrar imágenes del tejido de la pared torácica con lo que se pueden distinguir quistes, que son sacos no tumorosos llenados con fluido, de masas sólidas, desempeñándose especialmente bien cuando trabaja con tejido denso, con una **sensibilidad** de 82 % y una **especificidad** de 84 % siendo una técnica con amplia disponibilidad, bajo costo y que tiene una incomodidad tolerable por los pacientes por lo que es empleada como suplemento a la mamografía [3, 26, 28].

Una de las desventajas de la técnica, es que demora aproximadamente 19 minutos en realizar la prueba bilateral, esto sin incluir los estudios previos, análisis del resultado con los pacientes y los requerimientos del protocolo [26].

En un estudio de cohorte prospectivo realizado en 529 mujeres asintomáticas, Kuhl *et al* [29] sugiere que aún con el uso combinado de mamografía y ultrasonido no se pudieron detectar la mitad de los cánceres que fueron encontrados (43 cánceres en 41 pacientes).

La **imagen de resonancia magnética**, siendo un estudio es recomendable en personas que tienen un alto riesgo de tener cáncer de mama; posee una sensibilidad de 90 % y 81 % de especificidad.

Sin embargo, el uso de la resonancia magnética es limitado debido a su alto costo, una limitada disponibilidad y experiencia al mejorarlo [26].

De igual forma, una de sus desventajas es que deben de utilizarse agentes radiológicos que permitan mejorar el contraste de la imagen con lo que se puedan optimizar ciertas propiedades físico-químicas [30]. Aunque se consideran biológicamente inertes, se ha mostrado que pueden llegar a tener efectos en los sistemas cardiovasculares y renales, así como consideraciones especiales cuando se trata con niños y mujeres embarazadas [30].

A su vez, tiene casi cuatro veces más rellamados de pacientes que la mamografía en mujeres con una historia familiar con mayor riesgo de padecer cáncer de mama con el 70% de ellos falsos positivos [31], la cantidad de tiempo que se necesita para su adquisición y la tolerancia del paciente para mantenerse en espacios cerrados [17].

Otros estudios que se emplean en esta labor son el escaneo de impedancia eléctrica (87% de sensibilidad y 82% de especificidad), así como la mamografía 3D (90% de sensibilidad y 92% de especificidad) y la termografía de la cual se hablará ampliamente en capítulos posteriores.

Cuando una paciente tiene implantes de mama, la mamografía no es algo ideal ya que el implante puede dañarse, o estar ocultando el tumor atrás de este; la termografía puede ser efectivo ya que los implantes generalmente son colocados en la base de la mama y no afectan significativamente la transferencia de calor en el resto de la mama, aunque es necesario desarrollar una técnica que sea más efectiva en este tipo de pacientes [3].

### 2.3. Tumores de Cáncer de Mama

Los hallazgos que se encuentran en la mama a partir de la mamografía, son clasificados mediante el sistema **BI-RADS** (*Breast Imaging Reporting and Data System*), el cual se considera el idioma universal del tamizaje de la patología mamaria con lo cual se pueden estandarizar y categorizar las lesiones al establecer el grado de sospecha y saber las acciones que se deban de tomar en cada uno de los casos; esto puede observarse en la Tabla 2.1 [19].

Forma	Contorno	Densidad
♠ Redondeada (R)	♠ Bien definido (D)	♠ Superior (+)
♠ Oval (O)	♠ Oscurecido (O)	♠ Similar (=)
♠ Lobulada (L)	♠ Mal Definido (I)	♠ Inferior (-)
♠ Irregular (X)	♠ Microlobulado (M)	
	♠ Espiculado (S)	

Tabla 2.1: Descripción mamográfica de nódulos en las mamografías

En cada categoría de patrón mamográfico existen 4 categorías para las cuales se tienen lecturas que son:

- a) *Grasa*: Mama de composición predominantemente grasa
- b) *Densidad media*: Mama con tejido fibroglandular disperso

- c) *Heterogénea*: Mama con tejido glandular heterogéneamente denso
- d) *Densa*: Mama con parénquima glandular extremadamente denso que puede ocultar lesiones

Se debe diferenciar a un **nódulo** o **masa** la cual es una lesión ocupante de espacio que puede ser vista en dos proyecciones diferentes, mientras que se considerará **densidad** o **asimetría** cuando la lesión solamente se ve en una única proyección [19]; el manejo de las lesiones mamarias según el BI-RADS puede verse en la Tabla 2.2; generalmente a partir de que se tiene un BI-RADS 3 en el ambiente clínico se recomienda que se realice una biopsia para tener un diagnóstico de la paciente.

BI-RADS 0	No concluyente por lectura incompleta Necesitan realizarse pruebas de imagen adicionales y/o mamografías previas para comparar
BI-RADS 1	Mama Normal No se identifican hallazgos mamográficos comprendidos entre las categorías 2 y 5 de sospecha Hacer mamografía en dos años
BI-RADS 2	Benigna, probabilidad de cáncer similar a la población en general Son aquellos que cumplen con el siguiente criterio: Nódulos con contenido graso, calcificados y solitarios, así como microcalcificaciones comprendidas en el grado de sospecha típicamente benignas Hacer mamografía en dos años
BI-RADS 3	Hallazgos probablemente benignos Se usa para un grupo de lesiones de mama con criterios no definitivamente benignos de acuerdo a los estándares establecidos La probabilidad de malignidad es bajo, menor al 2 % Se recomienda un control a los 6, 12 y 24 meses para valorar si hay cambios de categoría o si se mantiene en esta para posteriormente pasar a la 2 A los 24 meses hacer una revisión anual
BI-RADS 4	Probablemente maligno Valor predictivo para cáncer del 24 % al 70 % <ul style="list-style-type: none"> <li>■ Categoría 4-A: Hallazgo mamográfico que requiere biopsia, baja sospecha de malignidad</li> <li>■ Categoría 4-B: Sospecha intermedia de malignidad</li> <li>■ Categoría 4-C: Preocupación moderada pero no clásica de malignidad, derivación hospitalaria</li> </ul>
BI-RADS 5	Sugerencia alta de malignidad Valor predictivo positivo por encima del 70 % Derivación hospitalaria
BI-RADS 6	Malignidad confirmada histológicamente, pero antes de iniciarse un tratamiento definitivo

Tabla 2.2: Manejo de Lesiones Mamarias Según BI-RADS

Una vez que se ha detectado un tumor, debe de conocerse el **grado de un tumor**, refiriéndose a la descripción basada de cómo se ven las células y el tejido de un tumor al microscopio y la rapidez con la que probablemente crecerá y se extenderá el tumor [32].

Por su parte, el **estado de un cáncer** indica el tamaño o etapa (extensión) del tumor original

y si las células se han diseminado en el cuerpo o si están afectados los ganglios linfáticos regionales, además del número de tumores presentes.

Los sistemas de graduación de los tumores difieren dependiendo del tipo de cáncer en:

- *Grado 1*: Células tumorales y la organización del tejido del tumor tienen una apariencia cercana a la normal. Tienden a crecer y a diseminarse lentamente; es bien diferenciado en un grado bajo.
- *Grado 2*: Moderadamente diferenciado (grado intermedio); crecen más rápido de lo normal.
- *Grado 3*: Tiende a crecer rápidamente, células muy anormales, grado alto.
- *Grado 4*: Indiferenciado, grado alto.
- *Grado X*: No es posible asignar un grado, es indeterminado.

Cuando se habla de cáncer de mama, los médicos usan con más frecuencia el **sistema de graduación Nottingham** basándose en [32]:

- *Formación tubular*: Qué tanto del tejido del tumor tiene estructuras normales de conductos de leche de la mama.
- *Grado nuclear*: Evaluación del tamaño y forma del núcleo en las células tumorales.
- *Rango mitótico*: Cuántas células que se dividen están presentes, lo cual es una medida de la rapidez con la que las células del tumor están creciendo y se están dividiendo.

A cada categoría se le asigna una puntuación del 1 al 3, siendo el 1 normal y 3 muy anormal en su mayoría; las puntuaciones se añaden luego dando un resultado entre 3 y 9 [32]:

- G1: Puntuación total entre 3 y 5
- G2: Puntuación total entre 6 y 7
- G3: Puntuación total entre 8 y 9

### 2.3.1. Tipos de Cáncer de Seno

Existen diferentes tipos de cáncer de mama, la gran mayoría son carcinomas (tumores malignos), los cuales se originan en las células que revisten los órganos y los tejidos; a menudo los cánceres de mama son un tipo de carcinoma llamado **adenocarcinoma** que se origina en las células que producen glándulas (tejido glandular), en este caso, originados en los conductos galactóforos o en los lobulillos (glándulas productoras de leche) [10].

Otros tipos de cáncer de mama también pueden ocurrir en las células del músculo, grasa o tejido conectivo; algunas veces el tumor puede ser una **combinación de diferentes tipos**; es posible que algunos tipos muy pocos comunes de cáncer de mama, las células cancerosas no formen un tumor en absoluto.

Cuando se hable de un **cáncer *in situ*** se refiere que no se ha propagado, mientras que el **cáncer invasivo o infiltrante** se ha propagado o invadido al tejido circundante del mama [10]. La información aquí vertida puede encontrarse en [2, 10, 33] donde se habla con más detalle de estos.

## Tipos comunes de cáncer de mama

### Carcinoma ductal *in situ*

Conocido también como carcinoma intraductal, es un cáncer de mama *no invasivo* o *preinvasivo*.

### Carcinoma lobulillar *in situ*

También se le denomina neoplasia lobulillar; este cambio de la mama *no es un cáncer*, sino que las células con aspecto de cancerosas crecen en los lobulillos de las glándulas productoras de leche de la mama, pero no atraviesan la pared de los lobulillos.

### Carcinoma ductal invasivo o infiltrante

Es el *tipo más común* de cáncer de mama; comienza en un conducto lácteo de la mama, penetra a través de la pared del conducto y crece en el tejido adiposo de la mama; normalmente es del 50 % al 75 % de los cánceres detectados. Se tienen algunos subtipos, dentro de los cuales se pueden dividir en dos formas:

- Mejor pronóstico que el carcinoma ductal infiltrante convencional:
  - Carcinoma quístico adenoide (adenoquístico).
  - Carcinoma adenoescamoso de bajo grado (tipo de carcinoma metaplástico).
  - Carcinoma medular.
  - Carcinoma mucinoso (o coloide).
  - Carcinoma papilar.
  - Carcinoma tubular.
- Pronósticos iguales o peores que el carcinoma ductal infiltrante regular:
  - Carcinoma metaplástico.
  - Carcinoma micropapilar.
  - Carcinoma mixto (tiene características de ductal invasivo y lobulillar).

### Carcinoma lobulillar invasivo o infiltrante

Comienza en las glándulas productoras de leche (lobulillas), se puede propagar a otras partes del cuerpo; son del 5 % al 15 % de los cánceres detectados.

## Tipos de cáncer menos comunes

### Cáncer de mama inflamatorio

Cáncer invasivo que representa del 1 % al 3 % del total de los cánceres de mama.

### Enfermedad de Paget del pezón

Comienza en los ductos de la mama y se propaga hacia la piel del pezón y después hacia la areola, es el 1 % de todos los cánceres de mama. La gran mayoría de Paget de seno tienen también uno o dos tumores en el interior de éste, los cuales son carcinomas ductales *in situ* o cáncer invasor de mama; las células de Paget son un signo revelador del cáncer, teniendo una apariencia grande y redonda al microscopio. Tienen síntomas como picazón, cosquilleo o enrojecimiento, piel descamada en o alrededor del pezón, pezón

aplanado, así como secreción del pezón que puede ser amarillenta o sanguinolenta. Muchas de las ocasiones tiene un diagnóstico inicial equivocado por lo que coadyuva a que la supervivencia con esta enfermedad sea precaria.

### **Tumores filoides**

Son poco comunes, se originan en el tejido conectivo (estroma) de la mama.

### **Angiosarcoma**

Ocurre muy poco en los senos, se origina en las células que revisten los vasos sanguíneos o los vasos linfáticos.

### **Carcinoma metastásico inflamatorio**

También conocido como carcinoma *erysipeloides*, es una rara forma de metástasis cutánea que se presenta como eritematoso; representa la diseminación linfática de un tumor primario, a menudo aparece después de la quimioterapia, radioterapia o cirugía de escisión tumoral.

### **Carcinoma en coraza *cuirasse***

Rara manifestación de cáncer de mama avanzado caracterizado por fibrosis extensiva y progresiva, además de rigidez de la piel del pecho, cuello, espalda y abdomen.

### **Carcinoma metastásico telangiectásico**

Se caracteriza por la aparición de pápulas y vesículas violáceas que se asemejan a un linfangioma circunscrito, que son lesiones en placa con numerosas vesículas rellenas de un líquido de aspecto transparente o hemorrágico.

### **Carcinoma metastásico nodular**

Usualmente aparece como nódulos firmes y múltiples, que se pueden asemejar a un queratoacantoma, la lesión es redondeada y tiene un centro costroso, así como un rodete periférico de piel con aspecto normal o eritematoso.

### **Alopecia neoplástica**

Puede aparecer como áreas aloécicas circulares que sugieren alopecia areata, los cuales refieren a parches sin cabello ovalados o redondeados de 2 a 5cm de diámetro en el cuerpo cabelludo). Típicamente son parches ovalados bien demarcados, a menudo con un tono eritematoso rosado y de superficie lisa, sin dolor ni prurito.

### **Carcinoma de masa metastásico del pliegue inframamario**

Se caracteriza por la aparición de un nódulo cutáneo, ubicado en el pliegue inframamario que sigiere carcinoma basocelular o espinocelular. A menudo ocurre en mujeres con mamas péndulas que imitan una dermatitis intertriginosa.

### **Carcinoma metastásico mamario del párpado con histología histiocitoide**

Clínicamente se presenta como un edema y ocasionalmente un nódulo indoloro discreto de la localización palpebral.

### 2.3.2. Etapas del Cáncer de Seno

El identificar las etapas que se tienen en el cáncer de mama es esencial para poder saber tanto su gravedad como el tratamiento, tales como el tipo, su tamaño y propagación. Los cánceres se identifican como etapas, si es 0 es un **carcinoma *in situ***, y posteriormente del 1 al 4 pudiéndose dividir en subetapas A, B y C; la información aquí vertida es tomada de [10, 34] donde puede consultarse también con mayor profundidad.

El cáncer de mama se clasifica por etapas con el **sistema TNM** de la American Joint Commission on Cancer (AJCC) el cual está basado en:

- **T:** Tamaño del tumor y si crece a áreas cercanas
- **N:** Si alcanza los ganglios linfáticos (nódulos) cercanos
- **M** Si ha hecho metástasis (propagación) a otras partes del cuerpo

#### **Etapas 0**

##### **Tis, N0, M0**

Se considera carcinoma ductal *in situ* (DCIS), un precáncer de mama o la forma más temprana de éste; hay células en el interior del conducto, por lo que no han invadido de manera profunda el tejido adiposo circundante de la mama.

Algunos tipos de este son el carcinoma lobulillar *in situ*, aunque no todos lo consideran cáncer o precáncer, así como la enfermedad de Paget del pezón. Para ninguno de estos se propaga en ganglios linfáticos ni áreas distantes.

#### **Etapas I A**

##### **T1, N0, M0**

El tumor mide 2cm o menos de ancho (T1), no se ha propagado a los ganglios linfáticos (N0) ni áreas distantes.

#### **Etapas I B**

##### **T0 ó T1, N1mi, M0**

El tumor no se detecta o mide menos de 2cm de ancho, los ganglios linfáticos de la axila miden más de 0.2mm de ancho y/o contiene más de 200 células, pero no mide más de 2mm (N1mi), el cáncer no se ha propagado a áreas distantes (M0).

#### **Etapas II A**

##### **T0 ó T1, N1, M0**

El tumor mide menos de 2cm de ancho o no es detectado (T0 ó T1), el cáncer no se ha propagado a áreas distantes (M0) y alguna opción:

- Cáncer propagado hasta debajo del brazo, mide más de 2mm de ancho en los ganglios linfáticos (N1a).
- Diminutas cantidades de cáncer en los ganglios linfáticos mamarios internos cercanos al esternón (N1b).

- Cáncer propagado a entre uno y tres ganglios linfáticos axilares o a ganglios linfáticos mamaros internos (N1c).

**T0 ó T1, N1, M0**

El tumor mide entre 2 y 5cm de ancho (T2), no se ha propagado a ganglios linfáticos (N0) ni a áreas distantes (M0).

**Etapa II B****T2, N1, M0**

El tumor tiene entre 2 y 5cm de ancho (T2), cáncer propagado a entre 1 y 3 ganglios linfáticos (N1), no se ha propagado a sitios distantes (M0).

**T3, N0, M0**

El tumor mide más de 5cm pero no ha invadido la pared torácica o piel (T3), el cáncer no se ha propagado a ganglios linfáticos (N0) ni sitios distantes (M0).

**Etapa III A****T0 a T2, N2, M0**

El tumor no se encuentra o mide menos de 5cm (T0 a T2), se propagó a entre 4 y 9 ganglios linfáticos axilares o agrandando los mamaros internos (N2), el cáncer no se ha propagado a sitios distantes (M0).

**T3, N1 ó N2, M0**

El tumor tiene más de 5cm de ancho, no ha invadido la pared torácica o de la piel (T3). Se propagó a entre 1 y 9 ganglios linfáticos axilares o mamaros internos (N1 ó N2), no se ha propagado a sitios distantes (M0).

**Etapa III B****T4, N0 a N2, M0**

El tumor ha crecido en la pared torácica o en la piel (T4), el cáncer no se ha propagado a sitios distantes (M0) con alguna de estas opciones:

- Cáncer no se ha propagado a ganglios linfáticos (N0).
- Cáncer se propaga a entre 1 y 3 ganglios linfáticos axilares y/o diminutas cantidades en ganglios linfáticos mamaros internos (N1).
- Se propagó a entre 4 y 9 ganglios linfáticos axilares o se ha agrandado los ganglios linfáticos mamaros internos (N2).

**Etapa III C****Cualquier T, N3, M0**

Cualquier tipo de tumor, el cáncer no se ha propagado a *siti*; se tienen situaciones como:

- Propagación del cáncer a 10 o más ganglios linfáticos auxiliares (N3).
- Propagación del cáncer a los ganglios linfáticos debajo de la clavícula (N3).
- Propagación del cáncer a los ganglios linfáticos sobre la clavícula (N3).
- Cáncer afecta a ganglios linfáticos axilares y ha agrandado los ganglios linfáticos mamarios internos (N3).
- Ganglios linfáticos axilares, cáncer se propagó a 4 o más, diminutas cantidades de cáncer en los ganglios linfáticos mamarios internos (N3).

#### **Etapa IV**

##### **Cualquier T, Cualquier N, M1**

Aquí el cáncer se ha propagado a órganos distantes o ganglios linfáticos distantes a la mama. Los lugares donde el cáncer se propaga con más frecuencia son los huesos, hígado, cerebro o pulmones.

Generalmente es difícil encontrarlo *in situ*, lo más seguro de encontrarlo en tacto es con tumores grandes y son metastásicos.

## Capítulo 3

# Termografía Infrarroja

De forma histórica, la temperatura se considera como un buen indicador de salud, Wunderlich *et al* [35] estableció que al tener un rango de temperatura entre 36.3 y 37.5°C se puede considerar como normal, por lo que cualquier anomalía entre este rango de temperatura puede ser un indicativo de enfermedad. Una cámara infrarroja permite detectar las firmas de calor y consiste en un lente y un arreglo detector posicionado en el plano focal de la lente. El lente está hecho de un material especializado que es transparente a la luz infrarroja pero tiene propiedades refractarias; el arreglo detector tiene la habilidad de absorber la energía de la luz infrarroja y producir una señal eléctrica [36].

Las diferentes ondas electromagnéticas están agrupadas dentro del espectro electromagnético; es posible organizarlas de acuerdo con su frecuencia o su longitud de onda los cuales tienen una relación inversamente proporcional [37] tal como se muestra en la Figura 3.1, es en el cercano infrarrojo (700nm a 1200nm) donde la luz tiene una penetración máxima en el tejido.

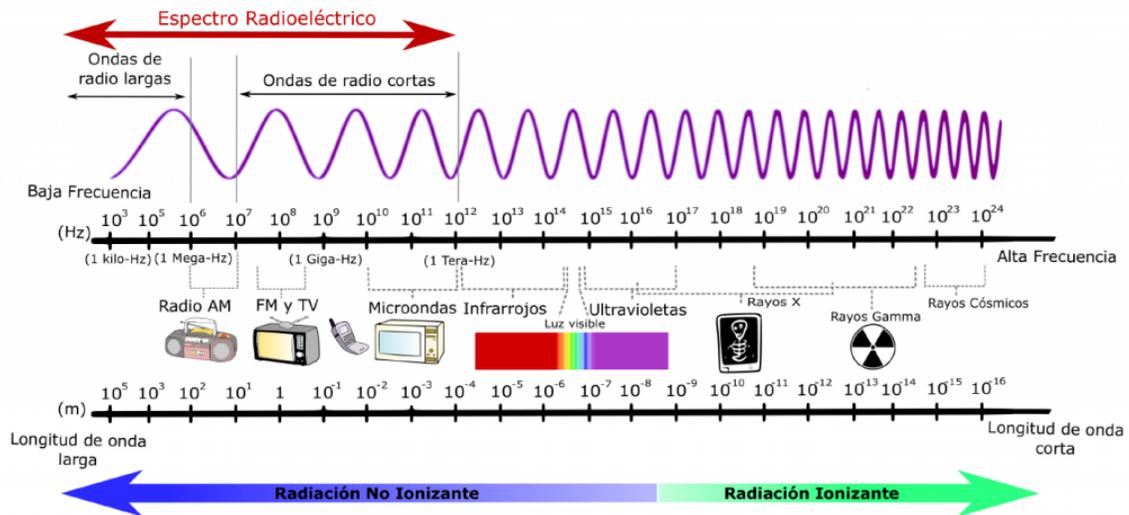


Figura 3.1: Espectro Electromagnético [38]

La radiancia (o resplandor, con símbolo  $L$ ) es la cantidad elemental de la radiometría la cual es la potencia por unidad de área y por unidad de ángulo sólido proyectado; la energía radiante tiene el

símbolo  $Q$  con unidades en Joules. El flujo radiante se refiere a la energía por unidad de tiempo con el símbolo  $\phi$  y unidades en Watts [39] y se toma en cuenta el ángulo  $\theta$  entre la normal al elemento fuente y la dirección de la observación. Por su parte, la irradiancia, o incidencia radiante, es el opuesto de la excitación que se refiere a la potencia por unidad de área que incide sobre la superficie, representado con el símbolo  $E$  y se calcula integrando la radiancia sobre el ángulo sólido proyectado de un hemisferio.

$$L = \frac{d^2\phi}{dA_s \cos\theta d\omega} = \frac{d^2\phi}{dA_S d\Omega} = \frac{d^2\phi}{dA_p d\omega} \quad (3.1)$$

$$E = \int_{\pi} L d\Omega \quad (3.2)$$

La radiancia es asociada a una fuente que puede ser activa (como la luminiscencia o térmica) o pasiva; un ejemplo de fuente térmica es un cuerpo negro, o un humano [39]. La radiancia es invariante en su orden para conservar la potencia, con lo cual, se tiene que la radiancia de la imagen en el plano del detector de una cámara es la misma que la radiancia de una escena si no hay pérdidas de transmisión debido a la atmósfera o la óptica, así como el que la radiancia del plano focal de un radiómetro es la misma que la del objetivo considerando que tampoco existen pérdidas de transmisión [39].

La **emisividad** se define como la tasa de radiancia de un cuerpo dado al de un cuerpo negro, se representa con el símbolo  $\epsilon$ ; algunos le consideran un factor de calidad ya que indica la capacidad de radiación de un material [39, 40]. Se considera que es la energía radiada por un material entre la energía radiada por un cuerpo negro [36].

Desde finales del siglo XIX se ha estudiado de forma amplia la radiación térmica, siendo en 1901 cuando Planck publicó una ecuación empírica en donde se incluyó la noción de que la energía existía en forma de “paquetes” discretos; verificaciones sobre esta ecuación y las constantes físicas continuaron hasta que fue posible medir la distribución espectral de la radiación de un cuerpo negro, la determinación de la constante de Stefan-Boltzmann y la confirmación de la escala de temperatura termodinámica [39]. La ley de Planck se presenta en la (Ecuación 3.3), tiene unidades de  $\frac{W}{m^2 sr \cdot m}$ , esta indica que cualquier objeto que tenga una temperatura por encima del cero absoluto emitirá una radiación electromagnética cuyo espectro e intensidad son función de la temperatura del cuerpo.

$$L_{\lambda} = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \quad (3.3)$$

Ecuación 3.3 Ley de Planck

Para determinar la radiancia total de radiación de un cuerpo negro, se integra la ecuación de Planck sobre todas las longitudes de onda con lo cual resulta la ley de Stefan-Boltzmann (Ecuación 3.4), en donde  $\sigma$  es la constante de Stefan-Boltzmann con valor de  $5.670 \times 10^{-8} \frac{W}{m^2 K^4}$ . La ley de Stefan-Boltzmann es más frecuente verla en términos de la radiación excitante con  $n = 1$ , de donde queda como se expresa en la Ecuación 3.5. El cuerpo humano a temperatura normal tiene un pico de emisión con una longitud de onda alrededor de  $10 \mu m$ , por lo que en aplicaciones médicas se utiliza la banda de longitud de onda entre los 8 y los  $10 \mu m$  [41].

$$L = \int_0^{\infty} L_{\lambda} d\lambda = \frac{n^2 \sigma}{\pi} T^4 \quad (3.4)$$

Ecuación 3.4 Ley de Stefan-Boltzmann

$$M = \sigma T^4 \quad (3.5)$$

Ecuación 3.5 Ley de Stefan-Boltzmann en términos de la radiación excitante

Un detector de radiación óptica es un transductor que transforma la energía óptica radiante en una forma diferente de energía que es más fácil de medir, los cuales pueden ser detectores de un punto o de un área [39]. Dentro de los detectores térmicos se encuentran los bolómetros y los pares termoeléctricos para la detección de un punto, mientras que para un área están los arreglos microbolométricos y arreglos lineales termoeléctricos.

En radiometría se le conoce como ruido al componente que surge de fluctuaciones aleatorias en el circuito detector, en la radiación entrante, o proveniente de otra fuente; la relación señal/ruido (SNR) es la relación de la corriente de señal a la corriente de ruido rms [39]. Otro concepto importante es la linealidad, la cual es la proporcionalidad de la salida y la entrada para un detector o un instrumento, lo cual implica que la responsividad de un detector o instrumento es constante sobre un rango definido de irradiancia o potencia de entrada.

Una figura de mérito es una cantidad usada para caracterizar el desempeño de un dispositivo, sistema o método con respecto a sus alternativas, algunas de las más comunes son:

- *Eficiencia cuántica sensible*: Es el número de salidas independientes de eventos por fotón incidente, generalmente se expresa como electrones por fotón y se encuentra con valores entre 0 y 1.
- *Eficiencia cuántica detective*: Es la tasa del cuadrado de la salida de la relación señal/ruido (SNR) con el cuadrado de la entrada SNR, siendo una medida de la degradación causada por el detector con un valor entre 0 y 1.
- *Potencia equivalente al ruido*: Es la señal entrante en Watts que produce una tasa de señal al ruido de 1. Es muy útil cuando la limitación de ruido es inherente en el detector y no es debido a la señal de entrada.
- *Equivalente a la potencia del ruido espectral*: Es la potencia equivalente al ruido para una longitud de onda sencilla monocromática de la señal de entrada.
- *Flujo de fotones equivalente al ruido*: Es la señal de entrada en fotones que produce una tasa de señal de ruido de unidad. Es la tasa del ruido de corriente rms a la sensibilidad de fotones.
- *Detectividad D*: Es el recíproco de la potencia equivalente al ruido.
- *Detectividad específica D\**: Muchos detectores muestran un ruido que es calculado como  $D\sqrt{A_d B}$  en donde  $A_d$  es el detector de área y  $B$  se refiere al ruido del ancho de banda; con esto se permite realizar una comparativa útil de los detectores de diferentes materiales, áreas desiguales y diferentes ruidos de ancho de banda, teniendo unidades de  $\frac{cm \cdot Hz^{1/2}}{W}$ .
- *BLIP*: Conocido así por sus siglas en inglés “Background-limited infrared photodetector” o el límite de fondo de un fotodetector infrarrojo. Indica la condición en la cual el ruido limitante en un detector surge la salida de fotones de fondo. Un detector de ruido interno BLIP ha sido reducido al punto en donde no es significativo, permitiendo que sea calculado la relación señal/ruido de forma fácil.
- *NETD*: Conocido así por sus siglas en inglés “Noise Equivalent Temperature Difference” o ruido equivalente a la diferencia de temperatura. Es la cantidad de diferencia de temperatura entre el objetivo y el fondo que produce una señal objetivo equivalente en potencia a la señal que es producida por el ruido en el sistema. Si la diferencia de temperatura es menor a esta cantidad, el detector no será capaz de discernir la señal objetivo de la señal de ruido. Es una medida de qué tan bien un detector de imagen térmica es capaz de distinguir entre pequeñas diferencias en radiación térmica en la imagen por lo que se le conoce como sensibilidad térmica, entre menor sea el valor de NETD será más sensible y puede tener un mejor desempeño. Generalmente es

expresado en miliKelvin, aunque es conocido en algunas ocasiones como un contraste térmico. Cuando el ruido es equivalente a la menor diferencia de temperatura medible, el detector ha alcanzado el límite de su habilidad para adquirir una señal térmica útil; entre más ruido exista, mayor será el valor NETD del detector. Las cámaras térmicas detectoras de microbolómetros son del orden de los 45mK, el valor del ruido medido debe ser especificado a una temperatura particular del objeto ya que esto afecta la medición. Si existe un contraste térmico significativo en la imagen y la temperatura de interés sea mucho mayor que la temperatura del fondo, no afectará tanto a la exactitud de medición.

Un bolómetro es un dispositivo cuya conductividad cambia con la temperatura [36], es una resistencia que posee un alto coeficiente de temperatura de resistencia y mide la cantidad total de radiación electromagnética proveniente de un objeto en todas las longitudes de onda, se representa con el símbolo  $\beta$  y su valor se calcula como  $\beta = \frac{1}{R} \frac{dR}{dT}$  con unidades de  $K^{-1}$ . Pueden ser fabricados por metales o semiconductores, la Ecuación 3.6 muestra la resistencia como una función de la temperatura, en donde  $R_0$  representa la resistencia a un cierto valor de temperatura [39]. Por su parte, un microbolómetro es un tipo específico de bolómetro utilizado como detector en cámaras térmicas y que permite detectar radiación infrarroja, generalmente son utilizadas en cámaras de bajo costo que no necesitan refrigeración.

$$R(T) = R_0(1 + \beta\Delta T) \quad (3.6)$$

La termografía infrarroja es un método que puede revelar anomalías en el cuerpo humano a partir de la detección de un cambio de temperatura de manera rápida, pasiva, sin contacto y de forma no invasiva. En esta técnica se utiliza una cámara termográfica en la cual se realiza una medición de la radiación emitida por una superficie para poder determinar la temperatura de un cuerpo, cuanto mayor sea la radiación, mayor será la temperatura [3, 8, 42].

La cámara opera utilizando un lente el cual enfoca la energía infrarroja en un sensor y proyecta la radiación infrarroja a través de microbolómetros en donde en función de la resistencia eléctrica de estos se obtendrá su temperatura al ser proporcional a la energía de la radiación infrarroja incidente tras convertir la parte infrarroja del espectro electromagnético proveniente del objeto en una imagen visible a través del detector [40]; un esquema representativo de la forma en la cual funciona una cámara infrarroja se puede apreciar en la Figura 3.2 [43], el píxel se obtura de forma electrónica abriendo o cortando el capacitor de integración; estas mediciones se muestran como una distribución bidimensional que correlaciona la temperatura con los niveles de la fuente. La resolución de una cámara infrarroja es baja en comparación con las de luz visible debido a que se necesitan detectores térmicos para sensar la energía que tienen longitudes de onda mucho mayores que las de luz visible, requiriendo que cada elemento del sensor sea significativamente más grande [36].

La termografía infrarroja es la transformación de una imagen infrarroja en una radiométrica que permite que los valores de temperatura puedan ser leídos en una imagen con lo cual cada píxel en la imagen radiométrica representa una medición de temperatura. Otro punto importante a considerar en las cámaras es el campo de visión, que representa el área máxima de una muestra que una cámara puede tomar, así como la longitud focal que describe la distancia entre la lente y la imagen enfocada en el sensor, la cual está determinada por la fuerza con la que converge la luz para enfocar el objeto que influye también el ángulo de la luz horizontal que pueda ser capturada por el lente, conocida como el campo de visión angular, que es el ángulo entre la luz capturada en la horizontal y la luz capturada en el borde [39].

En algunas ocasiones es complicado el poder registrar la temperatura con las cámaras infrarrojas debido a que se tienen objetivos con baja emisividad, por lo cual se requiere que ésta se incremente en los puntos a medir. La emisividad se refiere a la eficiencia con la cual un objeto emite radiación

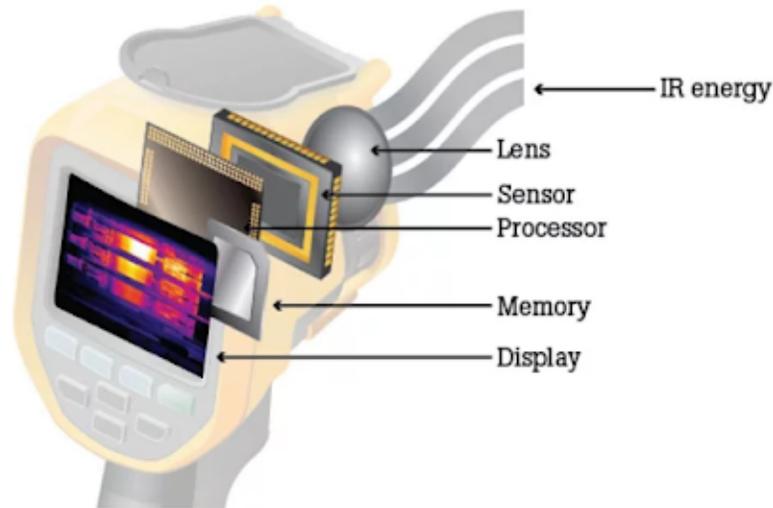


Figura 3.2: Esquema de funcionamiento de una cámara infrarroja [43]

infrarroja comparado con un emisor perfecto llamado cuerpo negro, el cual tiene una emisividad de 1 siendo un objeto ideal o a cual únicamente nos podemos aproximar en la naturaleza, estos son excelentes fuentes de radiación térmica debido a que el espectro emitido es directamente proporcional a su temperatura [36] y en las cámaras es importante poder ajustarlas a un valor correcto de emisividad para poder tener mediciones correctas de temperatura [43]. En tanto sea mayor su temperatura, será mayor la cantidad de radiación emitida.

Cuando no se tiene una emisividad correcta, o se quiere estar completamente seguro de que la temperatura que se quiera registrar sea la correcta es que puede pegarse cinta de aislar en el material, ya que esta tiene una emisividad de 0.95, siempre y cuando la cinta sea de alta calidad y que no sea de vinil, ya que puede no ser la mejor opción para realizar la medición. Esto se recomienda hacer principalmente en metales y en mediciones eléctricas, o cuando no se realice el ajuste en la emisividad en la cámara de acuerdo a los materiales que se están midiendo; normalmente las cámaras ya tienen cargadas una base de datos de la emisividad de diferentes materiales, además de que existen tablas que permiten conocer la emisividad de los materiales que se están midiendo.

La emisividad de un cuerpo puede afectar la temperatura medida por la cámara infrarroja, siendo dependiente de la longitud de onda, en el caso de la piel humana la emisividad varía por muy poco dependiendo del fototipo de piel que se tenga, ya que cuando se trata de piel blanca o de piel quemada es de  $0.97 \pm 0.02$ , mientras que para la piel negra es de  $0.98 \pm 0.01$  [3, 42, 44].

### 3.1. Termografía Infrarroja en Cáncer de Mama

Lawson y Chughtai fueron los primeros que utilizaron escáners infrarrojos en los que determinaron que la temperatura de superficie de la región que rodea a un tumor es de alrededor de  $2^{\circ}\text{C}$  mayor a la superficie de temperatura de la misma región colateral de la mama sano [3, 45].

Sin duda el estudio de Gautherie [46] es el más célebre en esta rama, al haber realizado un estudio a cerca de 58,000 pacientes en casi 12 años en el cual concluye que la termografía es una técnica que puede utilizarse para un tamizaje preliminar de mujeres asintomáticas para poner atención en aquellas

que deben de ser examinadas de manera más cuidadosa de tener cáncer de mama, así como resaltar la importancia de tener habilidad técnica y experiencia en la interpretación de las imágenes térmicas.

El uso de la termografía fue aprobada en 1982 por la FDA como una **herramienta adjunta** a la mamografía para el tamizaje en la detección de cáncer de mama que permite incrementar la sensibilidad hasta en un 10 % al utilizarse en combinación con la mamografía ya que no estas técnicas no siempre llegan a detectar la misma lesión por lo que se pudieron encontrarse cánceres que se presentaron antes de que las células adquirieran la habilidad de invadir el estroma circundante y que existiera evidencia morfológica de carcinoma *in situ*, inclusive entre 8 y 10 años previo a que puedan detectarse en la mamografía, además de ser útil cuando se tienen senos densos [8, 21, 45].

Los avances que se han tenido en las cámaras infrarrojas a partir del nuevo milenio, así como en la computación y simulación, condujeron a los investigadores a usar y explorar de nuevo esta técnica, esto debido a que las cámaras infrarrojas modernas pueden lograr sensibilidades por debajo de 20mK, lo que ha permitido la captura de termogramas con la capacidad de detectar pequeñas variaciones de temperatura.

Diferentes estudios han reportado la **sensibilidad** de la termografía infrarroja, que varía entre **71 %** y **94 %**, así como de la **especificidad** que oscila entre **14 %** y **85 %** [47]. La principal diferencia entre la termografía y técnicas como la mamografía, ultrasonido o la imagen por resonancia magnética, es que no busca diferencias en las características del tumor con respecto al tejido normal, sino que se enfoca en el calor producido debido a un incremento en la circulación de sangre como resultado de la neoangiogénesis y cambios metabólicos asociados con la generación y crecimiento del tumor [48].

Existe una relación entre la sensibilidad y la especificidad para la termografía infrarroja, coincidiendo en que al reportar una alta sensibilidad se tiene un umbral bajo para la clasificación de termogramas como anormales por lo que presentan una baja especificidad, mientras que al reportar una especificidad alta se han reportado bajas sensibilidades [3, 49-52]. Una de las limitaciones de este método ha sido la falta de estandarización tanto en los procedimientos de adquisición, que algunos no siguen un protocolo, así como en la evaluación de estos.

A causa de la multiplicación descontrolada de las células se genera una tasa metabólica de calor y una tasa de perfusión de la sangre que son mayores a los del tejido normal, disipándose así a los alrededores del tejido y manifestándose como picos de temperatura en la superficie de la mama, la cual en la proximidad del tejido afectado es de cerca de 2.5°C mayor a la del tejido en la región contralateral que no está afectada, con los que se pueden apreciar anomalías térmicas a través de la cámara termográfica en la que se mapea la temperatura de la mama y así una anomalía térmica que puede manifestar la presencia de un tumor, inflamación o infección, siendo éste un procedimiento no invasivo, indoloro y económico con lo que pueden ser detectados problemas potenciales en la fisiología del paciente sin importar la edad de esta a diferencia de la mamografía [3, 8, 9, 44, 45].

La distribución de la superficie de temperatura de la mama tumoral está determinada por el *tamaño, posición, profundidad y etapa del tumor*. Es más fácil identificar un tumor cuando es superficial; para tumores profundos, la superficie de temperatura puede tener variaciones muy sutiles con respecto a una mama sana, sin embargo, se ha demostrado que los tumores ubicados por debajo de los 38mm de profundidad con respecto a la superficie no pueden ser fácilmente detectados con una cámara infrarroja [3], por lo que la efectividad del uso de imágenes infrarrojas para detectar tumores de mama dependen de la capacidad para resolver la diferencia en temperaturas en la superficie de la piel generada por el tumor [42].

La magnitud de los gradientes de temperatura es usado para distinguir entre tumores benignos y malignos; se considera maligno cuando el gradiente es mayor a 1°C, mientras que cuando sea menor es asociado con enfermedades benignas de la mama [3, 8, 20], en su estudio, Kerserlingk *et al* [45] reporta que se tiene un incremento de hasta 2°C al presentarse un cáncer.

La termografía puede detectar la **angiogénesis** (desarrollo de nuevos vasos sanguíneos) debido al incremento en la demanda de sangre para abastecer los nuevos vasos sanguíneos y al incremento en la actividad metabólica ya que las mujeres que tienen un incremento en el número de microvasos en los senos (angiogénesis) tienen 7 veces mayor riesgo de desarrollar cáncer de mama en comparación con mujeres con densidad normal de microvasos [3, 45].

La tasa metabólica se traduce en el aumento de temperatura detectado; el calor metabólico generado por un tumor está asociado al tiempo de duplicación del volumen del tumor que potencialmente puede dar una indicación de su malignidad, por lo que uno de los aportes más importantes de la imagenología infrarroja digital de cáncer de mama es la posibilidad de **estimar la producción metabólica de calor de un tumor** de una manera no invasiva [20, 42].

Para la evaluación de los termogramas se sigue el sugerido por González [42] a partir de un puntaje térmico mediante la adición de la cantidad de la vascularidad a la diferencia de temperatura en la superficie en grados Celsius en el sitio de la lesión comparado con la mama contralateral, coincidiendo con lo planetado en la escala de Ville Marie. En este trabajo [42] se da un puntaje en función del nivel de vascularidad que se tenga:

1. Ausencia de patrones vasculares
2. Patrones vasculares simétricos o moderados
3. Asimetría vascular significativa
4. Asimetría vascular extendida en al menos  $\frac{1}{3}$  del área de la mama

Posterior a esto, se elige el punto de mayor temperatura para compararlo con el punto simétrico de su parte contralateral, generalmente esto puede realizarse con el software FLIR Tools, en donde se selecciona el botón de  $\Delta T$  y a esta diferencia de temperatura se le suma el puntaje que se le había dado con la vascularidad; en caso de tener más de 2.5 puntos en el score total se dice que está en peligro y se recomienda que se debe de realizar otro estudio, ya sea mamografía o biopsia.

Estudios como el realizado por Ng *et al* [20] sugieren que la imagen de cada mama sea seccionado en cuatro áreas previo a la comparación por simetría. Antes de realizar este tipo de análisis, es importante que se conozca la temperatura normal de los individuos estando sanos ya que sino puede ser interpretado como un signo de enfermedad [44].

## 3.2. Protocolo De Adquisición de Termogramas

Para poder hacer la adquisición de los termogramas existen protocolos que se deben de respetar, los cuales incluyen varios aspectos y de los cuales depende que la esta técnica pueda ser aplicada de buena manera; el protocolo más célebre es el de Glamorgan *et al* [53], aunque otros estudios como Kandlikar *et al* [3] y González [44] también lo mencionan, siendo tomada de estas fuentes la información presentada en esta sección.

El paciente debe de evitar al menos dos horas antes el ingerir comidas grandes o bebidas calientes como el té o café, así como el fumar o consumir alcohol. En el día de la adquisición, debe de evitar aplicarse cosméticos, lociones o antitranspirantes, así como los baños de sol al menos 5 días antes de la evaluación. Debido a los cambios sufridos por la paciente al estar menstruando o durante el embarazo en el tejido mamario [1] debe evitarse realizar el estudio en estas etapas.

Se debe de contar con una sala de examinación en la que el paciente pueda tener un reposo ahí durante 15 minutos con ropa suelta, poca ropa o sin esta para poder irse aclimatando al medio ambiente, la cual debe de estar entre los 18 y los 25°C con humedad relativa entre 40 y 75%. Se recomienda que en el lugar de adquisición de los termogramas se tenga un control de la luz exterior, así como mantener las puertas y ventanas cerradas.

Las cámaras de imagen térmica usadas para obtener los termogramas tienen un tiempo de inicio antes de que las imágenes sean estables; se recomienda que esté encendida al menos 15 minutos antes de iniciar las evaluaciones, la posición de la cámara debe ser claramente indicada si las imágenes son de diversos ángulos y el paciente no debe de moverse mientras que estas se obtienen.

Recordando que la **emisividad** es una propiedad que indica la eficacia con que una superficie emite una radiación térmica, será un parámetro importante al momento de adquirir las imágenes, por lo que las cámaras termográficas tienen ajustes tanto para la emisividad y la reflectancia; para esto, puede adherirse a una superficie una banda de cinta aislante negra con lo cual la emisividad estará ajustada cuando la temperatura de la superficie sin cinta sea igual a la de la superficie con cinta.

El estudio de Garduño *et al* [54] indica que se deben de tomar tres puntos de referencia en el termograma para identificar la estabilización térmica, que serán la axila, la mama y el abdomen. Una vez adquiridas, las imágenes deben de ser estandarizadas al momento de identificar regiones sospechosas detectando puntos calientes o asimetría entre los senos.

### 3.3. Análisis Manual de Termogramas

Para poder realizar el análisis de los termogramas de forma manual deberán de ser cargados los archivos en el software *FLIR Tools*. En base a la experiencia del análisis realizado, puede ser más sencillo trabajar con termogramas que están en blanco y negro sin tener la influencia de la paleta de colores aplicada a los termogramas. Lo anterior debido a que de esta forma pueden distinguirse de una mejor forma las regiones que se consideren con una mayor vascularización; habrá que considerar a su vez los límites de temperatura que se le apliquen a la imagen y con lo cual puede variar la apreciación del observador, pero puede recomendarse que se pongan en un rango de los 25°C a los 40°C para segmentar de mejor forma y descartar errores.

Según lo mencionado en la sección 3.2 coinciden en que las imágenes adquiridas deben de tener los senos de la paciente de manera simétrica y perpendicular sin importar su complejión; el estudio [44] indica que a cada termograma se le debe de dar primero un cierto puntaje en función a la vascularización que presente como:

1. Vascularización muy uniforme (regular) en toda el área
2. Baja vascularización; es el puntaje otorgado más usual
3. Moderado
4. Muy exagerado; rara vez se da este puntaje

Una vez otorgado este puntaje debido a su vascularización, ahora se procede a analizar la imagen en el programa *FLIR Tools* de los puntos que tengan ya sea un punto con temperatura más alta o más baja en el área del busto y se compara con la temperatura que se tiene en espejo.

A continuación se selecciona el botón  $\Delta T$  para saber la diferencia de temperatura que se tiene entre el punto con temperatura más alta o baja de los senos con respecto al punto simétrico contralateral; esta diferencia se suma al puntaje dado en función de la vascularización para obtener así el

**puntaje térmico.** En caso de que este valor sea mayor a 2.5 se clasifica como un termograma anómalo con lo cual el médico decide el estudio que se le deba aplicar, ya sea mamografía o biopsia en la mayoría de los casos.

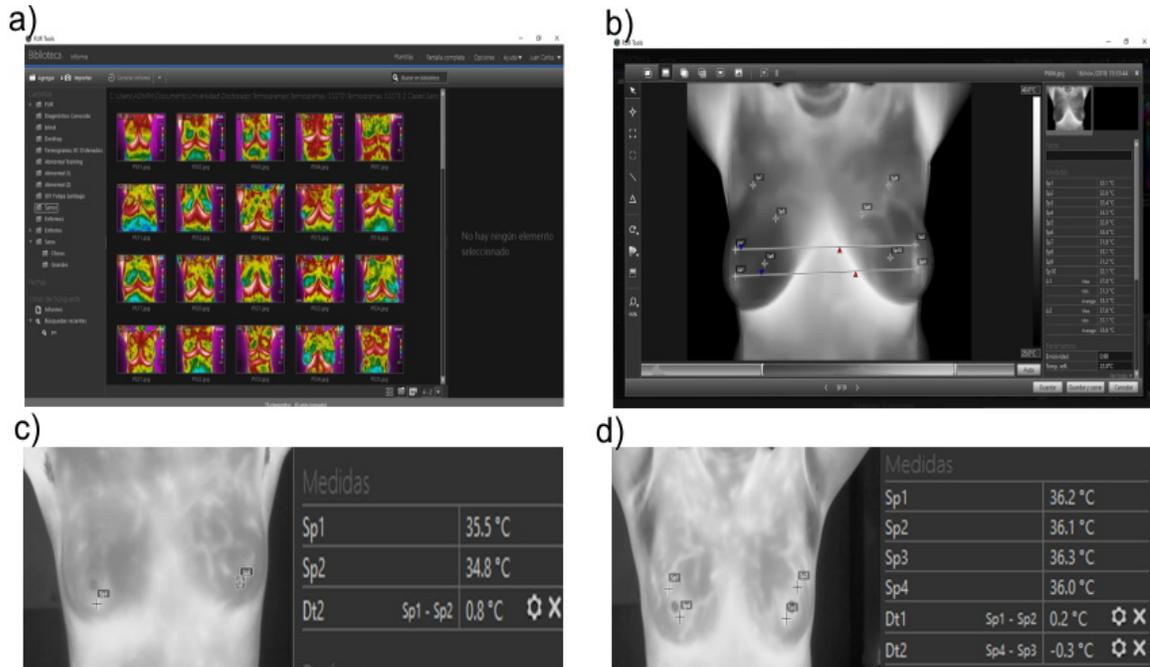


Figura 3.3: Capturas de pantalla de *Flir Tools*. Mediante este software se puede realizar el análisis de termogramas de forma manual

Para realizar estas mediciones en el software, se debe de buscar el botón “agregar herramienta de medición de cuadros” para que iniciar un análisis de esta forma y tras seleccionar un área, poder conocer cuál es el punto con mayor o menor temperatura de la mama para posteriormente dando clic en “agregar medida de punto” se obtenga el valor de la temperatura.

Se vuelve a ejecutar el procedimiento para la otra mama y es comparado con la parte en espejo de mismo busto analizando si su temperatura cambia; el mismo programa tiene también una herramienta para trazar una línea recta que facilita el ir al punto contralateral. En base a los puntos obtenidos, se eligen estos puntos que se tienen se selecciona el botón “agregar una función delta” y se compara la temperatura de manera automática en la parte derecha.

En la Figura 3.3 se puede apreciar la forma en la cual se evalúan los termogramas de forma manual apoyándose del programa *Flir Tools*. En el panel a) se muestra la pantalla principal del programa al iniciar, el panel b) indica la forma en la que aparece cuando una imagen es seleccionada para realizar el análisis en donde se permite hacer la comparativa entre los puntos simétricos contralaterales del termograma.

El panel c) muestra a una paciente que tuvo un puntaje de vascularización de 1, y al momento de realizar la comparativa de la temperatura de dos puntos, la diferencia entre los puntos contralaterales es de 0.8, por lo que es considerado en el grupo control. Por el contrario, en el panel d) se muestra una paciente que tiene 3 puntos al analizar la vascularización, mientras que la mayor diferencia de temperatura es de 3.3, teniendo la clasificación de anómalo.

Al realizar este tipo de análisis es común que el criterio para poder otorgar un cierto puntaje de vascularización que corresponderá al termograma sea diferente según la persona que lo analice, así como

tener dificultades al momento de realizar la búsqueda de los puntos con mayor o menor temperatura en la mama y el poder acertar al hacer que coincida con el punto contralateral, acentuándose estas diferencias cuando quien realice este protocolo tenga poca experiencia en la evaluación de las imágenes.

Con lo anterior, se puede identificar una **ausencia de una estandarización** para la valoración de las imágenes. Es por ello que el contar con un **método** que proporcione **uniformidad en el juicio** de los mismos ayudará a que la técnica pueda detectar de manera automática a las pacientes que tienen una anomalía térmica en los senos sin importar la experiencia de la persona que evalúe las imágenes o la carga de trabajo que tenga.

## Capítulo 4

# Aprendizaje Automático

En el campo de la inteligencia artificial, los algoritmos de aprendizaje se utilizan para realizar labores rutinarias, así como en labores de investigación, transcribir textos, identificación de objetos en imágenes, mostrar resultados relevantes en una búsqueda y reconocimiento de patrones [55, 56], siendo estos los usos principales del *aprendizaje automático*.

Estos algoritmos tratan de seguir el aprendizaje humano para realizar una tarea similar a lo que haría el cerebro, sin embargo, hay algunas cosas que quedan sin entender al 100 % sobre el por qué un algoritmo aprende de esa forma, tal y como pasa con las personas. El **reto actual** de la inteligencia artificial es que se puedan realizar no solamente tareas que son complicadas para el humano y que pueden realizarse siguiendo fórmulas matemáticas, sino que puedan realizar tareas que son sencillas para las personas de desempeñar pero difíciles de describir de una manera formal [57].

El aprendizaje automático o de máquina, es más conocido por su nombre en inglés como *machine learning*, tomando como base lo mencionado por Arthur Samuel [56, 58] como un algoritmo en el cual se tenga un aprendizaje de experiencias con respecto a una cierta función en donde, sin necesidad de que se esté programando la computadora, ésta tenga la habilidad para aprender a realizar una cierta actividad actualizando sus estimaciones de la solución mediante procesos iterativos.

*Machine learning es una forma de estadística aplicada* con un énfasis en el uso de computadoras para realizar un estimado estadístico de funciones complicadas y un énfasis menor en el proveer intervalos de confianza alrededor de estas funciones [57]. En el *machine learning* se utilizan operaciones que incluyen la optimización y la solución de sistema lineales de ecuaciones las cuales se utilizan para alcanzar la habilidad de desempeñar una tarea, lo cual recibe comúnmente el nombre de aprender; las tareas son descritas en términos de cómo el sistema de aprendizaje de máquina procesa un ejemplo, lo cual se refiere a un conjunto de características que han sido medidas cuantitativamente de algún objeto o evento que se quiere procesar en el sistema de aprendizaje de máquina [57].

Cabe destacar que un algoritmo será tan bueno como sean los datos con los cuales se entrenan, ya que si la adquisición de datos es deficiente, por muy bueno que sea el algoritmo de aprendizaje automático, este no podrá tener una buena generalización para cuando se ingresen más datos; por ello es que cobra importancia tener la mayor cantidad de datos posibles para que el algoritmo funcione mejor [56, 57, 59].

A su vez, al estar limitada en su habilidad para procesar datos que están en su forma natural, es decir, sin preprocesamiento, el construir un reconocedor de patrones requiere un buen dominio al extraer las características de los datos para poder detectar o clasificar patrones en una entrada [55]. Una de las razones por las cuales estas técnicas pueden ser utilizadas ahora, es porque además de

una mayor cantidad de datos que se pueden tener, se cuenta con computadoras que tienen una mayor capacidad para realizar las operaciones.

Hay tres formas en las cuales se puede realizar el aprendizaje, que son aquellos en los cuales los datos cuentan con una etiqueta en función a la clase a la cual pertenecen, llamándose **aprendizaje supervisado**, destacando entre ellos la regresión lineal, análisis de discriminantes, máquina de vectores de soporte (mejor conocido por su nombre en inglés como *support vector machine*), bosques aleatorios (reconocido como *random forest* o árboles de decisión), regresión logística y las redes neuronales como algunos de los más usados [59].

Por otra parte, el enfoque en el cual se buscan asociaciones que tienen estos sin estar diciéndole al algoritmo a cual clase es a la que pertenecen recibe el nombre de **aprendizaje no supervisado** [56, 60], siendo algunas de las más reconocidas análisis de componentes principales (PCA), descomposición en un solo valor (*Single Value Decomposition* SVD), análisis de componentes independientes (ICA), agrupación jerárquica y media-K de agrupación (*k-means clustering*).

Uno más es el **aprendizaje auto supervisado** los cuales aprenden características visuales de datos que no tienen una etiqueta dada [61], los cuales se entrenan mediante funciones de aprendizaje objetivo. Una de las aplicaciones más populares de este tipo de aprendizaje es en la predicción de textos como los utilizados en el celular o al completar frases en buscadores o correos.

Dada la naturaleza del trabajo realizado, nos centraremos en la técnica de aprendizaje supervisado, el cual es el más común, ya que recoge una gran cantidad de datos etiquetando imágenes para cada categoría que se requiera clasificar, por lo que es de suma importancia que estos se encuentren ordenados. En caso de tener valores faltantes, se recomiendan las técnicas descritas por Kotsiantis *et al* [62], destacando las de ignorar esos elementos, sustituirlos con un valor medio o aplicar métodos de regresión o clasificación.

La elección del clasificador que se utilizará estará en función de varios puntos, tales como el poder computacional que se tiene, la cantidad de datos disponibles para este fin, así como la exactitud de predicción del clasificador [56, 59, 62].

Existen diversos enfoques para poder separar los datos, el más tradicional es el de separarlo en 70 % para entrenamiento y 30 % para validación, otro indica que se formen varios conjuntos de datos para entrenar el algoritmo, utilizando validación cruzada y dejando uno fuera [62], mientras que uno más indica que debido a la disponibilidad de datos se tenga 70 % para entrenamiento, 15 % para validación, en donde pueda hacerse la validación cruzada, y 15 % para prueba del algoritmo [59].

A largo plazo, se espera que el aprendizaje no supervisado vaya tomando una mayor relevancia, debido a que las personas normalmente aprenden de una manera subjetiva e intuitiva con una gran cantidad de conocimiento de manera informal y sin que necesariamente se tenga que realizar un etiquetado de los datos.

## 4.1. Matemática Aplicada Al Aprendizaje Automático

En el aprendizaje automático se requiere estar familiarizado con algunos conceptos de álgebra lineal, probabilidad y computación numérica. El objetivo de esta sección es poder explicar parte de la matemática involucrada en el *machine learning* que permita entender un poco más la forma en la cual trabaja la técnica. Mucho de lo aquí presentado tiene su base en algunos libros de álgebra lineal, cursos de aprendizaje de máquina y libros especializados en el tema.

En el álgebra lineal se habla comúnmente de algunos objetos matemáticos tales como los valores escalares, vectores, matrices y tensores. Un **escalar** es un número cualquiera, mientras que un **vector**

se denomina como un arreglo de números que están organizados en un cierto orden y que pueden ser identificados por un índice [63] tal como se puede apreciar en la Ecuación 4.1 donde se indica que el elemento está solamente en una dimensión.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (4.1)$$

Ecuación 4.1 Representación de un vector

Por su parte, una **matriz** es cualquier arreglo de dos dimensiones de números  $m$  que son los renglones y  $n$  correspondiente a las columnas, con los cuales se podrá identificar cada elemento [63] como se aprecia en la Ecuación 4.2

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,n} \end{bmatrix} \quad (4.2)$$

Ecuación 4.2 Representación de una matriz

Por último, tendremos los **tensores**, los cuales representan arreglos de números que están en  $n$ -dimensiones; de forma general, cuando los números están arreglados en una rejilla regular con un número variable de ejes, estos son conocidos como tensores [63], los cuales podrán ser identificados en las coordenadas  $\mathbf{A}_{i,j,k}$ .

En el caso de los vectores es común que en el aprendizaje de máquina o automático se deba de medir el tamaño del vector utilizando la función de **normalidad** mostrada en la Ecuación 4.3, en donde se hace una función de mapeo de los vectores a valores no negativos midiendo la distancia del origen a un punto cualquiera.

$$\|x\|_p = \left( \sum_i |x_i|^p \right)^{1/p} \quad (4.3)$$

Ecuación 4.3 Norma  $L^p$

La norma  $L^2$  es conocida como la *norma Euclidiana* en donde se mide la distancia del origen a un punto indicado y que generalmente se le denomina  $\|x\|$ . Otra norma que también es aplicada es la norma  $L^1$  debido a que puede discriminar entre elementos que son exactamente cero y elementos que son pequeños pero no son cero, donde se tiene un valor de  $p = 1$ .

Por último, otra norma que se acostumbra a utilizar para los vectores en *machine learning* es cuando  $L^\infty$ , conocida como la norma máxima, la cual simplifica el valor absoluto del elemento con la magnitud más grande en el vector, tal como se expresa en la Ecuación 4.4

$$\|x\|_\infty = \max_i |x_i| \quad (4.4)$$

Ecuación 4.4 Norma  $L^\infty$

*Machine learning* se basa en la probabilidad para saber si el algoritmo está haciendo una predicción correcta, así como para analizar su comportamiento. Cuando se hable de qué tan posible es que ocurra algo tras una repetición de eventos, se le conoce como *probabilidad frecuentista*, mientras que cuando se refiere a los niveles de certeza, se le conoce como **probabilidad Bayesiana** [57].

Otros conceptos que se manejan son la expectativa, varianza y covarianza; el primero se refiere al valor promedio que una función toma cuando un valor es extraído de una distribución de probabilidad. La varianza da una medición de qué tanto los valores aleatorios de una función varían mientras se toman muestras de diferentes valores de la variable con su distribución de probabilidad; al aplicar la raíz cuadrada al valor se obtiene la desviación estándar. Por último, la covarianza da un sentido de qué tanto dos valores lineales están relacionados entre sí, entre más altos sean los valores absolutos, significa que los valores cambian demasiado y se encuentran alejados de sus medias [57, 63].

La teoría de la información es una rama de las matemáticas aplicadas que se utiliza en *machine learning* debido a que cuantifica cuánta información está presente en una señal [57]. Se define la autoinformación de un evento  $x=x$  para ser:

$$I(x) = -\ln P(x) \tag{4.5}$$

Los  $I(x)$  tienen unidades de nats que son la cantidad de información ganada por observar un evento de probabilidad  $1/e$ , aunque también pueden considerarse bits o shannons. Se puede cuantificar la cantidad de incertidumbre en una distribución entera de probabilidad usando la entropía Shanon. La entropía cruzada (*cross-entropy*) entre dos distribuciones de probabilidad en un mismo espacio mide la media de bits necesarios para identificar un evento de un conjunto de posibilidades, si un esquema de codificación está basado en una distribución de probabilidad dada  $q$ , más que en la verdadera distribución  $p$  [55, 57].

$$H(p, q) = E_p[-\log q] = H(p) + D_{KL}(p||q) \tag{4.6}$$

Los algoritmos de *machine learning* a menudo tienen distribuciones de probabilidad que involucran interacciones directas entre algunas variables, por lo cual se utilizan distribuciones de probabilidad en bastantes factores para poderlos multiplicar juntos [57].

Al trabajar con *machine learning* es necesario tener una función de optimización en la cual se esté minimizando la función  $f(x)$  que es la que altera el valor de  $x$ . La función que se va a minimizar tiene el nombre de **función objetivo, de pérdida o de costo**, y mide el error (o distancia) entre las puntuaciones de salida y los patrones deseados de puntuaciones para modificar así sus parámetros internos en los cuales se tienen características como valores de entrada que van de  $x_1$  hasta  $x_n$  [56].

La función de costo puede expresarse de diferentes formas, una de las más utilizadas es el desempeño del error medio entre el valor observado con el predecido al cuadrado (*mse* por sus siglas en inglés, Ecuación [4.7]), sin embargo, existen algunas otras funciones de desempeño tales como el error medio absoluto (*mae*, Ecuación [4.8]), suma del error absoluto (*sae*, Ecuación [4.9]), error de sumatoria al cuadrado (*sse*, Ecuación [4.10]) y entropía cruzada (Ecuación [4.11]).

$$J(w, b) = \frac{1}{2n} \sum_{i=1}^n |y(x) - a(x)|^2 \tag{4.7}$$

Ecuación [4.7] Función de costo error medio al cuadrado

$$J(w, b) = \frac{1}{n} \sum_{i=1}^n |y(x) - a(x)| \tag{4.8}$$

Ecuación [4.8] Función de costo error medio absoluto

$$J(w, b) = \sum_{i=1}^n |y(x) - a(x)|^2 \tag{4.9}$$

Ecuación [4.9] Función de costo suma de errores absolutos

$$J(w, b) = \sum_{i=1}^n (y(x) - a(x))^2 \quad (4.10)$$

Ecuación 4.10. Función de costo suma de errores al cuadrado

$$J(w, b) = -\frac{1}{n} \sum_{i=1}^n [y(x) \ln a(x) + (1 - y(x)) \ln (1 - a(x))] \quad (4.11)$$

Ecuación 4.11. Función de costo de entropía cruzada

La **entropía cruzada** es ampliamente empleada en problemas de clasificación y reconocimiento de patrones, ya que evita el problema de tener un lento aprendizaje a diferencia de las de costo cuadrático siempre que la neurona de salida sea una sigmoide, entre otras [56, 59].

La elección de la función de costo que se usará para ser minimizada en el algoritmo influirá en la predicción de la la media de salida para cada valor de entrada que se utilice, por lo cual la estadística será diferente. El perceptrón realiza el cálculo de la suma de los pesos tanto de entrada como de salida (Ecuación 4.18) y toma la decisión entre dos límites ajustables; si la suma es mayor a 1, entonces la salida es 1, sino, será 0 [62].

Las derivadas indican la pendiente de una función  $f(x)$  en el punto  $x$  especificando la forma en la cual los pequeños cambios en las entradas pueden obtener un cambio en las salidas [57]; es por esto que para ajustar apropiadamente los vectores de peso, los algoritmos de aprendizaje computan un vector gradiente que, para cada peso, indica por qué cantidad de error podrá incrementarse o decrecer si el peso fuera incrementado por una pequeña cantidad, el vector de peso es luego ajustado en dirección opuesta al vector gradiente [55], entre mayor sea el error, la neurona aprenderá más rápido, siendo esto lo que se espera de manera intuitiva [56]. De esta forma la derivada nos permitirá minimizar una función ya que dice cómo el cambio en  $x$  hace pequeñas mejoras en  $y$ .

Se le llama **mínimo global** al punto que obtiene el valor absoluto más bajo de  $f(x)$ , cuando hay varios mínimos globales se tiene que encontrar aquél que sea el óptimo global; en *deep learning* se optimizan funciones que pueden tener varios mínimos locales que no sean óptimos, lo cual es algo complicado, en especial cuando las entradas de la función son multidimensionales [57].

El **gradiente estocástico descendiente** consiste en mostrar los vectores de entrada para algunos ejemplos, computando las salidas y los errores, así como el promedio del gradiente para estos al reducir  $f(x)$  moviendo  $x$  en pequeños pasos en la dirección del gradiente negativo. De esta forma se repite el procedimiento para el ajuste de los pesos hasta que el promedio de la función objetivo detiene el decremento [55], controlando así la función de pérdida de la red.

$$\frac{\partial C}{\partial w_j} = \frac{1}{n} \sum_{j=1}^n x_j (\sigma(z) - y(x)) \quad (4.12)$$

Ecuación 4.12. Derivada parcial de la función de costo con respecto a los pesos

$$\frac{\partial C}{\partial b} = \frac{1}{n} \sum_x (\sigma(z) - y(x)) \quad (4.13)$$

Ecuación 4.13. Derivada parcial de la función de costo con respecto a los sesgos

$$\frac{\partial a(x)}{\partial w} = \begin{cases} 0 & \because \frac{\partial a(z)}{\partial w} = 0^T \quad \text{Cuando } a(z) \leq 0 \\ 1 & \because \frac{\partial a(z)}{\partial w} = x^T \quad \text{Cuando } a(z) > 0 \end{cases} \quad (4.14)$$

Ecuación 4.14 Vector de la regla de la cadena para los pesos

$$x' = x - \epsilon \nabla_x f(x) \quad (4.15)$$

Ecuación 4.15 Gradiente descendiente

El valor de  $\epsilon$  es conocido como la tasa de aprendizaje (que se menciona también posteriormente) el cual determina el tamaño del cambio que se debe de hacer en la dirección contraria al gradiente en la Ecuación 4.15; este convergerá cuando cada elemento del gradiente sea cero 57.

$$J_{i,j} = \frac{\partial f(x)_i}{\partial x_j} \quad (4.16)$$

Ecuación 4.16 Matriz Jacobiana

La **matriz Jacobiana** contiene todas las derivadas parciales de una función cuyas entradas y salidas son ambos vectores 57, esta queda descrita como se muestra en la Ecuación 4.16. También se puede estar interesado en obtener la segunda derivada, la cual indica cómo cambiará la primera derivada en relación a la variación de la entrada, por lo cual se puede pensar en ella como una medida de curvatura 64.

La **matriz Hessiana** es cuando la función tiene múltiples dimensiones de entrada, siendo el Jacobiano del gradiente tal como se muestra en la Ecuación 4.17 debido a que la matriz Hessiana es real y simétrica, esta puede descomponerse en un conjunto de eigenvalores reales y en bases ortogonales de eigenvectores. La segunda derivada puede darnos un máximo, mínimo o un punto de silla (o inflexión) 64.

$$H(f)(x)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad (4.17)$$

Ecuación 4.17 Matriz Hessiana

En *machine learning* generalmente se está interesado en medir tanto la **exactitud** de la proporción de ejemplos para los cuales el modelo produce una salida correcta, así como la medición de la **tasa de error** 57; estas métricas se emplean también en datos que no ha visto el algoritmo para determinar su desempeño en el mundo real; cuando el algoritmo logra tener un buen desempeño en datos que no ha visto, se dice que ha logrado una buena **generalización** 56.

La descripción de los datos puede hacerse mediante una matriz de diseño en donde cada columna corresponde a una característica diferente, por lo cual cada ejemplo puede describirse como un vector, y cada una de ellas debe ser del mismo tamaño, sin embargo esto no siempre es posible, por lo cual los datos pueden describirse como conjuntos de datos que contienen  $m$  elementos.

Al realizar un entrenamiento del conjunto de datos, normalmente se realiza un cómputo de la medición del error en el entrenamiento, el cual se debe de ir reduciendo, esto debe de realizarse también con el conjunto de prueba de datos de manera que se tenga una generalización del error en ambos conjuntos de datos 56, 59.

Los parámetros en los algoritmos se deben de ajustar para el conjunto de entrenamiento para reducir el error de entrenamiento y probarse en el conjunto de prueba; los **factores** que determinarán qué tan buen **desempeño** tendrá un algoritmo de aprendizaje de máquina serán su habilidad de **disminuir el error de entrenamiento** y **reducir** la brecha entre el **error de entrenamiento** y el **error de prueba**, siendo estos los retos centrales del *machine learning* al hacer un sobre ajuste (*overfitting*), el cual ocurre cuando la brecha entre el error de entrenamiento y el de error es muy grande y solamente se ajusta a un conjunto de datos en específico, o un ajuste bajo (*underfitting*), que es cuando el modelo no puede obtener un error de entrenamiento lo suficientemente bajo [56, 57, 59].

La regularización se refiere a cualquier modificación que se realiza a un algoritmo de aprendizaje que tiene como intención el reducir su error de generalización pero no su error de entrenamiento [57]; esta se aplica a la función de costo, existen diferentes funciones de regularización, por lo cual debe de elegirse aquella que sea más adecuada a la tarea particular que se quiere resolver.

Un **hiperparámetro** sirve para controlar el comportamiento de un algoritmo de aprendizaje, estos también pueden ser un factor que propicie el *overfitting* por lo cual deben de seleccionarse bien; se recomienda que se tenga un conjunto de validación que el algoritmo no haya observado en el entrenamiento, por esta razón del 100 % del conjunto de datos, tradicionalmente se divide en 70 % para entrenamiento, 15 % para validación y 15 % de prueba [56].

En ocasiones no se pueden tener cientos de ejemplos que integren a una base de datos, por lo que se puede implementar la técnica de *validación cruzada*. En esta técnica se repite el cómputo del entrenamiento y la prueba en diferentes subconjuntos o divisiones de la base de datos original para probar del primero al  $k$ -ésimo subconjunto de datos [56, 57, 59].

El **gradiente estocástico descendiente** es una extensión del algoritmo gradiente descendiente. Uno de los problemas es que en *machine learning* se requieren grandes conjuntos de datos para poder hacer una buena generalización, pero que estos son más caros en lo computacional [56, 57, 59]. La función de costo utilizada por un algoritmo de *machine learning* normalmente descompone en sumas sobre ejemplos de entrenamiento de funciones de pérdida, es por ello que el gradiente es una expectativa que puede ser estimada de forma aproximada utilizando un pequeño conjunto de muestras llamado **minibatch** o minilote  $m'$  de ejemplos del conjunto de entrenamiento el cual se selecciona de un pequeño número de ejemplos tal como se mostró en la Ecuación 4.15 [56, 57, 59].

El gradiente descendiente estocástico tiene varios usos importantes ya que es la forma principal de entrenar modelos lineales en conjuntos de datos grandes. El número de actualizaciones requerido para alcanzar la convergencia normalmente incrementa con el tamaño del conjunto de datos, sin embargo, el modelo eventualmente convergerá al mejor error de prueba posible antes de que el gradiente estocástico descendiente haya visto cada ejemplo en el conjunto de entrenamiento; incrementar el tamaño del *minibatch* no extenderá la cantidad de tiempo de entrenamiento necesitado para encontrar el mejor error de prueba posible del modelo [57].

## 4.2. Aprendizaje Profundo

Conocido también como *deep learning*, es un método de representaciones de aprendizaje de múltiples capas de procesamiento que realizan diferentes transformaciones en cada uno de los niveles que le permiten aprender representaciones de datos con múltiples niveles de abstracción para poder descubrir estructuras en grandes conjuntos de datos [55].

La mayoría de los algoritmos de *deep learning* están basados en un algoritmo de optimización llamado **gradiente estocástico descendiente**. Debido a que normalmente están representadas por

varias funciones diferentes juntas es que se les considera una red, mientras que la longitud de qué tantas capas son las que tendrá es por lo que se les consideran modelos profundos [57].

En estos modelos se tiene el objetivo de aproximar una función  $f^*$ ; si se habla de clasificadores,  $y = f^*(x)$  mapea una entrada  $x$  en una categoría  $y$  en donde la red define un mapeo como  $y = f(x, \theta)$  y aprende el valor de los parámetros  $\theta$  que resultan en la mejor aproximación de una función [56, 57]. El algoritmo de aprendizaje debe decidir cómo el uso de esas capas debe de producir la salida deseada para implementar de mejor forma la aproximación de  $f^*$  [57]. Dependiendo del tipo de alimentación se pueden definir redes neuronales, redes neuronales de alimentación hacia adelante, convolucionales, redes pre-entrenadas no supervisadas, redes de aprendizaje de refuerzo, redes neuronales recurrentes y redes neuronales recursivas [56].

El aprendizaje profundo es una forma especializada del aprendizaje automático, la principal diferencia es que el **aprendizaje automático** empieza con la extracción manual de las características (llamadas *features*) relevantes que incluya información de la representación del problema a resolver, en el caso de este trabajo una imagen, y se crea un modelo que categoriza los objetos de la imagen, mientras que el **aprendizaje profundo** extrae las características importantes directamente de los datos de entrada (imágenes en el caso de nuestro problema), y además de realizar un ‘aprendizaje completo’, se proporcionan datos sin procesar y una tarea que realizar como puede ser una clasificación, lo cual aprende a realizarlo de una manera automática, siendo el número de datos lo que va a determinar qué tanto puede llegar a mejorar, es decir, el aprendizaje profundo puede ser más útil si la cantidad de datos de entrenamiento disponibles se han incrementado [65].

Las **redes neuronales** son una técnica relativamente nueva, capaz de clasificar conjuntos linealmente separables de casos, sin embargo, es complicado concluir que se ha desarrollado la mejor red neuronal posible para una aplicación en particular debido a que no están limitadas por una relación matemática predefinida entre variables dependientes e independientes, siendo los predecesores del aprendizaje profundo moderno [20, 62, 65].

Son especialmente útiles en los casos donde no se tiene una separación lineal, ya que pueden modelar cualquier relación no lineal compleja, siendo así una alternativa a técnicas como la regresión logística en donde las variables normalmente son seleccionadas por inclusión a través de algunas técnicas de selección de variables escalonadas [20, 62, 66]. Normalmente están compuestas por un perceptrón multicapa, los cuales son funciones matemáticas que mapean un conjunto de valores de entrada a valores de salida formadas por un compuesto de varias funciones simples.

Uso de Redes Profundas	
Ventajas	Desventajas
♣ Menos entrenamiento estadístico formal	♣ Son una “caja negra” natural
♣ Habilidad de detectar implícitamente relaciones complejas no lineales entre variables dependientes e independientes	♣ Mayor carga computacional
♣ Habilidad para detectar todas las posibles interacciones entre variables predictoras	♣ Es propenso a sobreajuste
♣ Disponibilidad de múltiples algoritmos de entrenamiento	♣ Naturaleza empírica del modelo de desarrollo
	♣ Pueden ser difíciles de usar en campo

Tabla 4.1: Ventajas y desventajas de las redes profundas

Los algoritmos son modelados bajo la estructura del cerebro humano, en el cual por medio de una serie de ecuaciones matemáticas, trata de imitar la operación de procesos biológicos como el aprendizaje y la memoria con el objetivo de predecir un resultado basándose en los valores de algunas variables predictoras en donde se combinan diferentes arquitecturas para conseguir una variedad de capacidades computacionales [20, 66]; algunos pros y contras de las redes neuronales pueden encontrarse en la Tabla 4.1.

Algo importante al diseñar una red neuronal profunda es al elegir la **función de costo**, en la cual se utiliza la entropía cruzada entre los datos de entrenamiento y las predicciones del modelo; generalmente se combinará una de las funciones de costo primarias con un término de regularización, una de ellas es la caída de los pesos; se aplica igual a lo que se explicó para *machine learning* [57].

Las **redes neuronales convolucionales** son **herramientas esenciales** para el aprendizaje profundo, y especialmente convenientes para el reconocimiento de imágenes. Las redes toman valores de los píxeles de entrada de las figuras que multiplican a los pesos, que son los elementos que conectan a la siguiente capa que son los sesgos, los cuales se suman a la función que tienen de entrada con lo que se forma la **función de activación** mostrada en la Ecuación 4.18 y representada en la Figura 4.1 elegida la función de activación se utilizará para hacer el cómputo de los valores de las capas ocultas.

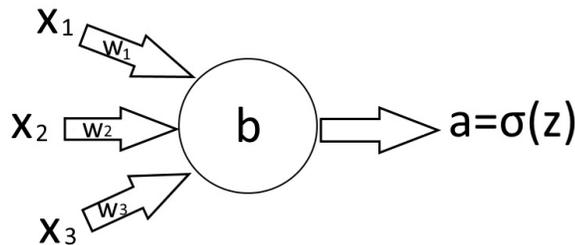


Figura 4.1: Cálculos de entrada y salida de una Neurona. Se presenta la forma en la que se tienen las entradas multiplicadas por los pesos y se añaden los sesgos.

En estos modelos cada neurona toma los valores de las neuronas conectadas multiplicadas por sus respectivos pesos que son sumados y aplican la función de activación [56]; es importante aclarar que la red normalmente tiene números aleatorios para los pesos y sesgos. Normalmente se les conoce como **modelos lineales** o combinaciones lineales siendo algunos de los más utilizados [57]. En estos se tiene un conjunto de vectores multiplicándose por un coeficiente escalar correspondiente y se suman los resultados, lo cual es realizado en la Ecuación 4.18, donde lo abarcado por un conjunto de vectores es el conjunto de todos los puntos obtenidos por una combinación lineal de los vectores originales [63].

$$a(x) = \sum_i^n w_i x_i + b = (w \cdot x) + b \quad (4.18)$$

Ecuación 4.18. Función de activación, en donde  $w_i$  es el vector de pesos,  $x_i$  el vector de entrada y  $b$  el sesgo (es un escalar)

Las redes son programadas para ajustar sus **pesos** internos basados en las relaciones matemáticas identificadas entre las entradas y salidas en un conjunto de datos de entrenamiento, es decir, ajustando estos pesos o **términos de error** mediante la diferencia entre las salidas objetivo y las neuronas de salida actuales para cada entrada  $x_i$ ; esto se logra corriéndolo de manera cíclica hasta que encuentre un vector de predicción que se desempeñe de forma correcta en todo el conjunto de entrenamiento al utilizar las variables predictivas y los resultados que están etiquetados al utilizar el gradiente para que disminuya la función de costo en una u otra forma a partir del algoritmo de entrenamiento [55, 56, 59, 62, 66].

Una vez realizado esto, se utiliza un conjunto de prueba que pueda medir su habilidad de generalización y producción de respuestas sensibles ante otros valores que no ha visto durante el entrenamiento para que puedan utilizarse en el reconocimiento de patrones o tareas de clasificación [55] [62] [66].

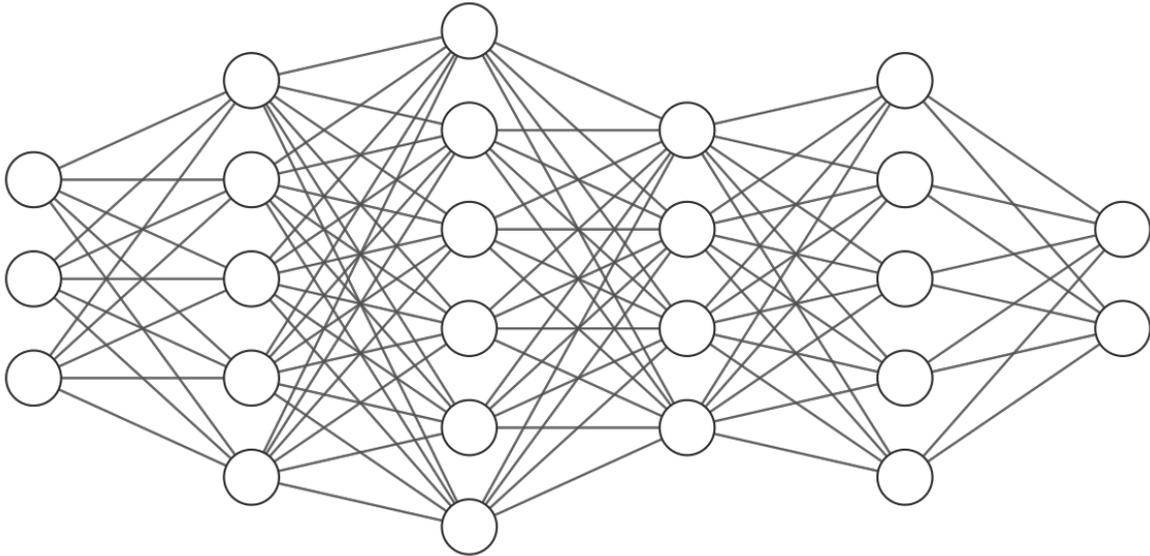


Figura 4.2: Representación de una Red Neuronal. Es posible apreciar la representación de una red neuronal, la cual está conformada por nodos y conectores de peso

La Figura 4.2 muestra cómo queda representada una red neuronal en función de los nodos (círculos) y conectores de peso (líneas), que pueden separarse en tres capas diferentes, la primera, que se encuentra a la izquierda, que es considerada la **capa de entrada**, la última, ubicada en la parte derecha, conocida como la **capa de salida**, su rol es la de proveer una transformación adicional a las características para completar la tarea que debe de realizar la red, dependiendo de la forma en la cual se represente es como se determinará la forma de la función de entropía cruzada; las capas que se encuentran entre la entrada y la salida, se les conoce como **capas ocultas** [66];

Es posible notar el hecho de que cada uno de los nodos de entrada se encuentra conectado a cada nodo oculto, pero que no hay interconexiones entre los elementos de una misma capa. La técnica es basada en perceptrones; en el entrenamiento la máquina muestra una imagen que produce una salida en forma de un vector de puntuación (**scores**) para cada una de las categorías en donde se quiere tener la mejor puntuación para cada una de ellas [55].

Se conectan mediante vectores de predicción llamados **pesos** (weights) desde  $w_1$  hasta  $w_n$ , siendo estos números reales con valores que oscilan entre  $-1$  y  $1$  [62] mediante un algoritmo de retropropagación (**backpropagation**), con el cual se modifican los parámetros internos de la función objetivo a partir de un gradiente con el que los pesos de las múltiples capas son modificados mediante la aplicación de la regla de la cadena (Ecuación 4.14) en las derivadas parciales con respecto a la entrada, yendo hacia atrás una vez que termina de realizar los cálculos en el orden natural de la red [55].

Debe tenerse claro el hecho de que *backpropagation* se refiere únicamente al método por el cual se computa el gradiente, mientras que otro algoritmo como el gradiente estocástico descendiente es utilizado para el desempeño del aprendizaje utilizando el gradiente [57].

Un ejemplo sencillo que ilustra la forma en la cual funcionan las redes neuronales se puede apreciar en la Figura 4.3, al mostrar una pequeña red con diferentes operaciones. La red tiene los

pesos ya establecidos, ilustrados en el panel a) con los colores dorado y azul por los cuales deben de multiplicarse los valores observados en la capa de entrada (3 y 2).

Para el nodo superior se multiplica  $3 \times 2$  y se suma al resultado de  $2 \times 5$ , con lo cual da el resultado de 13, se repite el procedimiento tanto para los pesos mostrados en color azul, como para la capa de salida arrojando un valor de 21; al realizar la resta del valor predicho menos el actual se obtiene el error. Se identifica que al modificar el peso que va de la capa intermedia a la salida por  $-2$  se obtendrá el valor deseado por lo que el error se convierte en cero, tal como se aprecia en el panel b).

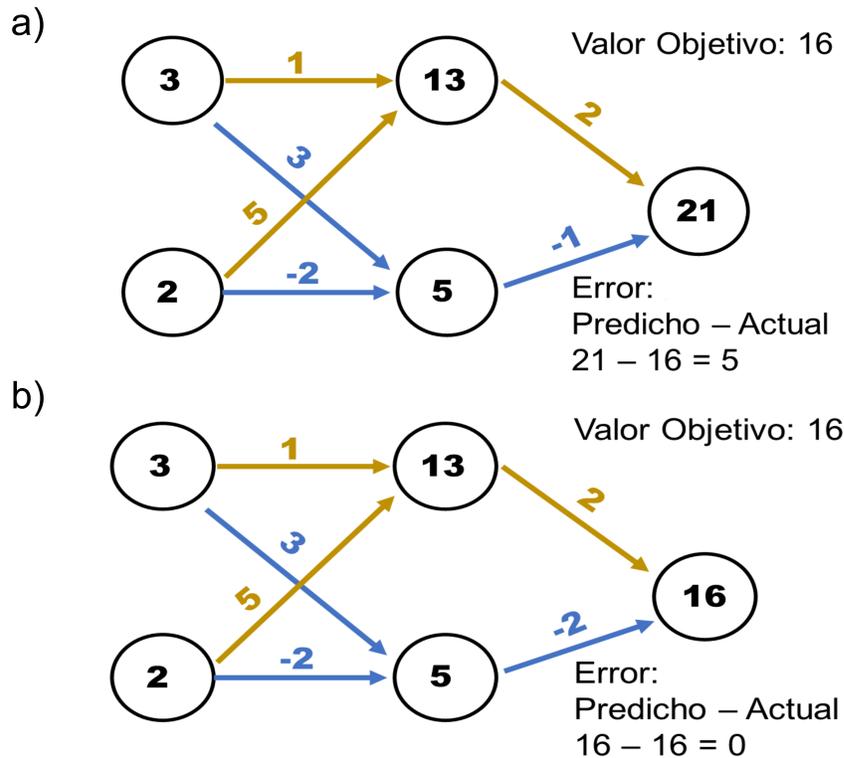


Figura 4.3: Representación Simple de Cálculos Realizados en una Red Neuronal. El panel a) muestra una red neuronal con los pesos dados originalmente, mientras que en el panel b) se ve la modificación aplicada a la red para obtener el valor objetivo.

Es de resaltar que la retropropagación se refiere solamente al método para computar el gradiente, mientras que otro algoritmo, como el gradiente estocástico descendiente, se utiliza para el desempeño del aprendizaje utilizando este gradiente [57]. De acuerdo con Caruana *et al* [67] las redes que cuentan con **retropopagación** son **insensibles al tamaño de la red** debido a que tienen una capacidad para aprender secuencias de modelos funcionales de manera similar a como aprenden las redes más chicas.

A su vez, el mismo estudio hace mención de la parada temprana (*early stopping*) para detener el entrenamiento en redes que sean muy grandes cuando han aprendido modelos similares a los aprendidos por redes pequeñas que tienen un tamaño óptimo. Esto significa una pequeña pérdida en el desempeño de la generalización de redes con una capacidad mayor si es utilizada, pero también previene a su vez el sobreajuste. Esto significa que se puede obtener un modelo con un mejor error en el conjunto de validación volviendo al ajuste del parámetro al punto en el cual el error de validación sea menor; cada vez que el error en el conjunto de validación se mejora se realiza una copia de esos parámetros, cuando el algoritmo de entrenamiento termina se regresa a esos parámetros en vez de a los últimos, con lo

cual, el algoritmo termina cuando ningún parámetro tiene un mejor récord sobre el error de validación para un número previamente especificado de iteraciones [57].

Una combinación lineal de peso de las entradas es sumada para determinar la entrada neta a cada nodo o capa y luego este resultado se pasa a través de una función de activación que es no lineal, la más popular es la unidad lineal rectificadora (**ReLU**) que es la rectificación de media onda [56, 59] y en la cual están basados la mayoría de las redes neuronales actuales.

$$f(z) = \max(0, a(z)) \tag{4.19}$$

Ecuación 4.19. Forma en la cual se calcula la unidad lineal rectificadora

**ReLU** es fácil de optimizar debido a que es similar a las unidades lineales, además de aprender más rápido en redes con muchas capas permitiendo que el entrenamiento de una red supervisada profunda sea a mayor velocidad induciendo la escasez y previniendo la desaparición del gradiente [55], tiene el problema de que todos los valores que son negativos se convierten automáticamente en cero, lo cual hace que la habilidad del modelo para ajustar o entrenar con los datos dados decrezca [68]. Los gradientes son consistentes debido a que la derivada de la operación de rectificación es 1 siempre que la unidad sea activa, lo que significa que la dirección del gradiente es más útil para aprender que las funciones de activación que requieren el efecto de una segunda derivada.

Otra opción es utilizar, la **función logística, sigmoideo o softmax** (Ecuación 4.20), la cual también existe de 0 a 1, y es ideal para aquellos modelos en los cuales se debe de predecir la probabilidad como una salida dentro de este rango [66]. La función es diferenciable, lo cual permite encontrar la pendiente en cualquier punto, puede causar que la red neuronal se quede trabada durante el entrenamiento [68]; generalmente la función de activación (Ecuación 4.18) va seguida de la función ReLU o logística.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{4.20}$$

Ecuación 4.20. Función sigmoide, también llamada logística o softmax

La activación de cada nodo oculto es multiplicado por un segundo conjunto de conectores de peso y sumado a un valor de sesgo. Una transformación logística de las entradas ponderadas a los nodos de resultado es aplicado para determinar el resultado final de la red, la salida de la red corresponde a la predicción de la red del resultado, y tendrá un rango de 0 a 1 cuando la función de activación del resultado de la red es la función logística; cada una de las redes es probada con un número diferente de nodos utilizando los mismos datos de entrenamiento para determinar el número de nodos que da un mejor aprendizaje [20, 66].

Cabe mencionar que se tienen variaciones de la función ReLU tratando de mejorar su desempeño, una de ellas es la función **Leaky ReLU** (ReLU defectuosa o con fuga) en la cual en lugar de tener una función que tenga un valor nulo en el origen, se tiene una pendiente diferente a la que posee después del origen con lo cual se incrementa el rango de la función dándole normalmente valores de 0.01 en adelante [68, 69]. Matemáticamente se define como:

$$y_i = \begin{cases} x_i & \text{si } x_i \geq 0 \\ \frac{x_i}{a_i} & \text{si } x_i < 0 \end{cases} \tag{4.21}$$

Ecuación 4.21. Definición de la función Leaky ReLU

En la Ecuación 4.14 se puede apreciar la forma en la cual se está desarrollando la regla de la cadena para los vectores de los pesos a partir de la función de activación dada por la Ecuación 4.18 con esto, se puede determinar que cuando  $a(z) > 0$  es como si la función de maximización expresada en la Ecuación 4.19 desapareciera y se obtiene la derivada de la función de activación con respecto a los pesos. De igual forma, puede apreciarse en la Ecuación 4.22 la aplicación de la derivada parcial al evaluar la función de activación con respecto a los sesgos siempre que se tenga una activación que sea diferente a cero 56, 59.

$$\frac{\partial a(x)}{\partial b} = \begin{cases} 0 & \because \frac{\partial a(z)}{\partial b} = 0 \quad \text{Cuando } a(z) \leq 0 \\ 1 & \because \frac{\partial a(z)}{\partial b} = 1 \quad \text{Cuando } a(z) > 0 \end{cases} \quad (4.22)$$

Ecuación 4.22. Derivada parcial de la función de activación con respecto a los sesgos

Las derivadas parciales descritas son las que se utilizan para describir la función de pérdida ya que, como se ha mencionado anteriormente, el entrenamiento de una neurona requiere que se tome la **derivada de la función de costo con respecto a los parámetros del modelo** esperando que la función de costo reduzca su valor de manera significativa y no necesariamente que lleve su valor hasta donde el gradiente sea cero 57. La Figura 4.4, tomada del curso impartido por Dan Becker 70 ilustra las funciones de activación más utilizadas.

Como se aprecia en la Figura 4.4, existen varias funciones de activación las cuales están dentro de las unidades ocultas de las redes de aprendizaje automático, generalmente se eligen las unidades ReLU, pero puede elegir cuándo usar cada una de estas puede ser una tarea difícil que llega a estar más en función de la experiencia e intuición de quien la diseñe, así como de hacerlo mediante la prueba y error 57.

Existen combinaciones que surgen a partir de las funciones de activación anteriormente descritas, tales como las unidades **maxout** que están compuestas por dos piezas que pueden aprender a implementar la misma función de la entrada  $x$  como una capa tradicional usando ReLU, la función de rectificación de valor absoluta o Leaky ReLU, así como aprender a implementar funciones que son totalmente diferentes 57.

Un fenómeno que puede ocurrir en las redes neuronales es el **olvido catastrófico**, en el cual la red neuronal olvida cómo desempeñar tareas para las cuales han sido entrenadas en el pasado, para evitar que esto ocurra cada unidad esté manejada por múltiples filtros 50.

Entre más capas tenga una red neuronal, los aspectos de la entrada que serán importantes para la discriminación se amplificarán y se podrán suprimir variantes que son irrelevantes 55, recordando que estas capas características son aprendidas por los datos y utilizan un procedimiento de aprendizaje de propósito general.

Al ajustarse gradualmente los pesos de la red y minimizar la diferencia entre el resultado predicho y el valor conocido de la variable de resultado; la diferencia es conocida como el **error de la red neuronal** 66 mostrada en la Ecuación 4.23.

$$e_i = w \cdot x_i + b - y_i(x) \quad (4.23)$$

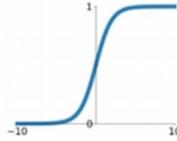
Ecuación 4.23. Término de error

Tanto la **tasa de aprendizaje** ( $\eta$ ), que es el monto de cambio de cada uno de los pesos después de cada caso de entrenamiento, como el **momentum**, que permite cambiar tanto los pesos como los sesgos en una forma proporcional a la magnitud del cambio previo, son utilizados en la propagación posterior al algoritmo de entrenamiento y ambos oscilan entre 0 y 1 66.

## Activation functions

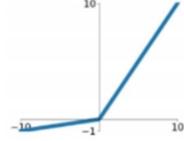
### Sigmoid

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



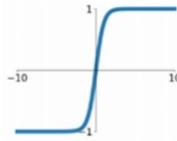
### Leaky ReLU

$$\max(0.1x, x)$$



### tanh

$$\tanh(x)$$

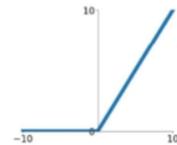


### Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

### ReLU

$$\max(0, x)$$



### ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

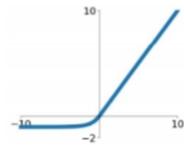


Figura 4.4: Funciones de Activación. Se ilustran las funciones de activación más utilizadas al crear redes neuronales [70].

$$w_{t+1} = w_t - \eta \frac{\partial C}{\partial w} \quad (4.24)$$

Ecuación 4.24. Forma en la cual se optimizan los pesos en función de la tasa de aprendizaje

$$b_{t+1} = b_t - \eta \frac{\partial C}{\partial b} \quad (4.25)$$

Ecuación 4.25. Actualización del sesgo de la neurona en la dirección opuesta al incremento de la función de costo

La tasa de aprendizaje dicta el monto de ajuste que se tendrá en las conexiones internas para mejorar la clasificación tras la falla, siendo este un parámetro importante en la precisión de la predicción [71] y probablemente sea el **hiperparámetro más importante** al momento de ajustar los valores para **mejorar el desempeño del algoritmo**, por lo que se recomienda que tiempo para únicamente ajustar un hiperparámetro sea éste el que se ajuste, ya que controla la capacidad efectiva del modelo en una forma más complicada que los otros hiperparámetros debido a que la capacidad efectiva del modelo es mayor cuando la tasa de aprendizaje es correcta para el problema de optimización y no cuando la tasa de aprendizaje es muy grande o muy bajo [57].

El sistema también puede ajustar la tasa de aprendizaje para optimizar su entrenamiento y evitar problemas como el sobreajuste; normalmente el valor de la tasa de aprendizaje es alto al inicio pero decrece en el transcurso del tiempo [71]. Cuando el valor de la tasa de aprendizaje es muy grande, el gradiente descendiente puede incrementar de forma inadvertida el error de entrenamiento, mientras que cuando es muy chico puede hacer que el entrenamiento sea muy lento y que se atasque de forma permanente en un alto error de entrenamiento [57].

Un patrón o caso de entrenamiento es seleccionado del conjunto de datos y posteriormente los valores de las variables predictivas son usadas como entradas a las redes para determinar los resultados

predichos de la red para ser comparados con las salidas deseadas. Si hay una **diferencia** entre lo predicho y lo deseado, se calcula una **señal de error** y propagación posterior a través de la red, la señal es un factor matemático que reduce la diferencia entre los dos valores [66]. Es común utilizar la **validación cruzada** para poder conocer el error mínimo tanto en el conjunto de datos de entrenamiento como en el de prueba, principalmente para poder evitar el sobreajuste [66].

La red de retroalimentación es más fácil de entrenar y generalizar que las redes con conectividad total entre las capas adyacentes, lo cual es la red neuronal de convolución (**ConvNet**), que ha sido ampliamente adaptadas por la comunidad de visión computarizada y se hablará de ellas a profundidad en la sección 4.3; ésta se estructura como una serie de etapas, las primeras están compuestas por capas convolucionales y la agrupación de las capas [55]. En estos casos se seguirá contando con capas de entrada y salida como se ilustra en la Figura 4.2 pero ahora se tendrán capas que realizan una función determinada en entre estas dos capas.

Las redes profundas tienen dos diferentes ventajas sobre los algoritmos clásicos de aprendizaje que no usan representaciones distribuidas, la primera es que permiten la **generalización de nuevas combinaciones** de los valores de las funciones aprendidas más allá de las vistas durante el entrenamiento.

En segundo lugar, la composición de capas de representación en una red profunda permite que las capas ocultas de una red neuronal multicapa pueda representar las entradas de las redes en una forma que haga fácil el **predecir las salidas objetivo** [55].

Un modelo puede describir una amplia gama de fenómenos al tener un gran número de parámetros libres, sin embargo la **verdadera prueba** para un modelo es su **habilidad para realizar predicciones** en situaciones en las cuales nunca había estado expuesto antes; sin embargo algunos de los problemas comunes son el *sobreajuste* (comúnmente conocido como *overfitting*) y *sobreentrenamiento*. Cuando se vea que la exactitud (*accuracy*) en los datos de prueba ya no se incrementa es cuando se debe de dejar de realizar el entrenamiento [56].

Para reducir el sobreajuste existen varias formas, una de ellas es incrementando la cantidad de datos de entrenamiento, mediante técnicas de regularización, abandono (conocido como *dropout*) en donde se van desactivando algunas capas con sus conexiones de entrada y salida, así como el aumento artificial de datos de entrenamiento [72].

La **regularización** se refiere a cualquier modificación que se realiza al algoritmo de aprendizaje que tiene como objetivo reducir el error de generalización luego de que se le ingresen los datos de prueba o validación, pero no su error de entrenamiento [57, 59]. La mayoría de las estrategias de regularización están basadas en regularizar estimadores, el cual trabaja cambiando un incremento en el sesgo por una reducción en la varianza.

Al usar la regularización, generalmente se elige una norma que penalice únicamente a los pesos de la transformación afín en cada capa y deja los sesgos sin regularizar debido a que generalmente estos requieren menos datos para un ajuste con mayor exactitud que los pesos.

El uso de *early stopping* es también considerada como una técnica de regularización, siendo una de las más utilizadas en *deep learning* debido a que es muy efectiva y simple debido a que no se está interfiriendo en la dinámica de aprendizaje como ocurre con la caída de los pesos. En este caso se está controlando la capacidad efectiva del modelo para determinar cuántos pasos se pueden tomar para ajustar el conjunto de entrenamiento lo cual se hace utilizando únicamente el hiperparámetro del tiempo de entrenamiento y el único costo significativo es el evaluar de manera periódica el conjunto de validación durante el entrenamiento [57]. Se debe de mostrar la curva de aprendizaje en donde el error del conjunto de validación tenga una curva en forma de ‘U’.

Por último, el *dropout* o la marginación, es otro método muy poderoso que es utilizado para la regularización. En esta técnica se dejan fuera aquellas unidades que no son de salida de una red de base subyacente; esto se realiza removiendo las unidades de una red al multiplicar sus salidas por cero con lo cual se hacen ligeras modificaciones a los modelos entrenando un conjunto consistente de todas las subredes que pueden ser construidas y omitiendo algunas de ellas en base a una probabilidad de 0.8 para las de entrada y 0.5 para las ocultas.

El modelo debe entrenarse al construir  $k$  diferentes conjuntos de datos utilizando un minilote del algoritmo de aprendizaje que realice pequeños pasos tal como el gradiente estocástico descendiente en donde cada vez que se carga un ejemplo en un minilote se muestrea de manera aleatoria una máscara binaria para aplicar a todas las entradas y las capas ocultas de la red donde la probabilidad de muestrear un valor de máscara de uno es un hiperparámetro ajustado antes de que el entrenamiento comience [57].

Usar *dropout* es más efectivo que otras regularizaciones, además de que no requiere demasiado costo computacional y no limita significativamente el tipo de modelo o procedimiento de entrenamiento que puede ser usado trabajando bien con casi cualquier modelo [56, 57].

### 4.3. Redes Neuronales Convolucionales

Las redes neuronales convolucionales, conocidas por sus siglas en inglés como CNN (*Convolutional Neural Networks*), son una forma especializada de redes neuronales que tienen una topología cuadrículada bien conocida siendo muy exitosas en aplicaciones prácticas [55, 57].

Su nombre indica que la red emplea operaciones matemáticas de convolución, la cual es un tipo de operación lineal especializada. Las CNN son simplemente redes neuronales que utilizan la convolución en lugar de una matriz general de multiplicación en al menos una de sus capas [57].

La convolución es una operación con dos funciones de un argumento real valorado. La ecuación 4.26 muestra la operación de convolución, en donde el valor de  $x$  se refiere a la entrada, mientras que  $w$  es conocido como el *kernel* (el cual es una función que evita los problemas de los cálculos de alta dimensión en un núcleo al transformar los datos no lineales a casi lineales al tener núcleos para el análisis de patrones), mientras que la salida es conocida como el mapa de características o *feature map*, suele usarse sobre más de un eje al mismo tiempo [56].

$$s(t) = (x * w)(t) \quad (4.26)$$

Ecuación 4.26: Operación de convolución

Las entradas normalmente son arreglos multidimensionales de datos y el *kernel* usualmente es un arreglo multidimensional de parámetros que es adaptado por el algoritmo de aprendizaje, estos arreglos multidimensionales serán conocidos como tensores. Debido a que las entradas y el kernel debe ser almacenado de forma separada, asumimos que estas funciones son cero en cualquier lugar, pero el conjunto finito de puntos para el cual almacenamos los valores [57].

La convolución involucra tres importantes ideas que pueden ayudar a implementar un sistema de aprendizaje de máquina (*machine learning*), las cuales son interacciones escasas, compartir parámetros y representaciones equivariantes, sin embargo, la convolución provee un medio de trabajo con entradas de tamaño variable [57].

La **limitación de interacción** se logra haciendo un *kernel* más pequeño que la entrada detectando solamente pequeñas características que sean significativas de los miles de píxeles que se tengan,

siendo estas bordes con *kernels* que ocupan solamente decenas o cientos de píxeles almacenando menos parámetros reduciendo así los requerimientos de memoria del modelo e implementando una eficiencia estadística con lo que se realizan menos operaciones computacionales [57]. Con esto lo que se propicia es que la red describa de forma eficiente interacciones complicadas entre muchas variables al construir sus interacciones de bloques simples construidos que describe cada uno solamente interacciones limitadas.

El **compartir parámetros** es el estar utilizando los mismos parámetros para más de una función en un modelo, se puede considerar que la red tiene pesos atados debido a que el valor de los pesos aplicados a una entrada está atado al valor de los pesos aplicados en otra parte [57]. Esto se refiere a que en vez de tener un aprendizaje de conjuntos separados de parámetros para cada ubicación, solamente se aprende un conjunto con lo que se reduce el almacenamiento de requerimientos del modelo para los  $k$  parámetros, siendo  $k$  de muchos órdenes de magnitud menores que los valores de entrada  $m \times n$ .

Al decir que una función es **equivariante**, se refiere a que si las entradas cambian, las salidas cambiarán en la misma forma. Una función  $f(x)$  será equivariante a una función  $g$  si  $f(g(x)) = g(f(x))$  [57]. La convolución crea un mapa de dos dimensiones en donde ciertas características aparecen en la entrada. Esto es útil cuando se conoce que una función de un número pequeño de píxeles vecinos es relevante al aplicarse a ubicaciones de múltiples entradas, por citar un ejemplo, esto ocurre cuando se detectan bordes en la primera capa de una red convolucional, estos mismos bordes pueden aparecer casi en cualquier lado de la imagen por lo que es práctico el compartir parámetros a través de toda la imagen, sin embargo esto no es algo que se quiera realizar en todas las capas ya que en algunas más profundas será necesario identificar otras características.

Una capa típica de **una red neuronal convolucional consta de tres etapas**: En la primera se realizan varias **convoluciones de forma paralela** para producir un conjunto de activaciones lineales, en la segunda cada activación lineal está corriendo a través de una función de activación no lineal como en una función de activación rectificadora lineal llamada **etapa de detección**, mientras que en la tercera se utiliza una **función de agrupamiento** (*pooling*) para modificar la salida de la capa [57].

En una red convolucional multicapa la entrada a la segunda capa es la salida de la primera capa, teniendo la salida de muchas convoluciones diferentes en cada posición; generalmente se trabaja en tensores de tres dimensiones con un índice en los diferentes canales y dos índices en las coordenadas espaciales de cada canal. Las implementaciones en software normalmente trabajan en modo de lotes, por lo que realmente usarán tensores de cuatro dimensiones, siendo esta el eje en donde se indexan diferentes ejemplos en el lote [57].

Debido a que las redes convolucionales normalmente utilizan múltiples canales de convolución, las operaciones lineales en las que están basadas no se garantiza que sean conmutativas, a menos de que cada operación tenga el mismo número de canales de salida como canales de entrada [57].

Las redes convolucionales pueden ser utilizadas a la salida de una alta dimensión estructurando un objeto en vez de solamente predecir una etiqueta de clase para una tarea de clasificación o un valor real para una tarea de regresión. En las redes utilizadas normalmente para la clasificación de un objeto en una imagen, la mayor reducción en la dimensión espacial de la red viene de las capas de agrupación con un paso grande. Una estrategia para etiquetar los píxeles de una imagen es el producir una estimación inicial de la etiqueta de la imagen para luego hacer un paso de ajuste fino de esta estimación al utilizar las interacciones entre los píxeles vecinos [57, 59].

El repetir este paso de ajuste fino varias veces corresponde a utilizar las mismas convoluciones en cada etapa compartiendo pesos entre las últimas capas de la red profunda con lo que se hace una secuencia de desempeño de cómputo seguido por las capas convolucionales sucesivas con pesos compartidos a través de las capas para un tipo particular de una red recurrente [57]. Una vez que la predicción para cada píxel es realizada, se pueden utilizar varios métodos para su procesamiento posterior para obtener una segmentación de la imagen en las regiones.

Los grupos grandes de píxeles contiguos tienen una tendencia a estar asociados con la misma etiqueta; la red neuronal convolucional puede ser entrenada para maximizar una aproximación del entrenamiento de un objeto en un modelo gráfico [73, 74].

En las redes neuronales los datos tienen varios canales que representan una cantidad diferente en algún punto del espacio o del tiempo, cada ejemplo de los datos de entrenamiento y de prueba debe de tener la misma dimensión espacial, una ventaja de las CNN es que pueden procesar entradas que varían en su extensión espacial; las entradas pueden no ser representadas mediante redes neuronales tradicionales basadas en la multiplicación de matrices [56, 57].

Cuando el ancho y la altura de una imagen varía en un conjunto de datos a analizar la convolución es algo sencillo de aplicar debido a que el kernel es aplicado simplemente un número diferente de veces dependiendo del tamaño de la entrada y la salida de la operación de convolución de acuerdo a su escala [57]. La convolución puede ser vista como una multiplicación de matrices, el mismo kernel de convolución induce a diferentes tamaños de una matriz circulante doblemente bloqueada para cada tamaño de entrada. En algunas ocasiones la salida de la red es permitido que tenga tamaños variables así como sucede a la entrada [57].

Generalmente la forma más tardada de un entrenamiento de una CNN es el aprendizaje de las características, siendo la capa de salida la menos costosa computacionalmente debido a los números tan pequeños de características proporcionadas como entradas a esta capa después de pasar a través de varias capas de agrupación (*pooling*). Una forma de reducir el costo del entrenamiento de una CNN es el utilizar características que no son entrenadas en una forma supervisada [57].

Las CNN son una forma especializada de redes neuronales para trabajar con datos que tienen una topología estructurada en cuadrícula y para escalar aquellos modelos que son de gran tamaño; esta aproximación ha sido de gran éxito cuando se tienen imágenes con una composición en dos dimensiones [57].

Las redes tendrán características distintas en su composición, modelado y tamaño de ingreso de las imágenes. Generalmente las imágenes se ajustan de manera automática a las dimensiones especificadas; a pesar de que existen varias arquitecturas, las que se utilizaron para analizar nuestro tema de investigación fueron *AlexNet*, *Inception V3*, *GoogLeNet*, *VGG 16*, *VGG 19*, *ResNet 50*, *ResNet 101*, *ResNet 152*, de las cuales se brinda también una semblanza. Al momento de comparar los diferentes modelos de aprendizaje automático profundo, se recomienda que los modelos de redes neuronales que utilizan convolución se comparen únicamente su desempeño estadístico únicamente con aquellos que utilicen este mismo principio.

Independientemente del tipo de red que se utilice, esta siempre estará limitada por la cantidad de **memoria disponible** y el **tiempo de entrenamiento** que se va a tolerar para cada una de estas. El entrenamiento en algunas de estas puede inclusive durar varios días o semanas dependiendo además del problema a resolver y de la cantidad de datos que se tengan.

### 4.3.1. Capas de Una Red Neuronal Convolucional

Una red neuronal convolucional está compuesta por diferentes capas que le permiten realizar diferentes tareas; pueden organizarse en capas de entradas, de extracción de características y de clasificación, de las cuales se obtendrá la predicción en ciertas categorías de las entradas de la red [75]; la Figura 4.5 tiene una representación esquemática de algunas capas de una red neuronal convolucional.

Las redes pueden tener capas de diferentes tipos y múltiples combinaciones, pero es posible clasificarlas en función de la tarea que desempeñan de la siguiente manera [76]:

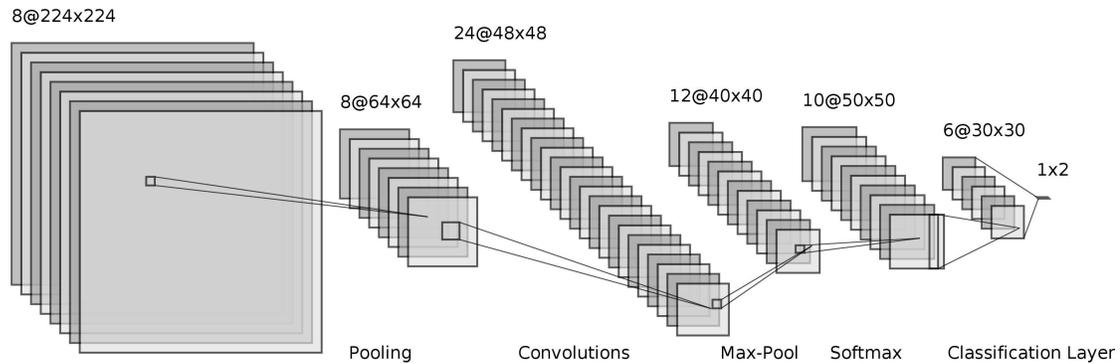


Figura 4.5: Representación de Capas de una Red Neuronal Convolutional. De forma abreviada se muestra la estructura de las capas de una red neuronal.

- ❖ **Capa de entrada:** Se encargan del ingreso de los píxeles de las imágenes, generalmente conformadas por el ancho, alto y número de canales que tendrán. Hay diferentes tipos al poder admitir una imagen de 2 ó 3 dimensiones, así como secuencias de datos o de reconocimiento de objetos de entrada.
- ❖ **Capas de convolución y totalmente conectadas:** Aprende las características; tiene varias capas de activación. Pueden aplicarse filtros convolucionales deslizantes a las imágenes de entrada, aplicar mapas de características de muestras ascendentes, así como capas totalmente conectadas, las cuales multiplican la entrada por un peso de la matriz para hacer la adición de un vector de sesgos.
- ❖ **Capas secuenciales:** Estas capas permiten que la red sea alimentada por medio de secuencias de datos a una red, estableciendo también el tamaño de entrada de un elemento de la secuencia por pasos de tiempo. La red tiene una memoria para recordar eventos importantes que ocurren de forma recurrente.
- ❖ **Capas de activación:** Es en donde ocurren las operaciones de activación al establecer capas ReLU, tangentes, exponenciales, entre otras, se pueden ver algunas de estas en la Figura [4.4](#). Introducen la no linealidad a la red, permitiendo que el modelo aprenda patrones más complejos en los datos, generalmente son utilizadas previo a las capas de salida para poder normalizar los puntajes (*scores*).
- ❖ **Capas de normalización, abandono (*dropout*) y de recorte:** Realizan una normalización de cada canal a través del mini conjunto de datos a estudiar, un canal de respuesta, una capa de abandono estableciendo una probabilidad de que ciertas capas sean anuladas, o un recorte a las entradas.
- ❖ **Capas de agrupación y desagrupación:** Reemplaza la salida de la red en una cierta locación con una suma estadística de las salidas cercanas. Esto se realiza seleccionando las características más fuertes al reducir el número de parámetros que le son más importantes al reducir la dimensionalidad de cada mapa, pero reteniendo la información importante, la cual puede ser de diferentes tipos. El agrupamiento máximo (*Max Pooling*) toma el elemento más grande del mapa de características rectificado, el promedio de agrupamiento (*Average Pooling*) divide las entradas en regiones de agrupamiento rectangulares para computar los valores promedio de cada región, y para el agrupamiento de la sumatoria (*Sum Pooling*) con la sumatoria de todos los elementos

en el mapa de características. Lo anterior puede suceder en 2 ó 3 dimensiones, según sea el caso. Debido a que el agrupamiento resume las respuestas sobre todo un conjunto de píxeles vecinos, es una posibilidad el utilizar menos capas de agrupación que las unidades de detección. Cuando el número de parámetros en la siguiente capa es una función del tamaño de entrada, esta reducción en la entrada puede arrojar una mejora en la eficiencia estadística y una reducción de los requerimientos de la memoria para almacenar los parámetros.

- ❖ **Capas de combinación:** Se tiene una concatenación de capas en las cuales se especifica que deben de tener el mismo tamaño en una dimensión especificada.
- ❖ **Capas de detección de objetos:** Son las capas enfocadas en la detección de mapas de características.
- ❖ **Capas de salida:** Es en donde se aplica la clasificación en una red neuronal donde se tienen las categorías a las cuales se van a predecir que pertenece una entrada. Algunas de las más características son la capa softmax, clasificación y regresión, sin embargo existen otras enfocadas en los píxeles de las entradas.

### 4.3.2. AlexNet

La arquitectura conocida como *AlexNet* [77] es una de las más importantes debido a que fue una de las primeras que se presentó y mejoró el estado del arte al conseguir una tasa de error en los datos de prueba de 17% en el 2010 en el concurso de clasificación de la base de datos de imágenes de *ImageNet* [78].

En dicho concurso se tienen más de 1.2 millones de imágenes en alta resolución para ser clasificados en 1000 diferentes categorías; de igual forma, esta arquitectura sufrió algunas variaciones en el 2012 resultando de nueva cuenta ganadora tras registrar una tasa de error del 15.3% a diferencia del segundo lugar que obtuvo 26.2%.

*AlexNet* revolucionó los modelos de *deep learning* con el uso que tuvo de redes neuronales convolucionales, debido a que estos son modelos que son fáciles de controlar y mucho más sencillas de entrenar, pero que tenían el problema de no poder utilizar imágenes de alta resolución y por el corto tiempo que debían de ser entrenadas. Es por esto que es una de las arquitecturas que, dependiendo de la cantidad de datos, puede ser entrenada también en una computadora que no tenga GPU. Para la competencia de *ImageNet* hizo uso de dos GPU's durando 5 días para poder dar los resultados con los cuales ganó la competencia en esos años.

Cuenta con 60 millones de parámetros y 650 mil neuronas que están acomodadas en 25 capas de las cuales 8 contienen capas de aprendizaje con sus respectivos pesos (5 capas convolucionales y 3 capas completamente conectadas), así como el uso de *dropout* [77].

Los filtros de la primera capa convolucional tienen una entrada de imagen de  $224 \times 224 \times 3$  con 96 kernels de tamaño de  $11 \times 11 \times 3$  con un paso de 4 píxeles (distancia entre el centro del campo receptivo de las neuronas vecinas en un mapa kernel) [77]. La segunda capa convolucional toma como entrada la salida de la primera capa convolucional y la filtra con 256 kernels de tamaño  $5 \times 5 \times 48$ , la tercera, cuarta y quinta capa convolucional están conectadas la una a la otra sin ninguna capa de agrupación o normalización [77].

La tercera capa convolucional tiene 384 kernels de tamaño  $3 \times 3 \times 256$  conectadas a las salidas (que son las capas agrupadas y normalizadas) de la segunda capa convolucional. La cuarta capa convolucional tiene 384 kernels de tamaño  $3 \times 3 \times 192$  mientras que la quinta capa convolucional tiene 256

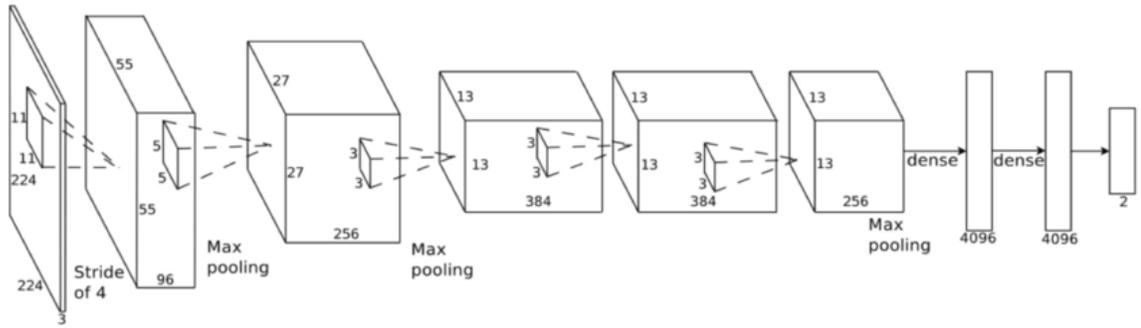


Figura 4.6: Esquema de la arquitectura *AlexNet*. Se puede apreciar la forma en la que se representa la arquitectura de *AlexNet* y el acomodo de sus capas [77].

kernels de tamaño  $3 \times 3 \times 192$ ; las caspas totalmente conectadas tienen 4096 neuronas cada una [77], la Figura 4.6 ilustra la forma en la cual están representados.

*AlexNet* utiliza *dropout* para evitar el sobreajuste; esto se realiza ajustando a cero las salidas de cada neurona oculta con una probabilidad de 0.5; aquellas neuronas que no son consideradas no contribuyen ni en la propagación hacia adelante ni hacia atrás, de forma tal que cada vez que se tiene una entrada la red neuronal muestra una arquitectura diferente, pero cada una de estas arquitecturas comparte los mismos pesos con lo cual se reducen las co-adaptaciones de neuronas debido a que una neurona no puede confiar en la presencia de otra neurona en particular, por lo cual es forzada a aprender características más robustas que son útiles en conjunto con muchas formas diferentes de subconjuntos de las otras neuronas [77].

### 4.3.3. GoogLeNet

Lleva este nombre en honor a Yann LeCun quien fue uno de los pioneros en *deep learning*. Introduce un nuevo nivel de organización en la forma en la que utiliza el módulo “*Inception*” o de principios, además de incrementar la profundidad de la red. El módulo *inception* se refiere a encontrar como una estructura escasa óptima local en una red convolucional de visión puede ser aproximada y cubierta por componentes densos ampliamente disponibles, lo cual implica que se construirá a partir de bloques convolucionales, los cuales deberán de buscarse las construcciones locales óptimas y repetirse espacialmente [79].

La construcción se realiza capa por capa en donde se debe de analizar la relación de la última capa y agruparlos en grupos de unidades con una alta correlación. Los grupos forman las unidades de la próxima capa y son conectados a las unidades en la capa previa [80]. Cada unidad de la entrada anterior corresponde a una región de la imagen de entrada y estas unidades son agrupadas en bancos de filtros; en las capas que están más cercanas a las entradas las unidades correlacionadas se concentran en regiones locales, con lo cual se tendrán varias se terminará con varias agrupaciones concentradas en una región sencilla y podrán ser cubiertas por una capa de convoluciones de  $1 \times 1$  en la siguiente capa; el tamaño que tengan las convoluciones variará según la región y de la conveniencia que se vea en cada una de estas [79].

De forma general, una red *inception* o de comienzos, consiste en módulos del tipo anterior apilados uno tras otro con capas ocasionales de *max pooling* con paso 2 para reducir a la mitad la resolución de la cuadrícula [79]. La Figura 4.7 muestra las 144 capas que conforman a la red de las cuales 22 de ellas son profundas.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Figura 4.7: Representación de las capas de la arquitectura GoogLeNet. Se pueden apreciar cada una de las capas que tiene la red, así como los tamaños de salida, profundidad, los pasos que tiene y los parámetros que la conforman [79]

Todas las convoluciones de la red, incluidas aquellas que están dentro de los módulos *inception* usan activaciones lineales rectificadas; el tamaño del campo receptivo de nuestra red es de  $224 \times 224 \times 3$  tomando canales de color RGB con una substracción media [79]. La red fue diseñada pensando en practicidad y eficiencia computacional.

*GoogLeNet* tiene la peculiaridad de que además de haber sido entrenada en *ImageNet* [78], fue también entrenada con la base de datos *Places365* [81, 82], la cual clasifica imágenes en 365 categorías de diferentes lugares.

#### 4.3.4. Inception V3

En esta red se sigue una idea muy similar a la que se planteó en *GoogLeNet*, debido a que esa fue la versión 1 de una red *inception*, con la diferencia de que en esta arquitectura se propone una factorización de una capa convolucional tradicional de  $7 \times 7$  en tres convoluciones de  $3 \times 3$ ; se cuenta con 3 módulos de *inception* tradicionales en  $35 \times 35$  con 288 filtros cada uno que es seguido de 5 instancias de los módulos *inception* factorizados y posteriormente se reduce a una rejilla de  $8 \times 8 \times 1280$  con una técnica de reducción de rejilla [83].

La calidad de la red se mostró relativamente estable a variaciones. A pesar de que la red tiene 42 capas de profundidad, el costo computacional es de solamente 2.5 veces mayor que *GoogLeNet*, sumando en total 316 capas [83]. Con el objetivo de evitar el sobre ajuste, la red cuenta con un mecanismo llamado regularización de suavizado de etiqueta el cual cambia la distribución de etiquetas correctas; esto es equivalente a reemplazar la pérdida de entropía cruzada [83].

Usualmente se tiene la creencia de que los modelos que emplean mayores resoluciones receptivas en el campo tienen como resultado un mejor desempeño en tareas de reconocimiento; de cualquier

modo es importante el distinguir entre el efecto del incremento en la resolución del campo receptivo de la primera capa y el efecto de los modelos de gran capacidad y cómputo. Lo anterior debido a que si solamente se cambia la resolución de la entrada sin ir más allá de un ajuste en el modelo se estarán utilizando modelos computacionalmente menos costosos para resolver tareas más difíciles [83].

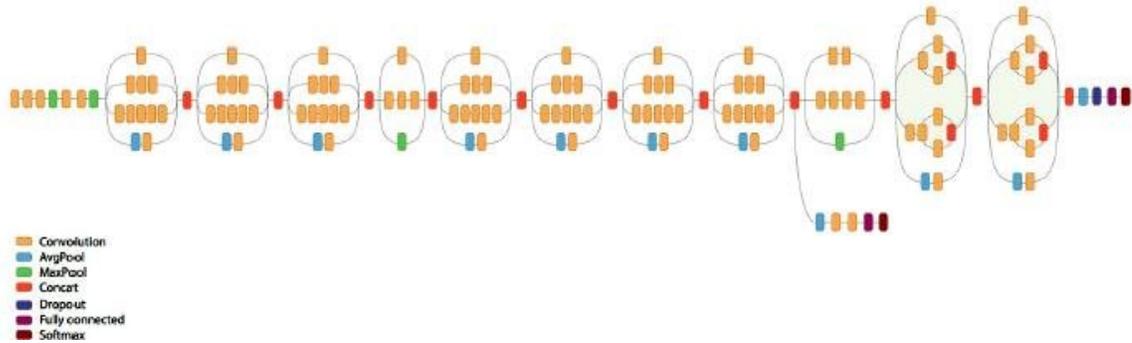


Figura 4.8: Representación de las capas de la arquitectura Inception V3. El esquema indica la forma en la cual se tienen módulos apilados en esta red [83]. Imagen cortesía [84].

De esta forma, es que esta red tiene como una de las novedades el uso de imágenes con un tamaño de entrada de  $299 \times 299 \times 3$ , incrementando así el tamaño a diferencia de otras redes que mantienen su tamaño de entrada de  $224 \times 224 \times 3$ . Es así que en la combinación de parámetros inferiores y la regularización adicional con normalización de lote de clasificadores auxiliares y suavizado de etiquetas permite para un entrenamiento de redes de alta calidad en conjuntos de entrenamiento de tamaño relativamente modestos [83].

#### 4.3.5. VGG

Tiene una simplicidad arquitectónica en las características pero requieren un algo costo computacional. Una de las principales contribuciones es la evaluación minuciosa de redes de profundidad creciente usando una arquitectura con filtros de convolución chicos de  $3 \times 3$  incrementando la profundidad de la red al añadir más capas convolucionales con lo cual se incrementa la exactitud [85].

Es precisamente gracias a que la red es más profunda en su diseño que se puede realizar esta tarea. La arquitectura tiene par de variantes, una de 16 y otra de 19 capas de pesos aprendibles para que en total tengan 41 y 47 capas respectivamente. El tamaño de entrada de las imágenes a cualquiera de las dos variantes es de  $224 \times 224 \times 3$  [85].

La red solamente tiene un pre-procesamiento al realizar la resta del valor medio RGB calculada en el conjunto de entrenamiento de cada píxel. La imagen pasa a través de las capas convolucionales apiladas donde se utilizan filtros con un campo receptivo muy pequeño de  $3 \times 3$  [85].

Un apilamiento de capas convolucionales con diferentes arquitecturas es seguido de tres capas totalmente conectadas en donde las primeras dos tienen 4096 canales cada una, mientras que la tercera realiza la tarea de clasificación y contiene 1000 canales, perteneciendo una a cada clase. La última capa es una capa *softmax*; es de resaltar que la configuración de las capas totalmente conectadas es la misma en todas las redes y se sigue una rectificación ReLU no lineal [85].

Al utilizar un apilamiento de capas convolucionales de  $3 \times 3$  en vez de una capa de  $7 \times 7$  se tiene la ventaja de una incorporación de tres capas no lineales rectificadas que hacen una función de decisión

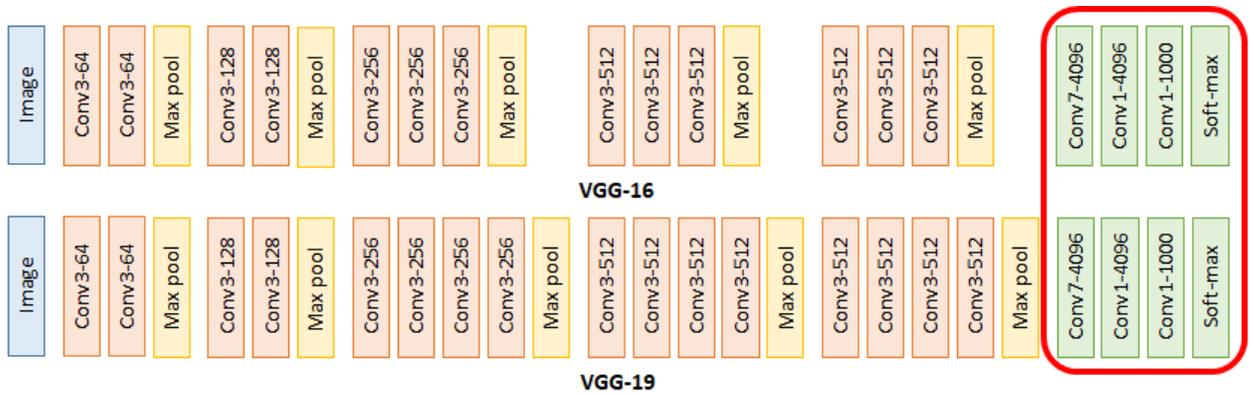


Figura 4.9: Representación de las arquitecturas VGG. Se muestran las formas en las cuales está la arquitectura VGG en sus dos variantes que son con 16 y con 19 capas de pesos aprendibles. Imagen cortesía [86].

más discriminativas, así como una reducción de parámetros a considerar [85]. Para el entrenamiento la red utilizó un tamaño de lote de 256 con un momentum de 0.9 y una regularización por caída de peso y *dropout* para las primeras dos capas totalmente conectadas; el tamaño de entrada de las imágenes es de  $224 \times 224$ .

La Figura 4.9 muestra las dos variantes de la red, que son con 16 y con 19 capas de pesos aprendibles, se puede apreciar que se le incrementa una capa extra de convolución después de la segunda capa de agrupamiento máximo en cada una de los siguientes tres bloques apilados. A pesar de que el tiempo de cómputo se incrementa, la red es una de las que arroja en diversos problemas resultados que tienen mayor exactitud, por lo cual llega a valer la pena el emplearla para problemas de clasificación de imágenes.

#### 4.3.6. ResNet

El aprendizaje profundo residual fue reportado por He *et al* [87] teniendo como una de sus principales características el ser una de las redes más profundas que existen. En estas redes, en lugar de esperar que pocas capas apiladas directamente se ajusten a un mapeo subyacente deseado, explícitamente se deja que estas capas ajusten un **mapeo residual** en las cuales se espera que las nuevas capas se ajusten a una nueva capa en una función identidad, con lo cual se puede tener un modelo con mejores soluciones para ajustar los datos del conjunto de entrenamiento para reducir así errores.

La formulación puede realizarse por redes neuronales previas con **conexiones de corto circuito** como se muestra en la Figura 4.10 [88] en los cuales pueden **saltarse una o más capas**, en este caso desempeñando un **mapeo de identidad** y sus salidas son añadidas a las salidas de las capas apiladas.

El mapeo subyacente deseado que se quiere obtener por el aprendizaje es  $f(x)$  para ser usado como la entrada de la función de activación; tradicionalmente el mapeo se aprende directamente como  $f(x)$ , mientras que en los bloques residuales se tiene un mapeo residual  $f(x) - x$ . En este caso, si el mapeo de identidad  $f(x) = x$  es el mapeo subyacente deseado, el mapeo residual es fácil de aprender debido a que solamente se necesita jalar los pesos y los sesgos a la siguiente capa de pesos a cero.

Se tiene la hipótesis de que es más fácil el optimizar el mapeo residual que optimizar el original; si un mapeo de identidad fuera óptimo, sería mucho más sencillo el empujar los residuos a cero que el

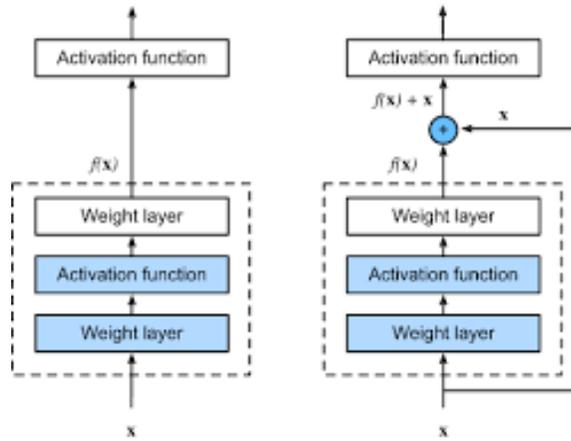


Figura 4.10: Comparativa de un bloque regular y un bloque residual. La imagen mostrada [88] indica la adición del operador llamado conexión de corto circuito.

ajustar un mapeo de identidad por una pila de capas no lineales. La red puede ser entrenada de inicio a fin por el gradiente estocástico descendiente con retropropagación.

Los corto circuitos identidad son libres de parámetros, la formulación propuesta siempre aprende funciones residuales ya que los corto circuitos nunca están cerrados, por lo cual la información siempre pasa a través de ella con funciones adicionales residuales para ser aprendidas. Los **corto circuitos de identidad** pueden ser directamente **usados cuando las entradas y las salidas son de las mismas dimensiones**. Cuando las dimensiones se incrementan se pueden adicionar ceros para incrementar las dimensiones sin incrementar parámetros, o que la proyección de corto circuito sea usada para que coincidan las dimensiones.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10 <sup>9</sup>	3.6×10 <sup>9</sup>	3.8×10 <sup>9</sup>	7.6×10 <sup>9</sup>	11.3×10 <sup>9</sup>

Figura 4.11: Representación de las capas de la arquitectura ResNet. Las variantes de la arquitectura y las capas que conforman a esta red pueden apreciarse en la siguiente imagen del estudio de He *et al* [87].

Con la reformulación del aprendizaje residual, si el mapeo de identidad es óptimo, los solucio-

nadores pueden simplemente conducir los pesos de las capas no lineales múltiples hacia cero para aproximar los mapeos de identidad. La red no utiliza la técnica de *dropout*, sin embargo cuenta con la normalización por lote (*batch normalization*) después de cada convolución y antes de una activación.

En la Figura 4.11 se aprecian las variantes que tiene la red, pudiendo apreciarse las diferencias entre cada una de estas. La red se conforma por 5 bloques de construcción que se muestran entre corchetes junto con el número de bloques que están apilados, siendo estas las que difieren. El tamaño de las imágenes de entrada es de  $224 \times 224 \times 3$ , independientemente de la variante de la red que se vaya a elegir.

## 4.4. Transferencia de Aprendizaje

Cuando se habla de una **arquitectura** se tiene una referencia a la estructura de una red [57], con esto se debe de determinar la cantidad de unidades que debe de tener, así como la forma en la cual se deben de conectar entre ellas. Las redes están organizadas en grupos que se llaman capas formando estructuras en cadena, siendo cada una de estas una función de la capa precedente a partir de lo cual se elige la profundidad de la red y lo ancho de cada una de las capas [56, 57].

Aunque de forma ideal se recomienda que las capas no estén conectadas en cadenas para que puedan saltarse conexiones entre ellas haciendo que se salten para que fluya el gradiente de las capas cercanas a la salida a las capas cercanas a la entrada [56, 57].

Al elegir un modelo profundo partimos de la creencia de que la función que queremos usar para el aprendizaje está compuesta por múltiples funciones simples con lo cual se siguen múltiples pasos donde cada paso hace uso de los pasos previos de salida. Las conexiones entre las capas son puntos importantes en el diseño de las arquitecturas; de manera predeterminada una red neuronal describe una transformación lineal por medio de una matriz  $W$ , cada unidad de entrada está conectada a una unidad de salida; en la medida en la cual las redes tengan menos conexiones teniendo pequeños subconjuntos de unidades en la capa de salida se podrán tener un menor número de parámetros y menor tiempo al evaluar la red [57].

El **teorema universal de aproximación** indica que existe una red lo suficientemente grande para conseguir cualquier grado de exactitud deseada, pero este no indica qué tan grande será la red. A su vez, independientemente de qué función se esta tratando de aprender, un perceptrón multicapa será capaz de representar una función, sin embargo, no está garantizado que el algoritmo de entrenamiento será capaz de aprender una función [57], lo cual puede ocurrir por dos diferentes razones:

1. El algoritmo de optimización usado para el entrenamiento puede no ser capaz de encontrar el valor de los parámetros que corresponden a la función deseada.
2. El algoritmo de entrenamiento puede elegir la función incorrecta debido a un sobreajuste.

Lo anterior se refiere a que no existe un procedimiento universal para examinar un conjunto de entrenamiento de ejemplos específicos y eligiendo una función que pueda generalizar los puntos que no están en el conjunto de entrenamiento.

La forma en la cual se encuentra una arquitectura de red ideal para realizar una tarea es en base a la experimentación guiándose al monitorear el error en el conjunto de validación. Existen diferentes arquitecturas de redes neuronales convolucionales que fueron previamente entrenadas con la base de datos *ImageNet* [78], en los cuales se tiene más de un millón de imágenes que son clasificados en una de

las mil categorías diferentes de objetos, por lo que han podido aprender características que son útiles para la clasificación en un amplio rango de imágenes [89].

Esta capacidad de aprendizaje es lo que se desea aprovechar para utilizarlo en otras tareas de clasificación en las cuales no se tenga un número tan basto de ejemplos para poder realizar esta labor, siendo algo comúnmente utilizado y en ocasiones hasta deseado en el aprendizaje profundo ya que con esta y un ajuste fino de los parámetros puede ser mucho más rápido y sencillo el entrenar una red con pesos iniciados de manera aleatoria [89, 90].

En esta técnica las últimas tres capas de la arquitectura es reemplazada con las que correspondan al problema en cuestión como se ilustra en la Figura 4.12; cada una de estas puede tener diferentes nombres, pero se referirán a las etiquetas de clasificación, la probabilidad de pertenecer a esta y la salida. De forma empírica se ha podido observar que cuanto más profundas sean las redes, éstas pueden hacer una mejor generalización para resolver diferentes problemas de reconocimiento y clasificación de imágenes, sin embargo, en tanto se tengan menores conexiones se reducirán los parámetros y el cómputo requerido para evaluar la red [56, 57].

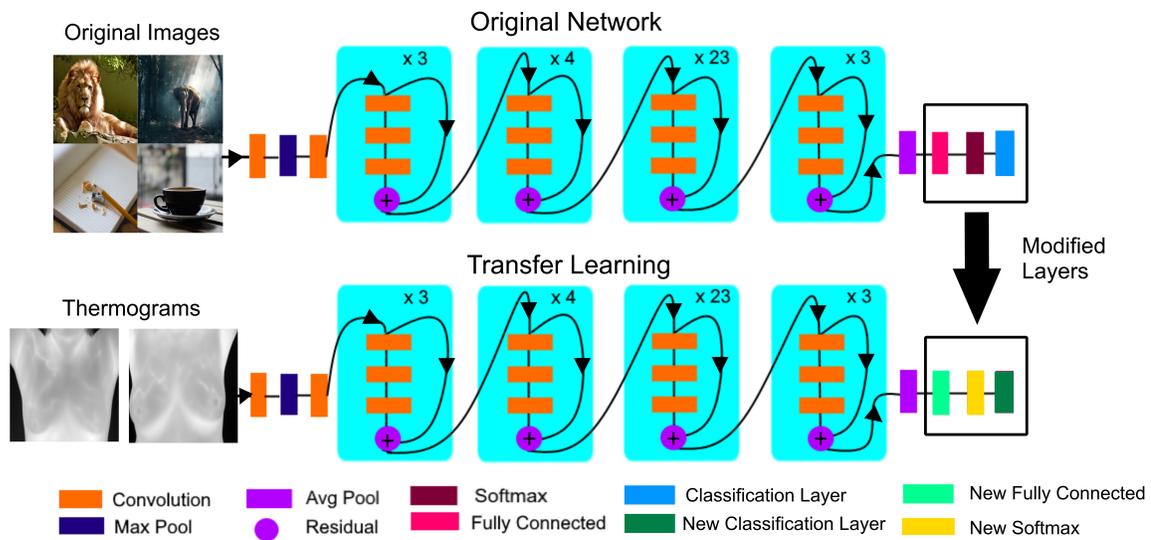


Figura 4.12: Representación de Transferencia de Aprendizaje. Se muestra la forma en la cual son modificadas las últimas capas en las capas para aprender una nueva labor de clasificación, en este caso para una arquitectura ResNet101.

Cuando se hable de transferencia de aprendizaje se tienen tres problemas de investigación principales, que son: ¿Qué?, ¿Cómo? y ¿Cuándo? transferir para poder realizar una tarea de manera general y no solamente para el problema originalmente analizado [90].

La transferencia de aprendizaje puede realizarse de diferentes formas:

- **No Supervisada:** Los datos no se encuentran etiquetados ni en la fuente ni en los objetivos.
- **Transductivo:** Los datos etiquetados están disponibles solamente en la fuente de datos original.
- **Inductivo:** Los datos etiquetados están disponibles en el objetivo, pudiendo ser un aprendizaje multi tarea si también están etiquetados los datos originales, o autodidacta si los datos originales no están etiquetados.

En todas las redes en donde se aplique la transferencia de aprendizaje se recomienda que en las nuevas capas transferidas se incrementen los factores de la tasa de aprendizaje de la capa completamente conectada para poder tener un aprendizaje que sea más rápido [91].

Para evitar la transferencia negativa de aprendizaje, se necesita verificar la factibilidad de que se pueda realizar entre la fuente original y la tarea objetivo a la cual se quiere aplicar [90].

Aunque los parámetros que varían en la transferencia de aprendizaje son mucho menores que cuando una arquitectura se elabora desde cero, estos deberán de tener un ajuste fino para que la red pueda optimizar su desempeño. Algunos de los parámetros que se pueden variar son:

- † **Weight Learn Rate Factor:** Tasa de factor de aprendizaje de pesos.
- † **Bias Learn Rate Factor:** Tasa de factor de aprendizaje de sesgos.
- † **Max Epochs:** Un *epoch* es un ciclo de entrenamiento completo de todo el conjunto de datos de entrenamiento terminando hasta cuando finaliza la modificación de los pesos. Debe de establecerse el máximo número de ciclos que se tendrán en el entrenamiento de la red. El **número** de *epochs* estará dado en **función de qué tan diversos son los datos**.
- † **Mini Batch Size:** Tamaño de mini lote. Debido a que un *epoch* es muy grande para poder ingresarlo a la memoria de la computadora, se debe dividir en lotes pequeños para realizar los cálculos. Se tiene un tamaño de lote o *batch size* que será el número de ejemplos presentes en un lote sencillo.
- † **Iteraciones:** Se refiere al número de *batches* que se necesitan para poder completar un *epoch*.
- † **Initial Learn Rate:** Tasa de aprendizaje inicial, juega un valor fundamental en la precisión de la predicción ya que indica el monto de ajuste entre las conexiones internas de la red. Generalmente el valor es pequeño para poder aprender en las capas transferidas con lo cual se puede mantener un aprendizaje rápido solamente en las nuevas capas y lento en las transferidas en donde el valor de la tasa de aprendizaje es alta.

La red neuronal se alimenta con todos los *batches* o lotes, usando todos los lotes una sola vez es un *epoch*, de manera que si se tienen 10 *epochs* se usarán los datos en 10 ocasiones. Si se tienen 1,000 ejemplos de entrenamiento, y el tamaño de lote es de 500, entonces se ejecutarán 2 iteraciones para 1 *epoch*.

Debido a que la neurona aprende por el cambio de los pesos y sesgos, al estar estas determinadas por las derivadas parciales de la función de costo con respecto a estas cantidades ( $\frac{\partial C}{\partial \omega}$  y  $\frac{\partial C}{\partial b}$ ), cuando se aprende muy lento es debido a que estas derivadas parciales son muy chicas.

## 4.5. Aumento de Datos

Debido a que los modelos de *aprendizaje profundo* requieren de tener un número elevado de datos para que estos puedan ser entrenados y generalicen de mejor manera los resultados obtenidos en la red, es que se utiliza la técnica del aumento de datos (**data augmentation**), la cual permite que se tenga una mayor cantidad de ejemplos de manera sintética, esta es una técnica muy efectiva para problemas de reconocimiento de objetos.

En esta técnica se inflan artificialmente los datos en el conjunto de entrenamiento previamente etiquetado conservando las transformaciones ayudando así a mejorar el desempeño de los clasificadores

de aprendizaje profundo, sobre todo cuando se tienen bases de datos pequeñas en los cuales no se logra generalizar los datos para la validación ni la prueba [92-94].

El número de variaciones que se aplican a cada una de las imágenes queda a criterio de quien realiza el estudio en función de la aplicación para la cual se realiza, pudiéndose aplicar diferentes combinaciones entre sí, sin embargo es importante que las transformaciones no deben de afectar la clasificación correcta en la clase a la que pertenezca (ejemplo: 'b', 'q', 'd'). Cabe aclarar que existe una diferencia en comparación con el incremento de datos, que consiste en multiplicar los datos que se tienen por un cierto factor.

Al aplicar un aumento de datos, puede realizarse de dos formas: por transformaciones geométricas en donde se cambie de posición y orientación, la cual llega a ser más recomendable al mapear los valores individuales de los píxeles a nuevos destinos, o de manera fotométrica en el cual se trata de hacer la imagen invariante a la luz y al color [92], adicional a esto existe la **generación de redes adversas** en donde se eligen colores para alterar la imagen, pero se ha demostrado que no tiene un desempeño que supere a los aumentos tradicionales, además de consumir casi el triple de tiempo de cómputo [94].

Existe la posibilidad de que se realicen las transformaciones en un mini lote justo antes de incluirlo en el modelo de aprendizaje, sin embargo, lo más recomendable es que se realicen antes de que ingresen a la red para que se tenga de esta manera un aumento verdadero de los datos que se tienen [95].

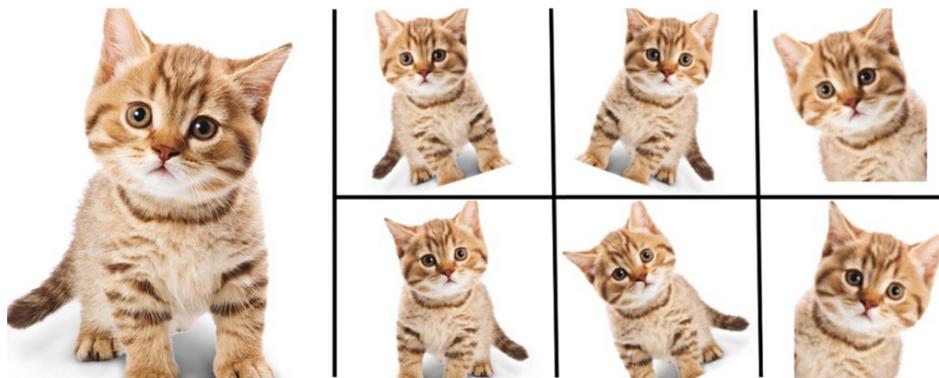


Figura 4.13: Representación de Aumento de Datos de una Imagen. De esta forma se realizan modificaciones en una imagen al aumentar de forma sintética los datos.

Tanto MATLAB, como varios de los programas en donde se utilizan las diferentes arquitecturas de aprendizaje profundo, permiten realizar diferentes variaciones a las imágenes que son:

- Rotación entre  $0^\circ$  y  $359^\circ$
- Reflexión en X
- Reflexión en Y
- Inclinación en X
- Inclinación en Y
- Traslación en X
- Traslación en Y

En el caso específico de MATLAB, se especifica en su documentación que con la instrucción `imageDatastore` se utilizan combinaciones de transformaciones de las imágenes para cada una de las iteraciones que se realicen del mini lote de imágenes durante el entrenamiento, sin embargo, según lo leído por diferentes opiniones en foros, es posiblemente más recomendable al momento de trabajar en MATLAB que se realicen las transformaciones geométricas de manera individual previo a ingresarlas a las redes ya que a diferencia de otros programas como Phyton, no se puede llegar a tener la seguridad de que aplique los cambios de manera aleatoria.

## 4.6. Métricas de Desempeño

Independientemente del tipo de algoritmo de aprendizaje automático que se vaya a elegir, es importante poder determinar las métricas de desempeño que se utilizarán para poder elegir un buen clasificador que pueda cumplir con la tarea que se quiera realizar para saber si este algoritmo es capaz de realizar esa labor o si es una mejor opción el cambiar de algoritmo.

Cuando se habla de un algoritmo es necesario determinar si en base al error de aprendizaje es que se puede determinar que se tiene una buena labor o si para tener un mejor desempeño se requiere un aumento de datos para que se reduzca el error; el error de Bayes define la mínima tasa de error que podría esperarse lograr al emplear un algoritmo [57]. Muchas de las ocasiones al querer resolver una pregunta de investigación se utiliza un punto de referencia fijo para que en función de los datos que se tienen, o que se podrán conseguir, para poder entrenarlo se pueda elegir el algoritmo que se va a utilizar.

La forma en la cual se puede determinar un nivel de desempeño que se puede esperar variará en función de los datos que han sido previamente publicados en la bibliografía si se va a enfocar a algo académico, o a lo que se ocupa en la realidad para que pueda ser seguro, efectivo y viable en costos tomando en cuenta la tasa de error que sea aceptable [57].

Por otra parte, la forma en la cual un clasificador indica cómo está desarrollando su labor se realiza mediante una **matriz de confusión**, ilustrada en la Figura 4.14 en la cual se colocan los datos de aquellos elementos a los que el clasificador predice que pertenecen a esa clase contra su condición verdadera dada en los datos etiquetados, en este caso pacientes, cada uno de estos cuadrantes tendrá los siguientes nombres [96, 97]:

- ♠ *Verdadero Positivo (TP)*: Aquellos pacientes que tienen la enfermedad y son clasificados como positivos por la prueba.
- ♠ *Falso Positivo (FP)*: El paciente no tiene la enfermedad, pero es clasificado por la prueba como que la tiene, también se considera a este un error tipo *I* o de falsa alarma.
- ♠ *Verdadero Negativo (TN)*: El paciente no tiene la enfermedad y el clasificador lo detecta como sano.
- ♠ *Falso Negativo (FN)*: El paciente tiene la enfermedad, pero la prueba lo cataloga como negativo; también considerado error tipo *II* o de falla al clasificar.
- ♠ *Positivos (P)*: Conjunto de personas que presentan una anomalía, integrado por la sumatoria de los verdaderos positivos y los falsos negativos.
- ♠ *Negativos (N)*: Conjunto de personas que pertenecen al grupo control, integrado por los verdaderos negativos y falsos positivos.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) <b>Type II Error</b>	<b>Sensitivity</b> $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) <b>Type I Error</b>	True Negative (TN)	<b>Specificity</b> $\frac{TN}{(TN + FP)}$
		<b>Precision</b> $\frac{TP}{(TP + FP)}$	<b>Negative Predictive Value</b> $\frac{TN}{(TN + FN)}$	<b>Accuracy</b> $\frac{TP + TN}{(TP + TN + FP + FN)}$

Figura 4.14: Matriz de Confusión. Los cuadrantes de la matriz de confusión y sus métricas se pueden apreciar en esta imagen; en ocasiones pueden intercambiarse de lugar las etiquetas de predicción y actual con lo cual los cuadrantes FN y FP se intercambian [98].

A partir de esto, se desarrollarán otros conceptos que sirven para evaluar la efectividad del clasificador, siendo algunos de estos más relevantes que otros al momento de saber su uso [96, 97], algunos de los conceptos más utilizados son:

- Sensibilidad:** También llamada por algunos como tasa de verdaderos positivos. Se refiere a la habilidad de la prueba para identificar a los pacientes con la enfermedad, por lo que es importante que este valor sea alto cuando se utilizan pruebas de tamizaje para identificar a pacientes enfermos. Se calcula de la siguiente manera:

$$Se = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{4.27}$$

- Especificidad:** Conocida también como tasa de verdaderos negativos. Es la habilidad para identificar correctamente a pacientes que no tengan la enfermedad, generalmente es utilizado para pruebas de diagnóstico y es utilizado como una segunda prueba. La forma de obtener este valor es:

$$Sp = \frac{TN}{TN + FP} = \frac{TN}{N} \tag{4.28}$$

- Valor predictivo positivo:** Indica la probabilidad de que se tenga la enfermedad en caso de que la prueba resulte positiva; la conocida además como precisión está dada por:

$$PPV = \frac{TP}{TP + FP} \tag{4.29}$$

- **Valor predictivo negativo:** Da la probabilidad de que el paciente no tenga la enfermedad cuando se da un resultado negativo en la prueba, se obtiene como:

$$NPV = \frac{TN}{TN + FN} \quad (4.30)$$

- **Tasa de falsos negativos:** Proporción de resultados positivos que son clasificados como negativos.

$$FNR = \frac{FN}{TP + FN} = \frac{FN}{P} \quad (4.31)$$

- **Tasa de falsos positivos:** Proporción de resultados negativos que son clasificados como positivos.

$$FPR = \frac{FP}{TN + FP} = \frac{FP}{N} \quad (4.32)$$

- **Prevalencia:** Indica qué tan seguido la condición positiva está presente en la muestra

$$Prevalencia = \frac{\sum Condición\ positiva}{P + N} \quad (4.33)$$

- **Tasa de descubrimiento falso:** Indica los casos que erróneamente se consideran positivos comparándolos contra todos los que el clasificador cataloga como positivos.

$$FDR = \frac{FP}{FP + TP} = 1 - PPV \quad (4.34)$$

- **Tasa de falsas omisiones:** Es la proporción que son erróneamente catalogados como negativos contra todos aquellos que el clasificador identifica como negativos.

$$FOR = \frac{FN}{FN + TN} = 1 - NPV \quad (4.35)$$

- **Exactitud:** Indica qué tan bien hace su labor el clasificador al reflejar la distribución de clase subyacente. Para que este tenga una mayor validez se recomienda que se utilice cuando se tienen el mismo número de elementos en cada una de las clases ya que sino puede tener valores diferentes por la forma en la que se calcula. Si bien es cierto que es una métrica importante, debe determinarse si es la ideal para un problema en específico.

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.36)$$

- **Medición F:** Conocido como *F1 score* o *F-measure* es la medida de la prueba de exactitud al considerar la media armónica de la precisión (PPV) y la sensibilidad de la prueba.

$$F_1 = \frac{2}{Se^{-1} + PPV^{-1}} = 2 \frac{Se * PPV}{Se + PPV} \quad (4.37)$$

- **Coefficiente de correlación de Matthews:** Medida de la calidad de clasificaciones binarias y que puede ser utilizada aunque las clases sean de diferentes tamaños. Es la media geométrica de los coeficientes de regresión del problema; toma en cuenta todos los datos de la matriz de confusión sin importar si el número de clases es balanceado, regresando un valor que oscila entre -1, que es un total desacuerdo entre la observación y la predicción, y 1 que es un clasificador perfecto; en caso de ser 0 es una clasificación debida totalmente al azar.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TN + FN)(P)(N)}} \quad (4.38)$$

- **Coefficiente Kappa de Cohen  $\kappa$ :** Indica qué tan bien se desempeña un clasificador en comparación con lo bien que se hubiese desempeñado por casualidad, esto es, la exactitud observada  $p_o$  y la exactitud esperada  $p_e$ ; entre mayor sea su valor indicará si existe una gran diferencia entre la precisión y la tasa de error nulo. Tiene un rango que va de  $-1$  a  $1$ , una de sus limitaciones es que no se ha determinado qué valor es el aceptable en investigaciones médicas.

$$p_o = \frac{TP + TN}{P + N} \quad (4.39)$$

$$p_e = \frac{P * \left(\frac{TP+FP}{P+N}\right) + TP * \left(\frac{FN+TN}{P+N}\right)}{P + N} \quad (4.40)$$

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4.41)$$

- **Exactitud balanceada:** Es el promedio aritmético de sensibilidad y especificidad; es mejor utilizarla cuando se tiene un conjunto de datos que tiene diferentes elementos en cada una de las clases, siendo así el promedio de las correcciones proporcionales de cada clase individualmente.

$$BalAcc = \frac{Se + Sp}{2} \quad (4.42)$$

- **Media Geométrica:** Tipo de media o promedio que indica la tendencia central o valor típico de un conjunto de números usando el producto de sus valores.

$$G_{mean} = \sqrt{Se * Sp} \quad (4.43)$$

## Capítulo 5

# Transferencia de Aprendizaje Automático en Termografía

El estudio realizado por Esteva *et al* [99] fue uno de los que más influencia tuvieron al hacer este trabajo, puesto que entrenó una red neuronal con imágenes clínicas que no sufrieron ninguna manipulación, comparando sus resultados con el de dermatólogos y su algoritmo fue capaz de clasificar cáncer de piel con un **nivel de competencia comparable al de un experto**, y pudo concluir que el uso de los dispositivos móviles pueden ampliar el alcance de los dermatólogos fuera de un ambiente clínico, sin embargo, hay diversos estudios que utilizan esta técnica.

La publicación de Nagendran *et al* [100] menciona algunos de ellos empleados en diferentes especialidades tales como la oftalmología, radiología, dermatología, histopatología, gastroenterología, ortopedia, cardiología y nefrología por mencionar algunos, aplicado a diferentes enfermedades como la degeneración macular, retinopatía diabética, cáncer colorrectal, cáncer de pulmón, arritmia, glaucoma, entre otros. Para que se pueda llevar a cabo el reconocimiento de imágenes en una red neuronal convolucional se deben de seguir los siguientes pasos:

- ✓ Cargar y explorar una imagen de datos
- ✓ Definir la arquitectura de la red
- ✓ Especificar las opciones de entrenamiento
- ✓ Entrenar la red
- ✓ Predecir las etiquetas de los nuevos datos
- ✓ Calcular la exactitud de la clasificación

Los datos fueron analizados mediante MATLAB, programa en el cual se recomienda el utilizar un almacén de datos de imágenes en donde los datos se encuentren recopilados en carpetas con el nombre de las respectivas categorías, las cuales serán las etiquetas de estos datos, pudiendo leer de manera eficiente los lotes de imágenes durante el entrenamiento de la red neuronal.

Al igual que Milosevic *et al* [8], la imagen se deja en escala de grises para poder enfocarse en las matrices de temperatura para la clasificación de termografías en normales y anormales; en este estudio, se extrajeron manualmente la región de interés pasándolo a escala de grises y recortaron individualmente los senos izquierdo y derecho utilizando tres diferentes clasificadores (*Support Vector Machine*, Naive Bayes y clasificador del vecino cercano k) evaluando el desempeño con una curva ROC, donde solamente tuvieron 40 imágenes (26 normales y 14 anómalos).

## 5.1. Primera Comparativa De Arquitecturas Pre-Entrenadas

Identificadas algunas de las diferentes arquitecturas que se tienen disponibles para realizar la transferencia de aprendizaje, se procedió a utilizarlas para la clasificación de termogramas de cáncer de mama.

El objetivo fue conocer la forma en la cual operan e identificar las diferencias que tienen tanto en tiempos de entrenamiento como en resultados que pueden arrojar con lo cual se pueda tener una primera aproximación en saber cuáles redes tienen un mejor desempeño y que posteriormente se puedan modificar los parámetros internos de las mismas; los menús de ayuda de MATLAB fueron un gran primer paso para realizar este paso.

La comparativa se realizó en las arquitecturas mencionadas en la sección 4.4 recalcando que las imágenes se reajustaron de manera automática en cada una de estas de manera automática. Inicialmente solo se consideró la base de datos de Silva *et al* [101] debido a que no se contaba aún con el resto de las bases de datos y esto sería una buena primera aproximación para nuestro problema.

Los termogramas fueron guardados en diferentes carpetas según la categoría a la que pertenecían (control o anómalo). Se le permitió al algoritmo que eligiera de manera aleatoria cuáles termogramas pertenecían al 70 % de los datos de entrenamiento y al 30 % de validación, así como aplicar de forma aleatoria el aumento de datos. En estos se utilizó un máximo de 4 epochs con una tasa de aprendizaje inicial de  $1 \times 10^{-4}$  y manteniendo la tasa de factor de aprendizaje tanto para sesgos como para pesos de 20.

Originalmente se trató de realizar el análisis en la laptop de uso personal, en donde algunas arquitecturas como *AlexNet* o ambas arquitecturas de *VGG* tardaron poco menos de 2 horas al realizar el análisis, otras tardaron más de un día en hacerlo, por lo que se realizó el análisis en la estación de trabajo HP Z820 con la que cuenta el LANCYTT, en donde el poder de cómputo es mayor gracias a la unidad gráfica de procesamiento (comúnmente conocida como GPU) NVIDIA Quadro K200 2GB VRAM.

El GPU es importante debido a que permite acelerar el procesamiento de imágenes con lo cual se explotan las capacidades del procesamiento paralelo al realizar operaciones de una forma más rápida debido a que tiene más núcleos de procesamiento siendo más rápida que la memoria RAM del CPU.

El cambio en la velocidad de cómputo fue drástico al hacer esta labor en un tiempo que oscilaba entre los 5 y los 20 minutos estando limitado solamente por el tamaño del mini lote que no podía tener más de 10 imágenes en cada uno.

Las métricas de desempeño descritas en la sección 4.6 fueron utilizadas para comparar el ejercicio que se tuvo en cada una de estas comparativas. Por las características propias tanto de la termografía como del estudio, se le dio una mayor importancia a los valores de sensibilidad en la prueba. La base de datos [101] cuenta una prevalencia mayor de personas pertenecientes al grupo de los controles que al de los anómalos, por lo que la exactitud balanceada fue un segundo punto a tomar en consideración.

Los resultados pueden observarse en la Tabla 5.1 los cuales fueron presentados en el congreso de *Photonics North 2018* celebrado en Montreal, Canadá, siendo publicados en [102]. Para presentar este trabajo se contó con el apoyo de “*Becas Nacionales CONACYT*”, así como de la *modalidad A* de la convocatoria 2018 de *Apoyos Especiales para Promover la Investigación, el Desarrollo Tecnológico y la Innovación del Consejo Potosino de Ciencia y Tecnología*, el programa 528 de “*Cátedras CONACYT*” y el proyecto 32 del “*Centro Mexicano de Innovación en Energía Solar*”.

Este primer experimento permitió obtener experiencia al momento de utilizar las redes pre-entrenadas, así como el poder identificar algunos errores que se hicieron al momento del análisis.

Red Pre-Entrenada	Se	Sp	AUC	PPV	NPV	BalAcc	G mean
VGG 16	100 %	82.35 %	0.8452	10 %	100 %	91.18 %	90.7 %
ResNet 101	67.67 %	83.67 %	0.8881	20 %	97.62 %	75.17 %	74.7 %
AlexNet	58.33 %	92.68 %	0.8744	70 %	88.37 %	75.51 %	73.5 %
GoogLeNet	50 %	82 %	0.7738	10 %	97.62 %	66 %	64 %
VGG 19	32.14 %	95.83 %	0.8786	90 %	54.76 %	64 %	55.5 %
ResNet 50	0 %	80.39 %	0.7143	0 %	97.62 %	40.20 %	0 %
Inception V3	NA	80.77 %	0.7328	0 %	100 %	NA	NA

Tabla 5.1: Resultados Primer Comparativa de Redes Pre-Entrenadas. Las métricas de desempeño de la primera comparativa de resultados obtenidos es mostrada en la tabla; el trabajo fue presentado en *Photonics North* [102].

Se descubrió que es fundamental el hacer la segmentación de los datos que integrarán a los conjuntos de entrenamiento y validación desde antes de ingresarlos al algoritmo para poder tener una certeza de que se analicen los mismos datos, siendo los termogramas anómalos los que se tienen en menor cantidad, que estos sean los mismos que integren tanto al conjunto de validación como al de entrenamiento, lo cual pudo ser una razón que justifique el por qué se tenía una variación tan amplia en los resultados arrojados por la métrica de desempeño entre las diversas arquitecturas.

A su vez, es importante poner los valores bajo los cuales se realizará el aumento de datos para que, si bien es cierto que las transformaciones geométricas que se apliquen serán aleatorias, se le coloquen límites dentro de los que oscilará.

También se reconoce que la limitante del número de pacientes con los cuales la prueba fue realizada, es otra causa que pudo influir en el desempeño de la misma, por lo que se refuerza el hecho de hacer desde antes la segmentación de los termogramas que conformarán tanto al entrenamiento como a la validación.

Hasta donde se tenía conocimiento, al momento de presentar los resultados de este trabajo en el congreso, era la primera vez que se utilizaba la técnica de transferencia de aprendizaje aplicada a la termografía en cáncer de mama, siendo así una forma de hacer esta evaluación de manera automática y eliminando la variabilidad debido al error humano.

Esta primera aproximación que se tuvo al momento de comparar los resultados de la Tabla 5.1 lucían promisorios al sugerir que la técnica podría ser utilizada como un método que podría identificar anomalías térmicas y podría ser utilizada de manera adjunta a la mamografía para la identificación temprana de un posible cáncer de mama.

## 5.2. Base de Datos de Termogramas

Se contó con el análisis de 3 bases de datos, una fue pública adquirida en Brasil por Silva *et al* [101], la cual contiene 219 termogramas, una más fue obtenida por el Dr. Antonio Ocegüera-Villanueva y el Dr. Jorge Luis Flores-Nuñez en el Instituto Jalisciense de Cancerología (IJC) contando con 101 termogramas, mientras que 170 fueron adquiridos en el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE) en un estudio encabezado por el Dr. Samuel Kolosovas. Las aprobaciones de los respectivos comités de ética fueron obtenidas en la adquisición de los termogramas en cada uno de los centros.

Del total de los 490 termogramas, tuvieron que ser excluidos 179 imágenes, esto debido a que no cumplían con algunos criterios, tales como diagnóstico desconocido, termograma fuera de foco, si el

paciente tenía cirugía plástica o prótesis de seno; la cantidad de termogramas que fueron descartados por cada una de estas causas pueden apreciarse en la Figura 5.1 indicándose además la edad y la desviación estándar de los mismos.

Para cumplir con lo estipulado en las redes neuronales para establecer los conjuntos de prueba y entrenamiento del algoritmo, se tomó el criterio de tener en cada una de las 3 bases de datos la norma del 70-30 en sendas categorías, esto es, que el 70 % de las imágenes de las pacientes clasificadas como anómalas fue utilizado en el entrenamiento de la arquitectura, mientras que el 30 % fue utilizado en la validación, aplicando este mismo criterio para las imágenes de las pacientes que integran al grupo control.

En la Tabla 5.2 se ve con detalle la composición de cada una de las bases de datos, así como el total analizado tanto para el grupo control, como para el anómalo, considerando el número que imágenes que integran al entrenamiento y la validación para cada una de estas, así como la edad, desviación estándar y los totales para cada categoría.

Composición Base Datos			Brasil	IJC	ISSSTE	Total
Control	Entrenamiento	n (70 %)	103	26	66	195
		Edad	57	51	56	56
		SD ( $\pm$ años)	11.5	9.2	9.9	10.9
	Validación	n (30 %)	38	18	16	72
		Edad	57	46	53	54
		SD ( $\pm$ años)	13.02	12.21	8.47	12.77
Anómalos	Entrenamiento	n (70 %)	24	5	2	31
		Edad	53	40	50	51
		SD ( $\pm$ años)	12.19	9.78	1.0	12.29
	Validación	n (30 %)	8	4	1	13
		Edad	62	39	56	54.46
		SD ( $\pm$ años)	13.6	13.32	0.0	16.65
Total	Entrenamiento	n (70 %)	127	31	68	226
		Edad	56	49	56	55
		SD ( $\pm$ años)	11.73	9.93	9.86	11.21
	Validación	n (30 %)	46	22	17	85
		Edad	58	45	54	54
		SD ( $\pm$ años)	13.2	12.7	8.2	13.5
Total	n	173	53	85	311	
	Edad	57	47	55	55	
	SD ( $\pm$ años)	12.1	11.4	9.6	11.9	

Tabla 5.2: Distribución de pacientes en cada base de datos. Se puede apreciar el número de pacientes que integran a cada grupo de sendas bases de datos para analizarse en las redes neuronales convolucionales.

La base de datos reportada por Silva *et al* [101] tiene datos de los pacientes tales como la edad en la cual se tomaron la termografía, primera y última menstruación, color de piel, hábitos alimenticios, parte de la historia médica, síntomas, así como pruebas con las cuales se determina si el paciente es sano o enfermo, ya sea con la aplicación de mamografía, biopsia o ambas en algunos casos.

Por otra parte, en las otras dos bases de datos mencionadas, se obtuvo la clasificación de estos en función de las lesiones mamarias según el BI-RADS, con lo que las lesiones 1 y 2 son clasificadas en el grupo control, mientras que del 3 al 5 se consideran anómalas, catalogando como desconocidos a los que sus resultados sean 0.

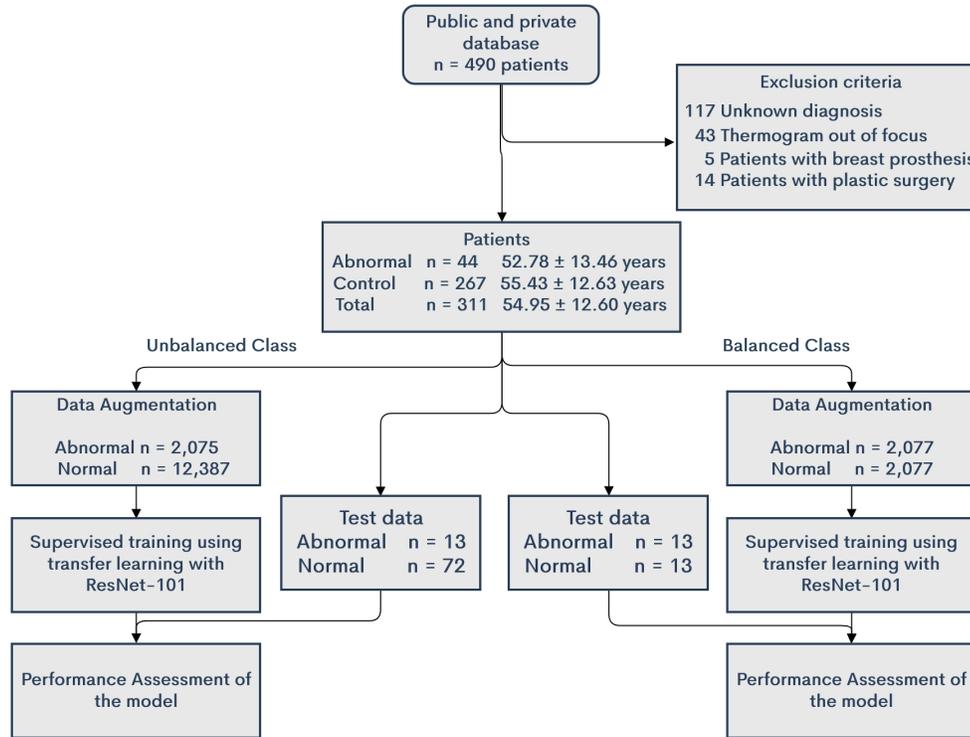


Figura 5.1: Diagrama de Flujo Análisis de Termogramas. Es posible apreciar la representación de una red neuronal, la cual está conformada por nodos y conectores de peso.

Aunque las bases de datos tenían diferentes vistas, para este estudio solamente fueron utilizadas las vistas frontales en donde podían verse ambas mamas. Todos los termogramas fueron ordenados asignándoles un número de paciente y separándolos en las carpetas de entrenamiento y validación en donde fueron clasificados como controles o anómalos según correspondía.

La Figura 5.2 muestra de forma gráfica la distribución de la Tabla 5.2 en función de cada uno de los centros en donde los datos fueron adquiridos, así como los totales para cada conjunto, ya sea de entrenamiento o validación para cada categoría.

Las imágenes fueron recibidas en un archivo .jpg, por lo que en MATLAB con la ayuda del paquete proporcionado por FLIR se convirtieron las imágenes a un archivo .tif en donde se almacenaron los valores de las matrices de temperatura y se guardaron en escala de grises para el análisis.

Cámaras FLIR	Brasil	IJC	ISSSTE
Modelo	SC620	T600	T400
Dimensión (píxeles)	640 x 480	480 x 360	320 x 240
Sensibilidad (°C)	0.04	0.04	0.045

Tabla 5.3: Datos de las Cámaras FLIR Utilizadas.

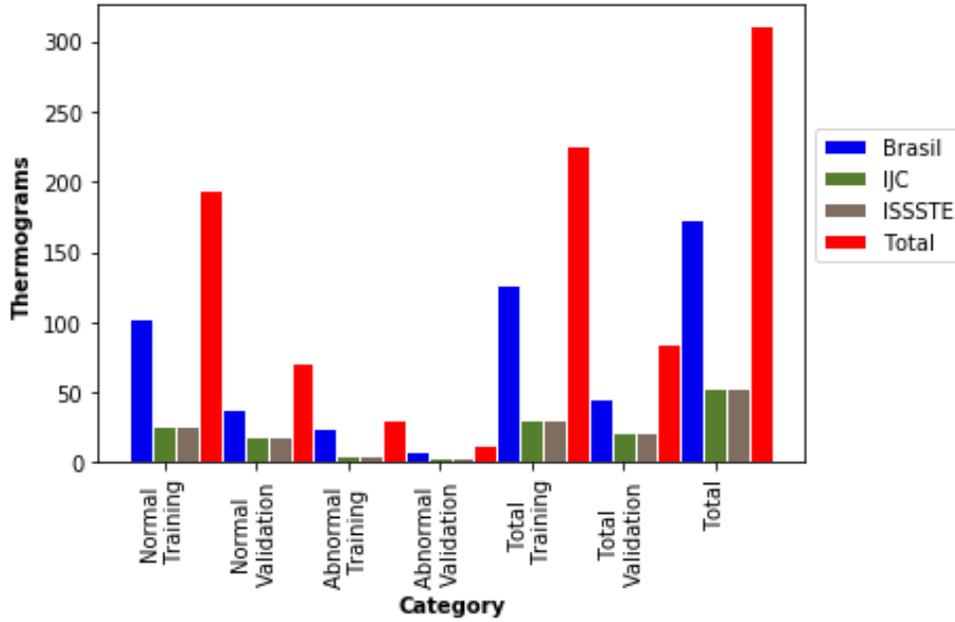


Figura 5.2: Bases de Datos Termogramas. Se aprecia la forma en la cual se distribuye la base de datos por centro y su conformación total.

De esta forma se evitan interpretaciones subjetivas debido al uso de paletas de colores y el rango en el cual se presente la matriz de temperatura de la paciente. La Tabla 5.3 muestra las diferentes cámaras FLIR que fueron utilizadas en la adquisición de los termogramas en las diferentes bases de datos.

### 5.2.1. Análisis Estadístico de Normalidad

Dadas las bases de datos, se requiere saber si la población tiene o no una distribución normal, por lo cual se realiza una **prueba no paramétrica** debido a que proviene de una distribución libre [103]. Una distribución normal es aquella en la cual la variable aleatoria tiene una distribución en forma de campana con un área bajo la curva de probabilidad igual a 1 que depende de dos parámetros: su media y la desviación estándar [104].

Para saber si los datos que se analizan son normales, se aplica la prueba de **Shapiro-Wilk** en caso de que la muestra aleatoria sea menor de 50 datos y en caso de ser mayor se usará la **prueba de Kolmogorov-Smirnov** [104, 105].

En caso de que no se tenga una distribución normal, por lo cual no se puedan aplicar los intervalos de confianza y contrastes en la *t de Student*, se utilizan pruebas como la **prueba de Wilcoxon de los rangos con signo**, que permite comparar el rango medio de dos muestras relacionadas para determinar si existen diferencias entre ellas que sean debido al azar, así como la **prueba de suma de rangos de Wilcoxon**, también llamada **prueba U de Mann-Whitney** para saber si hay diferencia entre las muestras o si vienen de la misma población [105].

El **valor p** se refiere al nivel de significancia más bajo en donde es significativo el valor observado del estadístico de prueba tomando normalmente el valor de 0.05 ó 0.01 como referencia [104]; es utilizado

como evidencia para probar una hipótesis con lo que puede determinar qué tan fuerte es una evidencia y evita cometer el error tipo I al rechazar la hipótesis nula cuando esta es verdadera; su valor oscila entre 0 y 1, entre menor sea el *valor p*, más fuerte será la evidencia en contra de la hipótesis nula y se dirá que es estadísticamente significativo (importante) [103].

Para realizarlo se recomienda que primero se establezca la hipótesis nula y alternativa (dependiendo del caso es a veces más fácil recolectar evidencia para rechazar una hipótesis que para aceptarla debido a que se requiere un número limitado de evidencias), elegir un nivel de significancia fijo para posteriormente seleccionar un estadístico de prueba adecuado para rechazar la hipótesis nula si el estadístico de prueba está en la región crítica o aceptarla [104].

La Tabla 5.2 muestra los valores de la media y la desviación estándar que tiene el conjunto de datos, mientras que tras realizar el análisis de Kolmogorov-Smirnov arroja un valor de 0.499, pero es diferente al coeficiente indicado en tablas que es de 0.05068 con un *valor p* de 0.05 el cual es deseado por lo cual, se debe rechazar la hipótesis nula indicando que no se sigue una distribución normal.

Ahora se debe realizar la prueba no paramétrica de suma de rangos de Wilcoxon (U de Mann-Whitney) para comparar dos poblaciones, se parte de la hipótesis nula en la cual ambas medianas son iguales para de esta manera considerar que se puede tratar como si tuvieran distribuciones normales. Al realizarla se obtiene un *valor p* de 0.06724 que es mayor a 0.05 concluyendo que la diferencia entre las medianas no es estadísticamente significativa por lo que no es posible rechazar la hipótesis nula.

### 5.2.2. Generación de Datos Sintéticos

Al combinar las tres bases de datos obtenidas de forma individual se espera un conjunto que pueda dar mayor diversidad al momento de ser analizada. Tomando como referencia el estudio realizado por Becket *et al* [106], se realizaron **dos enfoques**, al considerar clases balanceadas y desbalanceadas.

Para la **clase balanceada** el objetivo es tener el mismo número de elementos que integren a cada una de las clases para que de esta manera se tenga la misma probabilidad de clasificar a cada imagen dentro de cada una de estas. Por otra parte, el tener **clases desbalanceadas** permite que se apegue más a lo que se encontrará en la realidad debido a que al hacer una prueba en la población, son pocas las personas que presentan una anomalía.

La segmentación de termogramas para cada uno de los enfoques fue realizada como se ilustra en la Figura 5.1, para el entrenamiento, la clase desbalanceada contó con 31 termogramas anómalos y 195 controles, mientras que para la clase balanceada quedó determinada por 31 termogramas en cada una de las clases.

Debido a que se quería tener un control sobre las transformaciones geométricas para aplicar a las imágenes, se decidió prescindir del comando que viene integrado en MATLAB prefiriendo hacer transformaciones que consistieron en rotaciones entre  $0^\circ$  y  $359^\circ$ , reflexión en ambos ejes, traslación entre 0 y 50 píxeles en ambos ejes, que fueron descritas con mayor detalle en la sección 4.5 y combinaciones aleatorias entre estas.

El resultado de esto fue la generación de 67 imágenes más en el conjunto balanceado conformando un total de 2,077 termogramas por clase mientras que para el conjunto desbalanceado al ser aleatorio, se tuvo un promedio de 67 copias de termogramas anómalos produciendo 2,075 termogramas, mientras que para el grupo control fue un total de 12,387 como resultado de aumentarlo en promedio 63.5 mostrándose esto también en la Figura 5.1.

Por otra parte, en el conjunto de prueba tanto de clases balanceadas como desbalanceadas, los termogramas no fueron modificados, siendo analizados tal cual como fueron adquiridos tras haberles

convertido a formato .tif; para el enfoque de clases desbalanceadas se tuvieron 13 pacientes con anomalía y 72 controles, mientras que para el de clases balanceadas fueron 13 termogramas las que integraron al enfoque balanceado. Las imágenes fueron guardadas en carpetas correspondientes a sus clasificaciones tanto para entrenamiento como para validación en sendos enfoques.

### 5.3. Selección de Mejor Arquitectura Pre-Entrenada

Tras conocer la forma en la cual opera una red neuronal en donde se pudieron identificar algunos errores, se procede a realizar otro análisis utilizando variaciones en las arquitecturas en el cual se utilizará la base de datos multicentro obtenida con las transformaciones geométricas descritas en [5.2.2](#) que permitieron hacer un aumento de datos sintético.

Debido a que las arquitecturas de aprendizaje automático son algoritmos iterativos en los cuales se modifican sus pesos y sesgos, un análisis cíclico fue realizado para determinar aquella iteración entre las redes que brinde un resultado óptimo.

Originalmente se realizaron algunas pruebas con 5 iteraciones para saber las diferencias que se tenían al modificar el tamaño del mini lote de 128, 100, 68, 32, 24 y 12 pero esto impedía que se corrieran los análisis en la memoria GPU de la computadora, obligando a utilizar la memoria RAM.

El tiempo de análisis de cada iteración de las arquitecturas se incrementó demorando para las de mayor lote 29 horas sin lograr terminar de hacer los cálculos completos, distando mucho del promedio de 30 minutos que tardaba cuando se usaba un tamaño de mini lote de 10 que permitía el uso del GPU. Realizando la comparación de los pocos resultados que se lograron obtener, se encontró que no era un factor decisivo en las métricas de desempeño de la arquitectura cuando se tenía un mini lote de 10 pero sí en el tiempo de análisis.

Determinando así el tamaño del mini lote que se utilizará para realizar las correcciones implementadas en las arquitecturas, se realizará una prueba entre ellas con la base de datos multicentro para conocer aquella que pueda dar un mejor resultado.

Para este análisis, los parámetros internos de la red se mantuvieron sin cambio con respecto al análisis previo ya que solo se quiere saber la arquitectura que pueda dar un mejor resultado y en el cual solo se deba realizar un ajuste fino. Las arquitecturas descritas cuentan con un límite de iteraciones que detienen el análisis de forma automática en caso de que se tenga una tendencia al sobreajuste en la exactitud.

De esta forma, en cada arquitectura se realizó el código estableciendo ciclos de 50 corridas guardando los resultados de las métricas de desempeño en una hoja de cálculo, permitiendo que se pudieran comparar los valores de cada corrida para determinar aquella que sea más conveniente para que se realice el ajuste fino.

La sensibilidad fue el valor con mayor relevancia para elegir la mejor arquitectura debido al enfoque del método cuya función es ser una herramienta en la cual se pueda identificar a personas que tengan una anomalía, dejando en segundo término la especificidad, seguido de la exactitud balanceada y por último el área bajo la curva ROC.

La Tabla [5.4](#) muestra los resultados obtenidos para el análisis de clases desbalanceadas, es de resaltar que la red *ResNet 101* aparece en par de ocasiones como uno de los mejores 4 resultados arrojados por las iteraciones y que en par de ocasiones al hacerlo con *ResNet 101* se obtuvieron los mismos valores en la métrica de desempeño.

Arquitectura	Se	Sp	AUC	PPV	NPV	Bal Acc	G mean
ResNet 101	92.3 %	25.0 %	0.697	18.1 %	94.7 %	58.6 %	48 %
VGG 19	92.3 %	25.0 %	0.481	16.9 %	90.0 %	54.8 %	53.7 %
GoogLeNet	76.9 %	37.5 %	0.511	18.2 %	90.0 %	57.2 %	53.7 %
ResNet 101	76.9 %	55.5 %	0.760	23.8 %	93.0 %	66.2 %	65.3 %

Tabla 5.4: Métricas de Desempeño de Diferentes Arquitecturas con Clases Desbalanceadas. Se muestran los mejores resultados en las diversas corridas de las arquitecturas tras haber aplicado el aumento de datos sintéticos con las transformaciones geométricas.

Arquitectura	Se	Sp	AUC	PPV	NPV	Bal Acc	G mean
ResNet 101	92.3 %	23.1 %	0.579	54.5 %	75.0 %	57.7 %	46.1 %
VGG 19	92.3 %	23.0 %	0.508	54.5 %	75.0 %	57.7 %	46.1 %
AlexNet	92.3 %	15.4 %	0.633	52.2 %	66.6 %	53.8 %	37.6 %
ResNet 50	92.3 %	23.1 %	0.373	54.4 %	75.0 %	57.7 %	46.1 %

Tabla 5.5: Métricas de Desempeño de Diferentes Arquitecturas con Clases Balanceadas. Se muestran los mejores resultados en las diversas corridas de las arquitecturas tras haber aplicado el aumento de datos sintéticos con las transformaciones geométricas.

También se aprecia que los valores arrojados por *VGG 19* son los mismos en sensibilidad y especificidad pero llegan a variar en el área bajo la curva y exactitud balanceada.

Por su parte, la Tabla 5.5 muestra los valores de las iteraciones hechas al analizar clases balanceadas teniendo resultados muy semejantes entre sí y equiparables con los obtenidos al tener un enfoque desbalanceado.

En ambos casos se aprecia que la red con mejores resultados fue *ResNet 101*, por lo que será en esta arquitectura en la cual nos enfocaremos en realizar el ajuste fino a los parámetros internos de la red.

## 5.4. Búsqueda de Parámetros Óptimos para la Red

*Deep learning* involucra optimización en demasiados contextos y es el problema más difícil al momento de realizar un entrenamiento de una red neuronal. En esto destaca el poder encontrar los parámetros de una red que puedan reducir la función de costo. Debido a que la optimización es una tarea extremadamente difícil, normalmente se trata de diseñar cuidadosamente la función objetivo. Generalmente se debe de elegir un modelo que sea fácil de optimizar que usar un algoritmo poderoso para optimización [57].

Desafortunadamente no existe un consenso acerca de la forma que sea ideal para poder elegir la optimización de un algoritmo, así como de las tasas de aprendizaje, aunque en este caso se sugiere un algoritmo que sea capaz de tener tasas de aprendizaje capaces de adaptarse tal como el gradiente estocástico descendiente con momento, dependiendo esto en mucho de la familiaridad que se tenga con el algoritmo para poder realizar un mejor ajuste de hiperparámetros [57].

Generalmente se puede seguir un procedimiento llamado en inglés *grid search*, otro que es el *random search* o búsqueda aleatoria de parámetros o el modelo de obtimización Bayesiana, sin embargo este último no se recomienda usar ya que sus resultados son muy variables y no siempre puede dar el mejor [57].

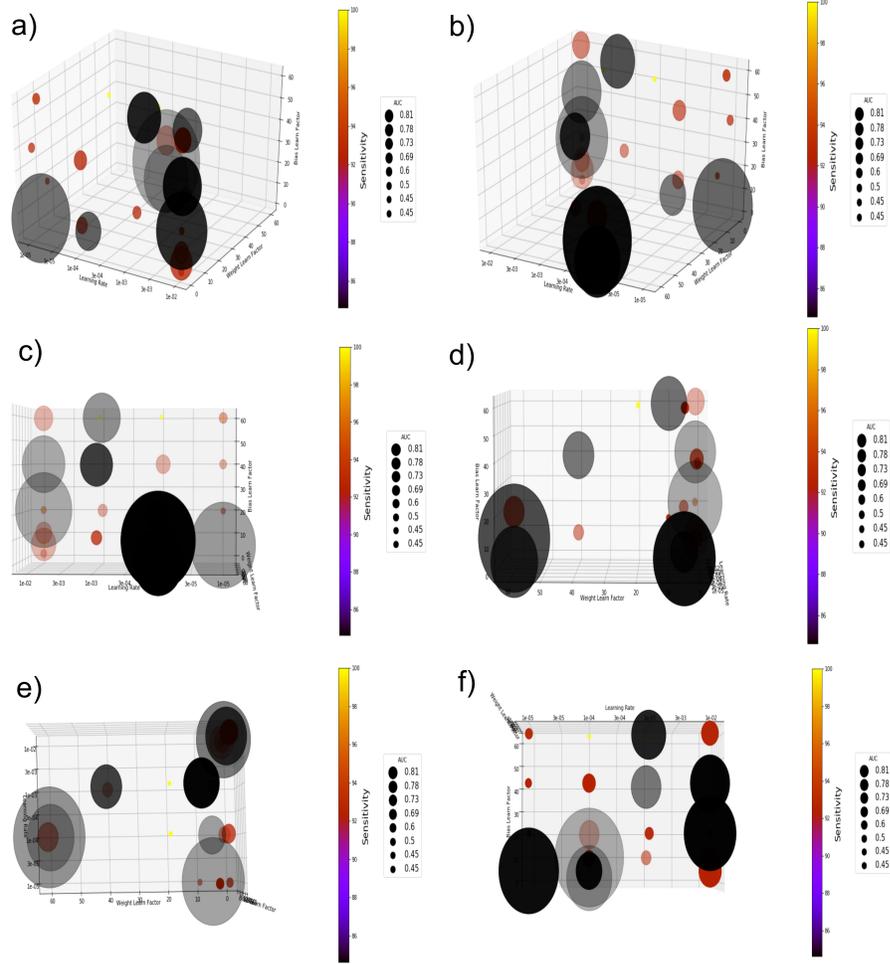


Figura 5.3: Grid Search Balanceado. Resultados de las iteraciones para buscar los mejores parámetros internos en el enfoque balanceado.

El elegido para este trabajo fue *grid search*, el cual se refiere a la búsqueda de los parámetros óptimos de la red en los cuales se tengan mejores resultados en una cierta combinación de variables. En el presente estudio nos enfocamos en las combinaciones de 3 variables, cada una de ellas modificada de manera sistemática realizando 10 iteraciones en la arquitectura *ResNet 101* para un total de 1800 iteraciones para determinar el ajuste fino, siendo estas:

☆ Tasa de aprendizaje inicial:

- $1e^{-2}$
- $1e^{-3}$
- $1e^{-4}$
- $1e^{-5}$

☆ Tasa de factor de aprendizaje de pesos:

- 1
- 5
- 10
- 20
- 40
- 60

☆ Tasa de factor de aprendizaje de sesgos:

- 1
- 5
- 10
- 20
- 40
- 60

La selección de la tasa de aprendizaje que pueda ser mejor para el algoritmo, se considera más un arte que una ciencia [57], por lo cual debe elegirse a prueba y error, así como monitoreando las curvas de aprendizaje que grafican la función objetivo como una función del tiempo.

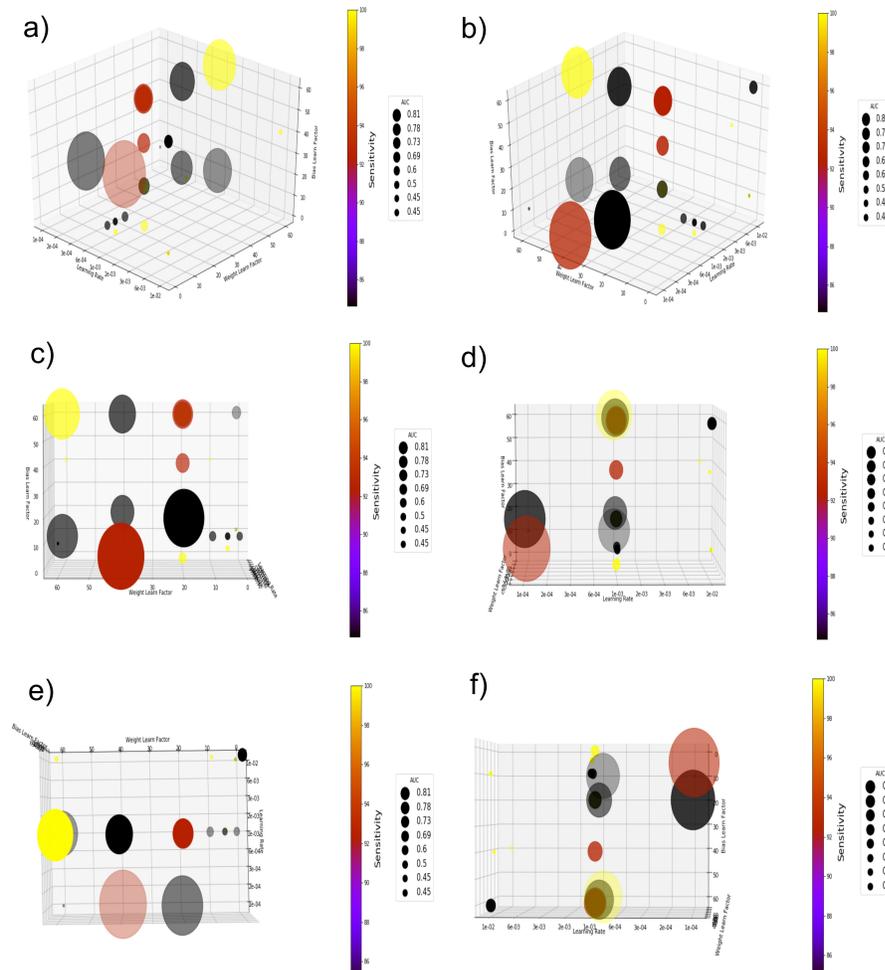


Figura 5.4: Grid Search Desbalanceado. Resultados de las iteraciones para buscar los mejores parámetros internos en el enfoque desbalanceado.

Debido al uso que se le quiere dar a la termografía infrarroja, solamente se conservaron aquellas iteraciones en las cuales fue obtenida una sensibilidad igual o mayor a 80%; diversas iteraciones cumplían con los parámetros dados, por lo cual fueron ordenados con el número de la iteración que correspondía para ser distinguidos y, al igual que en el análisis para determinar la arquitectura que pudiera darnos los mejores resultados, se priorizó la sensibilidad, seguido de la especificidad, área bajo la curva y exactitud balanceada.

Para poder determinar también de forma gráfica la región en la cual se tienen los mejores resultados, se hizo un gráfico tridimensional especificando cada una de estas variables en donde cada uno de los puntos tiene un color diferente en función de la sensibilidad iniciando con el valor de 80%, así como su tamaño que indica el valor del área bajo la curva.

La Figura 5.3 muestra diferentes vistas del análisis gráfico para el enfoque balanceado, es posible apreciar que existen regiones en donde el área bajo la curva será mayor, pero esto no representa que la sensibilidad, que es el parámetro al cual le hemos dado una mayor importancia, corresponda a este mismo.

Se puede apreciar que cuando se tiene una tasa de factor de aprendizaje de pesos que van entre 0 y 10 es cuando arroja valores de sensibilidad altos sin llegar a considerarse un clasificador ‘tonto’ al tener por azar un 100%, para la tasa de factor de aprendizaje de sesgos existe una mayor cantidad entre 0 y 20, siendo lo más complicado en determinar el valor óptimo para la tasa de aprendizaje, por lo que para la decisión fue necesario apoyarse en los valores de la especificidad, resultando en la tasa de aprendizaje de  $1e^{-5}$ .

La Figura 5.4 hace lo propio con el enfoque desbalanceado el cual fue un poco menos complicado que para el anterior. Es fácil observar que la tasa de aprendizaje con los mejores resultados se apilan cuando se tiene un valor de  $1e^{-3}$ , así como los valores de la tasa de aprendizaje de los pesos que es claro verlo cuando se deja un valor de 10, mientras que los valores de tasa de factor de aprendizaje de sesgos tiene buenos resultados entre 10 y 20 siendo este el valor más complicado para determinar.

Es un poco más sencillo poder interpretar los resultados y argumentos previos si analizamos la Tabla 5.6 en ésta se presentan los mejores resultados obtenidos tras buscar los parámetros óptimos en la arquitectura de *ResNet 101* con cada uno de los enfoques.

Enfoque	Se	Sp	AUC	PPV	NPV	Bal Acc	G mean	LR	Weight	Bias
Balanceado	92.3 %	53.8 %	0.722	66.7 %	87.5 %	73.1 %	70.5 %	$1e^{-5}$	10	20
Desbalanceado	84.6 %	65.3 %	0.814	30.6 %	95.9 %	74.9 %	74.3 %	$1e^{-3}$	5	10

Tabla 5.6: Resultados Parámetros Óptimos. Se muestran los mejores valores tras la búsqueda de los mejores parámetros en el ajuste fino. LR = Tasa de aprendizaje inicial, Weight = Tasa de factor de aprendizaje de pesos, Bias = Tasa de factor de aprendizaje de sesgos.

La curva ROC de ambos enfoques es mostrada en el panel a) de la Figura 5.5 siendo el color negro la correspondiente al enfoque con clases balanceadas y con verde desbalanceadas con sus respectivos valores de área bajo la curva en la simbología. Con sendos puntos se representan a su vez los valores que se obtuvieron para la sensibilidad y especificidad en el estudio realizado.

De igual forma, en los paneles b) y c) se muestran las matrices de confusión con el número de termogramas que fueron clasificados en esa categoría sobre el total utilizado en el conjunto de prueba, así como los porcentajes que se tienen en cada cuadrante y una barra de color que indica el máximo y mínimo de imágenes que pertenecen a los mismos.

En el panel a) de la Figura 5.5 se pueden apreciar otros símbolos en donde se especifican otros estudios que han utilizado termografía para el tamizaje de cáncer de mama con los respectivos valores

de sensibilidad y especificidad. Cada uno de estos estudios hacen uso de diferentes formas para realizar la clasificación.

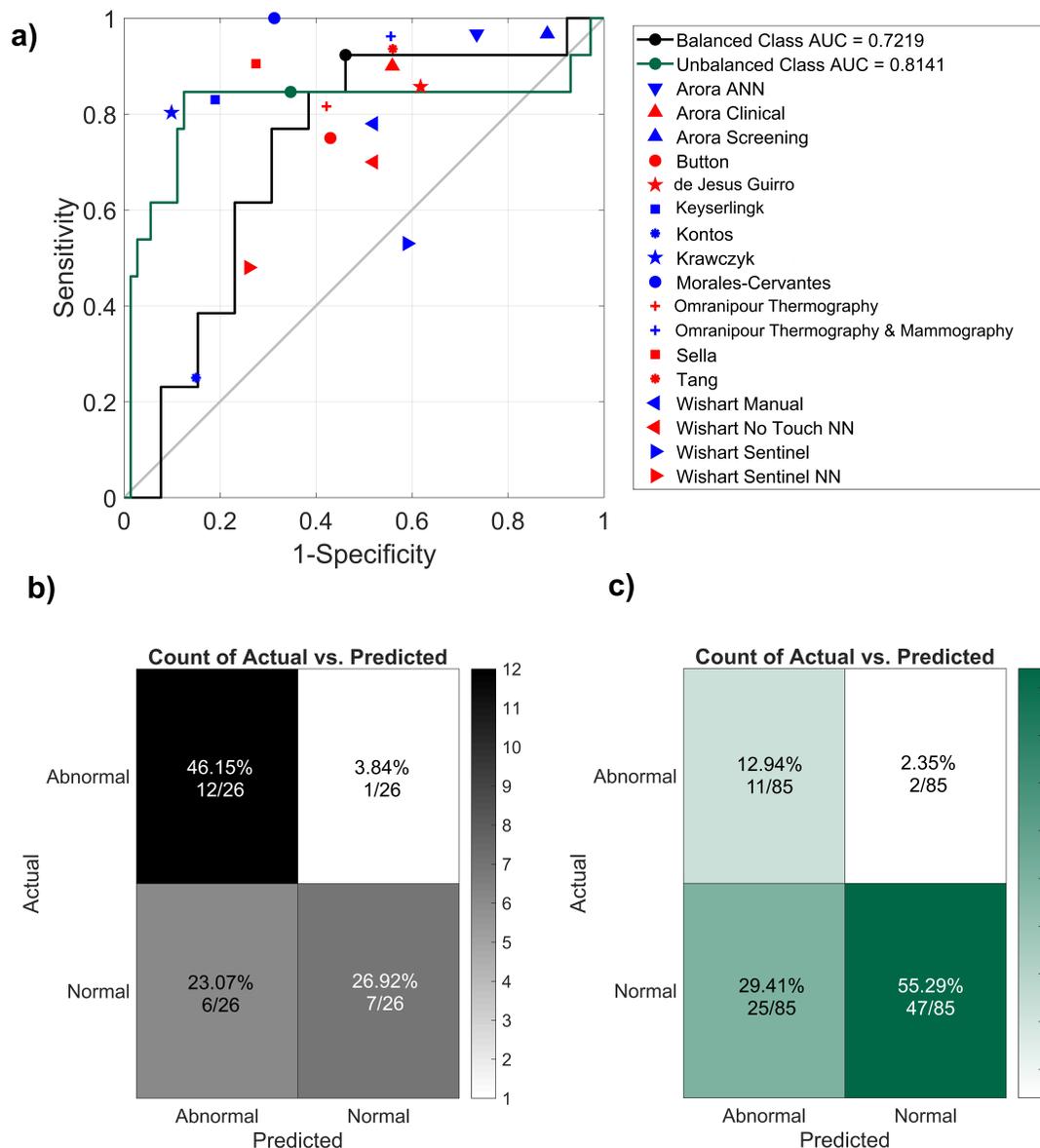


Figura 5.5: Curva ROC y Matrices de Confusión. Comparación de los resultados obtenidos contra otros estudios donde se utiliza la termografía como método de tamizaje.

El estudio de Omranipour *et al* [107] es probablemente uno de los más relevantes por el enfoque que tiene, ya que compara los valores obtenidos por la mamografía, termografía (mostrado en la Figura 5.5.a) con 81.6 % de sensibilidad y 57.8 % de especificidad), el uso de termografía con la prueba en frío y la combinación de termografía con mamografía, lo cual es una de las ideas que se tendría para utilizar, debido al incremento que se tiene en la sensibilidad (96.2 %) y especificidad (44.4 %) mostrando este último también en el gráfico de la comparativa.

Arora *et al* [51] realizó la prueba con 92 pacientes que fueron sometidos a estrés frío, al tener aire

directamente soplando sobre los senos y apoyados por un software llamado *Sentinel BreastScan* dan tres respuestas combinando el riesgo que tienen de tener la enfermedad (Arora Screening) con 96.7% de sensibilidad y 11.8% en especificidad, la ubicación de la lesión basada en las imágenes adquiridas previamente con una sensibilidad de 90% y especificidad de 44.1% (Arora Clinical) y por último, utilizando redes neuronales la evaluación de si en lo analizado se tienen hallazgos positivos o negativos, siendo esta última la que obtuvo mejor resultado con una sensibilidad de 96.7% y una especificidad de 26.5% (Arora ANN).

Button *et al* [108] utiliza termografía dinámica obteniendo sensibilidad de 75% y 43% de especificidad, mientras que de Jesús Guirro *et al* [109] utilizó el software *ThermaCAM Research Professional* con la ayuda de par de expertos en termografía obtuvo una sensibilidad de 85.7% y 38.2% especificidad.

Keyserlingk *et al* [45] utiliza un software la escala de Ville Marie para la evaluación de termogramas logrando una sensibilidad de 83% y 19% de especificidad y reafirma la idea que estos al ser interpretados por especialistas pueden dar información objetiva. Kontos *et al* [52] es quien presenta el desempeño más pobre en sensibilidad logrando solamente el 25% a cambio de 85% de especificidad al evaluar a 63 pacientes ayudado de par de expertos.

Es interesante el método de Krawczyk y Schaefer [110] que extrae características que describen simetrías bilaterales para posteriormente usar un clasificador que pone todas las características en conjunto para tomar una decisión tras aplicar diversas técnicas tales como *support vector machine*, selección de características, ajuste de pesos de acuerdo a la exactitud individual de cada clasificador, entre otros, logrando una sensibilidad de 80.35% y 90.15% de especificidad con una exactitud de 90.03% como uno de sus mejores resultados.

La metodología usada por Morales-Cervantes *et al* [111], si bien es cierto que da un 100% de sensibilidad, también es de resaltar que un método que de el 100% es posible que se considere como un clasificador ‘tonto’, además que la paleta de colores que se elija para los termogramas puede influir en la decisión que tenga para seleccionar la vascularización; su especificidad es del 68.68%.

Por su parte, Sella *et al* [112] utilizó imágenes infrarrojas en 3D permitiendo observar la vascularización con mejor detalle y siendo clasificada mediante *Support Vector Machine* con 90.5% de sensibilidad y 72.5% de especificidad, mientras que Tang *et al* [50] tratando de enfocarse en la vascularización en función de qué tanto brillan los píxeles en la imagen y determinando la amplitud de estos, logrando valores de 93.6% y 44% en sensibilidad y especificidad respectivamente.

Por último, Wishart *et al* [113] muestra 4 diferentes enfoques en su estudio, por un lado, *Sentinel BreastScan* mide 7 diferentes parámetros de temperatura, de los cuales 4 son de comparativas de puntos contralaterales de los senos, el *Sentinel BreastScan Neural Network* utiliza redes neuronales en los cuales arroja solamente si hay hallazgos positivos o negativos, a su vez evalúa los termogramas por un experto de forma manual.

Finalmente, utiliza un programa de inteligencia artificial llamado *NoTouch BreastScan* midiendo los mismos parámetros que *Sentinel BreastScan* pero los algoritmos que identifican anomalías son modificados por una puntuación de 0 a 10 según su nivel de significancia.

Es posible apreciar que los resultados obtenidos en nuestro trabajo tras la búsqueda de los parámetros óptimos con la arquitectura *ResNet 101* son equivalentes al menos al de un experto en termografía, siendo este uno de los objetivos al emplear el aprendizaje automático. A su vez, los resultados sugieren tener un mejor desempeño que varios de los estudios mencionados en la Figura 5.5 de la curva ROC, con lo que podría ser utilizado como una técnica que permita dar resultados confiables en el análisis de termogramas.

Con esto también da posibilidad de que se utilice la termografía para el tamizaje de mujeres con tejido denso y se analicen sus resultados ya que los obtenidos por la mamografía en este tejido no son los mejores.

Lo anterior se puede verificar en el estudio realizado por Berg *et al* [26], quien comparó la forma en la cual se incrementan los valores de sensibilidad en el tamizaje de personas con posibilidad de tener cáncer de mama cuando solamente se utiliza la mamografía a cuando se utiliza de manera conjunta con el ultrasonido; los resultados pueden apreciarse en la Tabla 5.7.

Estudio	Se	Sp	AUC	PPV	NPV
Mamografía	50.0 %	95.5 %	0.68	14.7 %	99.2 %
Mamografía + Ultrasonido	77.5 %	89.4 %	0.9	10.1 %	99.6 %

Tabla 5.7: Comparativa de Berg *et al* con senos densos. El estudio [26] revela cómo mejora la sensibilidad y especificidad con el uso de la mamografía y ultrasonido en conjunto.

## 5.5. Comparativa Resolución de Imágenes

Tras realizar las pruebas del desempeño del algoritmo utilizando diferentes bases de datos, se generó la duda: ¿La sensibilidad térmica de la cámara afecta en el desempeño de la red neuronal? Para saber si se presentan diferencias se realizó una comparativa redondeando la matriz de temperatura de una muestra de 31 termogramas tomados de manera aleatoria para verificar la forma en la cual estos pueden cambiar su desempeño con las redes que tienen mejores resultados para desbalanceado y balanceado respectivamente.

Las tablas presentadas a continuación indican diferentes termogramas con su respectiva etiqueta y su cambio al momento de redondear las matrices de temperatura de 40mK que es la sensibilidad más baja, a 45mK, 50mK y al incrementar su redondeo a 80mK y 100mK.

La Tabla 5.5 muestra que el termograma 2 modificó su predicción de ‘Normal’ a ‘Anómalo’ cuando se redondeó a 45mK, pero también se aprecia que la red clasificó tuvo una variación de 2% con respecto a una y otra clasificación y cuando se redondea a 50mK la matriz de temperatura, este vuelve a la predicción cuando se tienen 40mK.

De igual forma, el termograma 12 cuando presenta un redondeo de 40mK como se tiene originalmente en la matriz de temperatura, tiene una posibilidad de 50% pero posteriormente cambia 1% a partir de que se redondea a 45mK y mantiene la predicción de ‘Anómalo’ desde ese momento siendo un error de la red clasificarlo así.

Por último, el termograma 16 cambia su predicción al redondearse 100mK, fuera de estos tres termogramas, la clasificación se mantiene prácticamente sin cambios en los porcentajes teniendo a su vez un buen desempeño en la clasificación de los termogramas.

Etiqueta de Termogramas		Redondeado 40mK		Redondeado 45 mK		Redondeado 50mK		Redondeado 80mK		Redondeado 100mK	
#	Clasificación Real	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob
1	Control	Normal	64 %	Normal	59 %	Normal	64 %	Normal	65 %	Normal	61 %
2	Control	Normal	51 %	Abnormal	51 %	Normal	51 %	Normal	56 %	Normal	54 %
3	Control	Normal	72 %	Normal	72 %	Normal	65 %	Normal	66 %	Normal	70 %
4	Control	Normal	61 %	Normal	67 %	Normal	64 %	Normal	66 %	Normal	63 %
5	Control	Normal	64 %	Normal	65 %	Normal	65 %	Normal	66 %	Normal	66 %
6	Control	Normal	54 %	Normal	55 %	Normal	57 %	Normal	56 %	Normal	56 %
7	Control	Normal	61 %	Normal	60 %	Normal	56 %	Normal	61 %	Normal	60 %
8	Control	Normal	62 %	Normal	61 %	Normal	61 %	Normal	58 %	Normal	60 %
9	Control	Normal	63 %	Normal	60 %	Normal	61 %	Normal	58 %	Normal	63 %
10	Control	Normal	60 %	Normal	67 %	Normal	61 %	Normal	63 %	Normal	63 %
11	Control	Normal	59 %	Normal	59 %	Normal	60 %	Normal	58 %	Normal	60 %
12	Control	Normal	50 %	Abnormal	51 %	Abnormal	53 %	Abnormal	56 %	Abnormal	61 %
13	Control	Normal	51 %	Normal	55 %	Normal	55 %	Normal	53 %	Normal	52 %
14	Control	Normal	56 %	Normal	53 %	Normal	58 %	Normal	61 %	Normal	58 %
15	Anómalo	Abnormal	74 %	Abnormal	77 %	Abnormal	75 %	Abnormal	76 %	Abnormal	75 %
16	Control	Normal	52 %	Normal	53 %	Normal	50 %	Normal	51 %	Abnormal	55 %
17	Anómalo	Abnormal	75 %	Abnormal	76 %	Abnormal	77 %	Abnormal	77 %	Abnormal	75 %
18	Anómalo	Abnormal	85 %	Abnormal	86 %	Abnormal	85 %	Abnormal	85 %	Abnormal	86 %
19	Anómalo	Abnormal	75 %	Abnormal	77 %	Abnormal	76 %	Abnormal	76 %	Abnormal	73 %
20	Anómalo	Abnormal	85 %	Abnormal	85 %	Abnormal	85 %	Abnormal	85 %	Abnormal	85 %
21	Anómalo	Abnormal	71 %	Abnormal	67 %	Abnormal	71 %	Abnormal	67 %	Abnormal	67 %
22	Anómalo	Abnormal	85 %	Abnormal	82 %	Abnormal	85 %	Abnormal	83 %	Abnormal	82 %
23	Anómalo	Abnormal	80 %	Abnormal	80 %	Abnormal	82 %	Abnormal	82 %	Abnormal	79 %
24	Anómalo	Abnormal	75 %	Abnormal	74 %	Abnormal	75 %	Abnormal	75 %	Abnormal	74 %
25	Anómalo	Abnormal	85 %	Abnormal	84 %	Abnormal	85 %	Abnormal	83 %	Abnormal	84 %
26	Anómalo	Abnormal	87 %	Abnormal	88 %	Abnormal	87 %	Abnormal	86 %	Abnormal	88 %
27	Anómalo	Abnormal	86 %	Abnormal	88 %	Abnormal	88 %	Abnormal	86 %	Abnormal	88 %
28	Anómalo	Abnormal	90 %	Abnormal	90 %	Abnormal	90 %	Abnormal	91 %	Abnormal	90 %
29	Anómalo	Abnormal	70 %	Abnormal	72 %	Abnormal	70 %	Abnormal	71 %	Abnormal	70 %
30	Anómalo	Abnormal	83 %	Abnormal	83 %	Abnormal	84 %	Abnormal	83 %	Abnormal	82 %
31	Anómalo	Abnormal	90 %	Abnormal	89 %	Abnormal	89 %	Abnormal	90 %	Abnormal	89 %

Tabla 5.8: Redondeo Matrices Temperatura Red Desbalanceada. Se muestran los termogramas con su respectiva etiqueta y las clasificaciones para la red desbalanceada con su respectiva probabilidad de que pertenezcan a esta categoría tras haberse redondeado a diferentes valores la matriz de temperatura.

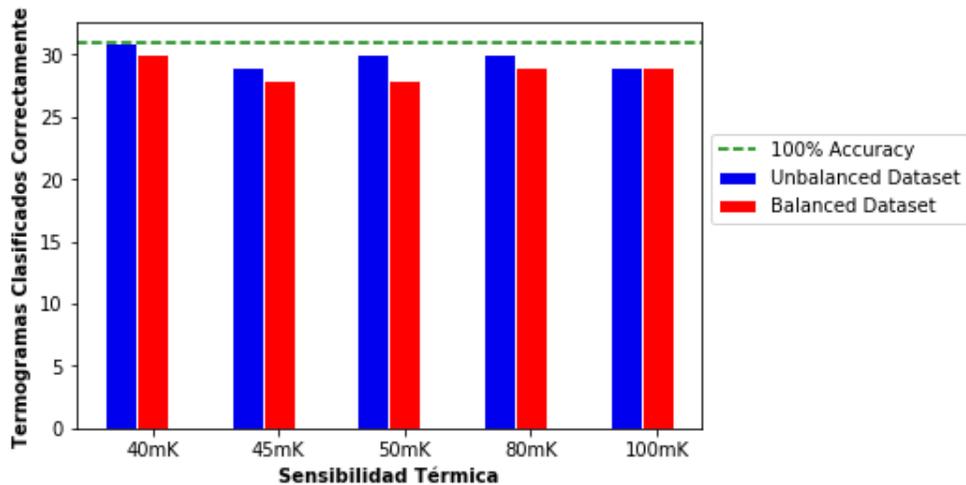


Figura 5.6: Clasificación Termogramas VS Sensibilidad Térmica. Exactitud de la clasificación de termogramas al variar la sensibilidad térmica de las imágenes como una comparativa al utilizar cámaras térmicas con una sensibilidad térmica diferente.

Etiqueta de Termogramas		Redondeado 40mK		Redondeado 45 mK		Redondeado 50mK		Redondeado 80mK		Redondeado 100mK	
#	Clasificación Real	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob	Predicción	% Prob
1	Control	Normal	59 %	Normal	69 %	Normal	69 %	Normal	68 %	Normal	63 %
2	Control	Normal	98 %	Normal	98 %	Normal	99 %	Normal	99 %	Normal	98 %
3	Control	Normal	98 %	Normal	99 %	Normal	97 %	Normal	99 %	Normal	99 %
4	Control	Normal	98 %	Normal	98 %	Normal	98 %	Normal	98 %	Normal	99 %
5	Control	Normal	93 %	Normal	96 %	Normal	94 %	Normal	95 %	Normal	94 %
6	Control	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %
7	Control	Normal	100 %	Normal	100 %	Normal	100 %	Normal	100 %	Normal	100 %
8	Control	Normal	100 %	Normal	100 %	Normal	100 %	Normal	100 %	Normal	100 %
9	Control	Normal	100 %	Normal	99 %	Normal	99 %	Normal	100 %	Normal	100 %
10	Control	Normal	100 %	Normal	99 %	Normal	100 %	Normal	100 %	Normal	99 %
11	Control	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %
12	Control	Normal	92 %	Normal	92 %	Normal	93 %	Normal	91 %	Normal	92 %
13	Control	Normal	97 %	Normal	97 %	Normal	96 %	Normal	94 %	Normal	97 %
14	Control	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %	Normal	99 %
15	Anómalo	Abnormal	97 %	Abnormal	98 %	Abnormal	99 %	Abnormal	98 %	Abnormal	98 %
16	Control	Normal	59 %	Abnormal	53 %	Abnormal	53 %	Normal	57 %	Normal	55 %
17	Anómalo	Normal	97 %	Normal	97 %	Normal	97 %	Normal	96 %	Normal	96 %
18	Anómalo	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %	Abnormal	97 %	Abnormal	98 %
19	Anómalo	Abnormal	100 %	Abnormal	100 %	Abnormal	100 %	Abnormal	100 %	Abnormal	100 %
20	Anómalo	Abnormal	92 %	Abnormal	92 %	Abnormal	88 %	Abnormal	93 %	Abnormal	93 %
21	Anómalo	Abnormal	100 %	Abnormal	100 %	Abnormal	100 %	Abnormal	100 %	Abnormal	99 %
22	Anómalo	Abnormal	98 %	Abnormal	98 %	Abnormal	97 %	Abnormal	99 %	Abnormal	99 %
23	Anómalo	Normal	57 %	Abnormal	50 %	Normal	68 %	Normal	59 %	Normal	57 %
24	Anómalo	Abnormal	52 %	Abnormal	50 %	Abnormal	56 %	Abnormal	56 %	Abnormal	61 %
25	Anómalo	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %
26	Anómalo	Abnormal	99 %	Abnormal	99 %	Abnormal	98 %	Abnormal	98 %	Abnormal	100 %
27	Anómalo	Abnormal	96 %	Abnormal	97 %	Abnormal	96 %	Abnormal	95 %	Abnormal	94 %
28	Anómalo	Abnormal	99 %	Abnormal	99 %	Abnormal	99 %	Abnormal	99 %	Abnormal	99 %
29	Anómalo	Abnormal	99 %	Abnormal	99 %	Abnormal	97 %	Abnormal	98 %	Abnormal	98 %
30	Anómalo	Abnormal	99 %	Abnormal	99 %	Abnormal	100 %	Abnormal	99 %	Abnormal	100 %
31	Anómalo	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %	Abnormal	98 %	Abnormal	96 %

Tabla 5.9: Redondeo Matrices Temperatura Red Balanceada. Se muestran los termogramas con su respectiva etiqueta y las clasificaciones para la red balanceada con su respectiva probabilidad de que pertenezcan a esta categoría tras haberse redondeado a diferentes valores la matriz de temperatura.

En la Tabla 5.5 también son 3 termogramas los que varían su predicción en el que de nueva cuenta se tiene una variación pero ahora mucho más marcada cuando se redondea la matriz de temperatura en 45mK y 50mK al ser clasificado con 53 % de probabilidad de ser anómalo y posteriormente vuelve su predicción a ‘Normal’. El termograma 17 es clasificado erróneamente teniendo en cada uno de los casos una probabilidad mayor al 95 % de ser clasificado como ‘Normal’.

Finalmente, el termograma 23 cambia en la predicción al ser redondeado a 45mK al disminuir a un 50 % la probabilidad de caer en una u otra categoría, aunque para el resto de los valores vuelve a tener una probabilidad mayor de pertenecer a la clase ‘Normal’, siendo errónea su clasificación.

En conclusión, de 31 termogramas utilizados para el análisis, 15 correspondieron al grupo control, y 16 de ellos al de anómalos, los cuales fueron redondeados a diferentes valores de temperatura. Los resultados observados en las Tablas 5.5 y 5.5 se pueden apreciar a detalle los resultados respectivos tras haberlos ingresado en la red.

La Figura 5.6 indica cuántos termogramas fueron clasificados correctamente en comparación con su etiqueta. Puede observarse que para ambos casos se tiene un escenario que sería el peor de estos al clasificar erróneamente a 3 de los 31 pacientes.

Basados en estos resultados y en la probabilidad de clasificarse en una u otra categoría, la sensibilidad térmica de una cámara es una variable que puede influir en el estudio para clasificar de manera incorrecta hasta en un 9.6 % y esto se añade al error que puede tener la arquitectura por sí

Enfoque	Grupos Termogramas	Se	Sp	AUC	PPV	NPV	Acc	Gmean
Desbalanceado	T600 y T400	37.5 %	39.2 %	0.360	13.8 %	71 %	38.4 %	38.3 %
	SC600 y T600	60.8 %	63 %	0.634	54.3 %	69 %	61.9 %	61.8 %
	SC600 y T400	37.8 %	63 %	0.552	15.8 %	84.5 %	50.3 %	48.6 %
Balanceado	T600 y T400	15.6 %	100 %	0.738	100 %	54.2 %	57.8 %	39.5 %
	SC600 y T600	0 %	25 %	0.039	0 %	20 %	12.5 %	0 %
	SC600 y T400	15.6 %	25 %	0.108	17.2 %	22.9 %	20.3 %	19.8 %

Tabla 5.10: Prueba Clasificación por Cámaras. Métricas de desempeño para verificar que los algoritmos aprendieran en función del problema a resolver y no de la cámara utilizada.

misma. A su vez, se ha comentado el hecho de la pérdida de información relevante al analizar las imágenes en una red neuronal para diversos problemas, esto debido a que generalmente son mucho más grandes que las permitidas por la arquitectura como entrada.

## 5.6. Comparativa Cámaras

Para verificar que el algoritmo no haya aprendido a identificar las imágenes en función de la cámara con la cual fueron adquiridos, se realizará un experimento en el cual los termogramas serán separados por centro de adquisición y guardados en las carpetas correspondientes.

Se sabe que los algoritmos pueden ser entrenados para clasificar en el número de clases que se desee, como ejemplo, están las mil categorías en las cuales se entrenan utilizando la base de datos de *ImageNet*, pero para que esto sea posible, deben de estar entrenados en identificar estas clases.

A causa de esto, y ya que está entrenado el algoritmo para clasificar solamente dos clases, se realizarán tres experimentos con las combinaciones de cada una de las cámaras que se utilizaron para la adquisición de los termogramas.

La comparación se realizó con las etiquetas dadas originalmente guardando los termogramas en sus respectivas carpetas llevando el nombre de *‘Abnormal’* y *‘Normal’* tanto con el enfoque de clases balanceadas como desbalanceadas empleando las redes que arrojaron los mejores resultados al evaluar los termogramas tras la búsqueda de los parámetros óptimos. En caso de que las métricas sean iguales a las arrojadas en la prueba inicial, se tendría un algoritmo que aprendió en función de las cámaras y no en función de las anomalías del termograma.

Los resultados de la prueba se pueden observar en la Tabla 5.10, indicando las diferentes combinaciones en función de las cámaras utilizadas para el análisis, siendo tres por cada enfoque, y los valores de la métrica de desempeño, mostrando así que ninguna de las métricas coincide con los arrojados en la Tabla 5.6, dando así motivos para tener una confianza para concluir que no se está clasificando en función de la cámara empleada para adquirir los termogramas, sino de los rasgos de la imagen del problema a estudiar.

## 5.7. Imágenes Ajustadas

Debido a que en algunas ocasiones pueden obtenerse los termogramas de manera incorrecta debido a la presión que se tiene de la parte clínica para poder ver al mayor número de pacientes

siendo limitados en un cierto tiempo, no es raro pensar que el área correspondiente a los senos no está optimizada según lo indicado en los protocolos de adquisición mencionados en la sección 3.2

Es así que se decide a realizar un recorte a las imágenes de manera manual a las imágenes con extensión .tif de forma que se centre únicamente en el área de interés para conocer si esto puede ayudar a tener a la red neuronal a tener una mejor decisión sin otros elementos que puedan distraerle.

Uno de los retos al realizar esto fue el tamaño de la imagen ya que era inferior al de 224x224 píxeles que son los requeridos para la entrada en *ResNet 101* por lo que cada termograma se completó con ceros alrededor hasta tener los 224 píxeles en total.

La comparativa se realizó con la arquitectura *ResNet 101* teniendo enfoques tanto balanceado como desbalanceado, con 100 iteraciones en cada una eligiendo las que dieron mejores resultados, de nueva cuenta con los criterios de sensibilidad, área bajo la curva ROC y especificidad mostrándose los mejores en la Tabla 5.11.

Enfoque	Se	Sp	AUC	PPV	NPV	Bal Acc	G mean
Balanceado	81.8 %	63.8 %	0.756	69.2 %	77.8 %	72.7 %	72.1 %
Desbalanceado	90.9 %	32.5 %	0.618	27.8 %	92.6 %	61.7 %	54.3 %

Tabla 5.11: Resultados Termogramas Ajustados. Los mejores resultados Tras iteraciones se presentan los mejores resultados con los termogramas ajustados.

Los valores de sensibilidad no fueron malos, sin embargo, quedó a deber en el área bajo la curva, así como en los valores de especificidad y el conjunto balanceado en el valor predictivo negativo.

Si bien es cierto que debería de hacerse énfasis en respetar el protocolo de adquisición de termogramas como un paso fundamental para el análisis de estos evitando perder información de la imagen en otros detalles, también es cierto que un área de oportunidad para utilizar las imágenes ajustadas permitiendo centrarse solamente en la solución del problema bajo dos enfoques.

El primero sería realizar un recorte solamente del área de interés de manera automática para que la arquitectura analice solamente los senos y determine si son anómalos o no. El segundo, multiplicar con píxeles sintéticos que sean similares a los que integran a la matriz de temperatura para que el área de interés de la imagen se llene al tamaño requerido aunque se corre el riesgo de deformar la imagen.

## 5.8. Clasificación Categoría de la Lesión en la Mama

Cuando una paciente presenta una anomalía en la mama y se le manda a biopsia, la resolución de esta puede ser el tener un tumor maligno, benigno, o no presentar anomalías. Es por ello que se realiza un análisis para poder identificar el tipo de lesión que se tiene utilizando el criterio descrito en la Tabla 2.2 como base para saber a qué categoría debería de pertenecer cada paciente.

El análisis se realiza utilizando la base de datos adquirida en el IJC, así como los termogramas pertenecientes al grupo control de Brasil debido a que Silva *et al* [101] no agrega los resultados del BI-RADS para los clasificados como anómalos en su base de datos, y que aún no se contaba con la base de datos del ISSSTE. A la base de datos que integra el estudio se le aplica la prueba de normalidad de Shapiro-Wilk arrojando que los datos siguen una distribución normal.

La Tabla 5.12 presenta el número de termogramas que integran a tanto a los conjuntos de entrenamiento, validación y totales para las tres categorías, así como el total analizado. En esta prueba también se realizó la comparativa cuando se tienen clases balanceadas como desbalanceadas.

Una de las limitantes que se tienen para esta prueba es el saber que el número de termogramas con los cuales se está entrenando la red es muy chico, por lo que se trata de compensar realizando un incremento de 500 veces cada uno de los termogramas para la clase desbalanceada, mientras que para la clase balanceada se realiza en 25 ocasiones la copia de sendas imágenes.

Comparación Base de Datos			Enfoque	
			Desbalanceado	Balanceado
Benigno	Entrenamiento	n (70 %)	25	8
		Edad	48	51
		SD ( $\pm$ años)	8.4	9.76
	Validación	n (30 %)	10	3
		Edad	45	44
		SD ( $\pm$ años)	12.95	7.5
Total	n	35	11	
	Edad	47	49.36	
	SD ( $\pm$ años)	9.83	9.33	
Maligno	Entrenamiento	n (70 %)	27	8
		Edad	52	46
		SD ( $\pm$ años)	12.96	8.73
	Validación	n (30 %)	11	3
		Edad	56	62
		SD ( $\pm$ años)	18.84	19.65
Total	n	38	11	
	Edad	53	50	
	SD ( $\pm$ años)	12.66	13.68	
Normal	Entrenamiento	n (70 %)	8	8
		Edad	52	52
		SD ( $\pm$ años)	7.74	7.74
	Validación	n (30 %)	3	3
		Edad	52	52
		SD ( $\pm$ años)	4.93	4.93
Total	n	11	11	
	Edad	52	52	
	SD ( $\pm$ años)	6.85	6.85	
Total	Entrenamiento	n (70 %)	60	24
		Edad	51	49
		SD ( $\pm$ años)	10.65	8.83
	Validación	n (30 %)	24	9
		Edad	51	53
		SD ( $\pm$ años)	15.89	13.24
Total	n	84	33	
	Edad	51	50	
	SD ( $\pm$ años)	12.27	10.1	

Tabla 5.12: Base de Datos Multiclases. Se presentan los termogramas que componen al conjunto de entrenamiento, validación y totales para cada categoría tanto para el enfoque desbalanceado como balanceado.

Es de resaltar que la matriz de confusión ahora será de 3x3, por lo que ahora se tendrán tantos valores como se tengan de clases, en este caso siendo 3, mientras que las métricas de desempeño del

clasificador se realizan como se hizo mención en la sección 4.6. Los valores de la matriz de confusión para cada categoría se calculan de la siguiente manera:

$$i = 1 : \text{número de clases} \tag{5.1}$$

$$TP(i) = \text{Matriz de confusión } (i, i) \tag{5.2}$$

$$FN(i) = \sum \text{Matriz de confusión } (:, i) - TP(i) \tag{5.3}$$

$$FP(i) = \sum \text{Matriz de confusión } (:, i) - TP(i) \tag{5.4}$$

$$TN(i) = \sum \text{Matriz de confusión } - TP(i) - FP(i) - FN(i) \tag{5.5}$$

Se realizaron 30 iteraciones para cada una de las arquitecturas en sendas clases, en el análisis cíclico con **clases balanceadas** arrojando los mismos valores en las métricas de desempeño en 28 ocasiones, 12 de ellas con *AlexNet*, otras 12 con *VGG 16*, par de ocasiones con *ResNet 50* y en una ocasión con *VGG 19* y *ResNet 101*. Por otra parte, el análisis con **clases desbalanceadas** la arquitectura con mejores valores en cada uno de los campos fue *ResNet 50*; los valores pueden apreciarse en la Tabla 5.13.

Enfoque	Se	Sp	PPV	NPV	Acc	G mean
Balanceado	77.7 %	88.9 %	86.6 %	91.6 %	77.7 %	78.4 %
Desbalanceado	71.7 %	90.4 %	90.4 %	92.6 %	83.3 %	77.5 %

Tabla 5.13: Clasificación Categoría de la Lesión en la Mama. Se muestran los mejores resultados cuando se tienen clases balanceadas y desbalanceadas al clasificar la lesión según el BI-RADS en maligno, benigno y control.

Aunque los resultados mostrados en sensibilidad, valor predictivo negativo y especificidad no lucen mal, el número de termogramas utilizados para el análisis es muy chico como para poder determinar que funcione para la clasificación. En 26 de las mejores corridas con clases desbalanceadas se obtuvo en promedio que el algoritmo clasifica correctamente a casi todos los benignos, sin embargo tiene problemas para distinguir a los pacientes del grupo control de los que tienen una lesión benigna, además de cometer algunos errores al momento de clasificar los termogramas malignos con benignos. Por su parte, cuando se tienen clases balanceadas llega a confundir termogramas del grupo control con benignos, pero la diferenciación entre los termogramas benignos y malignos no tiene problemas para diferenciarlos.

El estudio realizado por Ribli *et al* [114] realiza una clasificación de lesiones de mamografías utilizando aprendizaje profundo en donde estas también son catalogadas como lesiones benignas, malignas y controles, sin embargo utilizan cerca de 86,000 imágenes para su entrenamiento. En imágenes infrarrojas destaca el realizado por Parisky *et al* [49] que se ayuda de sistemas computarizados que evalúan termogramas dinámicos para clasificar las imágenes en benignos, malignos y controles a partir de una base de datos de 769 pacientes de 5 instituciones diferentes para lo cual le tomó 4 años el recabar los datos para el estudio en los cuales contó con radiólogos que evaluaron las mamografías adquiridas de forma conjunta con los termogramas.

De esta forma, los resultados sugieren que la técnica luce como una opción para ser empleada pero que requerirá de un mayor número de imágenes que puedan servir para entrenar los algoritmos y que permitan tener una certeza que los resultados obtenidos no están relacionados con un pequeño número de termogramas de entrenamiento que le son proporcionados.

## 5.9. Análisis de Individual de las Bases de Datos

Una vez que se analizaron las bases de datos proporcionadas por los diferentes centros de manera conjunta surge el interés por conocer si el algoritmo podrá mantener los mismos valores de la métrica de desempeño cuando se analiza cada centro de manera individual.

Base de Datos	Enfoque	Se	Sp	AUC	PPV	NPV	Acc	Gmean
Brasil	Balanceado	87.5 %	37.5 %	0.812	58.3 %	75 %	62.5 %	57.2 %
	Desbalanceado	87.5 %	52.6 %	0.847	28 %	95.2 %	70 %	67.8 %
IJC	Balanceado	50 %	75 %	0.687	66.6 %	60 %	62.5 %	61.2 %
	Desbalanceado	75 %	94.4 %	0.888	75 %	94.4 %	84.7 %	84.1 %
ISSSTE	Balanceado	0 %	100 %	0		50 %	50 %	0 %
	Desbalanceado	0 %	100 %	0.812		94.1 %	50 %	0 %

Tabla 5.14: Resultados por centro de adquisición. Los enfoques balanceado y desbalanceado son integrados en este análisis.

De esta forma, los datos plasmados en la Tabla 5.2 se ingresan a la red *ResNet 101* que dio mejores resultados con clases balanceadas y desbalanceadas respectivamente, para verificar la efectividad de los mismos. Los resultados son mostrados en la Tabla 5.14 que muestra un mejor desempeño cuando se analiza la base de datos de Silva *et al* [101], mientras que al analizar los termogramas adquiridos en el IJC tiene un desempeño rescatable en el enfoque desbalanceado, algo que no repite al analizarse de manera balanceada; a su vez, con los datos del ISSSTE da valores que no podrían tomarse como válidos y en donde los valores de PPV no pueden calcularse.

Al revisar la Figura 5.2 es posible observar que se tiene un número limitado de termogramas anómalos de validación en el IJC con 4 por 18 controles, mientras que en el ISSSTE es solamente un anómalo contra los 16 del grupo control, a diferencia de la base de datos adquirida en Brasil que es probada con 8 termogramas anómalos y 38 controles, esto en el enfoque desbalanceado, mientras que en el balanceado los termogramas analizados están limitados por el número de termogramas anómalos.

Debido a que existe una incertidumbre entre saber si los resultados obtenidos son debido al número de termogramas analizados en cada centro o si en verdad influye la forma en la que fueron adquiridos es que se realiza un segundo experimento realizando tres grupos distribuyendo los termogramas como se muestra en la Tabla 5.15. Los grupos contenían datos adquiridos en diferentes centros de manera aleatoria para verificar la variación con respecto al número.

Enfoque	Clasificación	Grupo A	Grupo B	Grupo C
Balanceado	Control-Enfermo	8	4	1
	Control	38	18	16
Desbalanceado	Anómalo	8	4	1

Tabla 5.15: Integración de Grupos por enfoque. Número de termogramas que se integraron en cada grupo para realizar la comparativa en su clasificación.

De esta forma, los tres grupos son evaluados para cada uno de los enfoques con la misma arquitectura, siendo presentados en la Tabla 5.16 especificando solamente los valores de sensibilidad, especificidad y área bajo la curva ROC debido a que en casi todos los campos quedaban vacíos el resto de las métricas.

Tras este experimento se refuerza nuestra hipótesis en la cual el número de termogramas analizados por el algoritmo influye en los resultados obtenidos en las métricas de desempeño más que con el centro en el cual fueron adquiridos.

Métricas	Balanceado			Desbalanceado		
	Grupo A	Grupo B	Grupo C	Grupo A	Grupo B	Grupo C
Sensibilidad	88 %	100 %	100 %	75 %	100 %	100 %
Especificidad	63 %	0 %	0 %	63 %	72 %	63 %
AUC	0.78	0.50	1	0.75	0.97	0.94

Tabla 5.16: Métricas de Desempeño por Grupos. Obtenidas a partir de la aplicación de la arquitectura *ResNet 101* a los diferentes grupos analizados.

## Capítulo 6

# Conclusiones

El presente trabajo muestra el uso de **transferencia de aprendizaje de redes neuronales convolucionales profundas** para poder **clasificar termogramas de cáncer de mama** teniendo como una de las **ventajas principales** el que **no se necesita el preprocesamiento** de la imagen, sino que se **ingresa** tal cual la **matriz de temperaturas** a la red. Otra ventaja es el hecho de poder hacer la **evaluación** de forma **estandarizada** y no dependerá de la persona que esté analizando el termograma para clasificarlo dentro de alguna categoría.

La tarea se realizó con dos enfoques diferentes, con clases balanceadas en donde se tiene la misma probabilidad de pertenecer a una clase u otra, así como con clases desbalanceadas, donde se tiene una prevalencia mayor de población que no tiene la enfermedad, tal como se espera ocurra en pruebas con la población.

Se obtuvieron valores de **sensibilidad** de **92.3 %** y **84.6 %** mientras que de **especificidad** fueron de **53.8 %** y **65.3 %**, y un AUC de 0.814 a 0.722 de acuerdo a la Tabla 5.6. Estos resultados son comparables a los que tienen expertos de acuerdo a lo descrito por de Jesus Guirro *et al* 109 y Omnaripour *et al* 107. Los resultados obtenidos con el enfoque desbalanceado son equivalentes a los obtenidos por expertos cuando los termogramas son analizados de forma manual mientras que al tener clases balanceadas se supera este rendimiento; las comparaciones con otros estudios similares se visualizan en la Figura 5.5

Debido a que el algoritmo se realizó con datos adquiridos en diferentes centros y usando dos modelos distintos de cámaras infrarrojas, así como al aplicar las técnicas de aumento de datos de manera sintética, se cree que el presente trabajo trata de realizar una buena generalización con respecto a la variabilidad con lo cual se tenga una cierta robustez en el método. A su vez, se tiene claro que cualquier modelo de aprendizaje de máquina necesita ser validado en diferentes estudios, investigaciones y bases de datos para que sea más robusto y pueda ser aplicado de forma clínica.

Existen diferentes métodos que pueden ayudar en el tamizaje de cáncer de mama, la **termografía** es uno de ellos que presenta varias ventajas tales como el ser un método no invasivo, que no utiliza radiación y que resulta ser cómodo para las pacientes por no causar dolor al momento de adquirir los termogramas. Su uso es **complementario** al de la mamografía, es decir, es una **herramienta adjunta** para ayudar a mejorar la capacidad de detección de cáncer de mama.

La termografía, al igual que otros métodos de imagenología, carecen de procedimientos estandarizados de evaluación, por ello, los valores obtenidos de sensibilidad y especificidad llegan a presentar una alta variabilidad en función de la experiencia y capacidad de la persona que analice el estudio. Contar con técnicas de aprendizaje automático tanto en el análisis de imágenes médicas tanto en este

como en otros problemas, están orientados a la búsqueda de la estandarización en función de un modelo entrenado que sirvan como herramientas a la parte médica en la toma de decisiones y diagnóstico brindado al paciente.

En lo personal coincido con lo expresado por Ferryman [115] sobre las herramientas de inteligencia artificial, indicando que no se tiene una seguridad en que las herramientas de inteligencia artificial en el cuidado de la salud garanticen beneficios para la población total y cumplan con todas las expectativas que se tienen en ellos. Esto debido a que influyen otros factores tales como la obtención de datos y que no tengan problemas de privacidad de los mismos, así como ser la forma más certera de diagnóstico.

Por otra parte, se tienen grandes expectativas en que sean una **herramienta** más para que los **expertos** en materia de salud **puedan utilizarlo** en diferentes áreas y permitan al **sistema de salud** ser **más justo y equitativo** al preocuparse por las personas.

De resaltar que se insista en lo expresado por diversas personas líderes en la materia tales como Andrew Ng y Lily Peng por mencionar a dos referentes en el área de inteligencia artificial, al mencionar que “*el aprendizaje de máquina nunca superará a un médico con experiencia*” [116], por lo cual deben considerarse como herramientas para la toma de decisiones por la parte clínica y que **previo** a que se realice un **diagnóstico seleccionando** solamente a los **pacientes de mayor riesgo** para su evaluación y se **reduzca la carga de trabajo clínico**, ayudando así a **expandir el acceso a la salud** en áreas marginadas y/o rurales.

Los problemas con el tamaño del conjunto de datos son comunes; Andrew Ng [59] comenta que en algunos entrenamientos de algoritmos, sobre todo en aquellos de transferencia de aprendizaje automático, es normal contar con análisis que tengan entre mil y 10 mil elementos y para que el algoritmo presente mejoras de 0.1% en su precisión se deberá de contar con más de 10 mil datos.

Una de las limitaciones del estudio, es el no haber podido llevar a comparar este algoritmo a un centro con datos adquiridos en ese momento y evaluar su desempeño, así como al hablar del tamaño de la base de datos, no haber tenido un mayor número de termogramas que cuenten con la clasificación equivalente de BI-RADS, o de termogramas malignos que detecten el tipo de lesión, así como el no incluir a pacientes que tengan cirugía plástica o prótesis mamaria, en el análisis.

## 6.1. Trabajo a Futuro

Un área de oportunidad que existe es el incrementar el número de termogramas en los cuales se tenga la etiqueta del BI-RADS para poder determinar la categoría a la que pertenece y sea considerado como benigno, maligno o control para tener de esta forma herramientas que permitan decidir si se tendrá que realizar una biopsia u otras medidas según lo considere la parte médica, ya que es un experimento presentado en la sección 5.8 donde los resultados sugieren que en caso de poder contar con una mayor cantidad de imágenes se podría realizar una contribución en este campo.

A su vez, si se consigue una fuerte cantidad de datos etiquetados de termogramas con los cuatro tipos de cáncer más comunes que son los carcinomas ductal *in situ*, lobulillar *in situ*, ductal invasivo y lobulillar invasivo, presentados en la sección 2.3.1 podría compararse directamente con los resultados arrojados por la biopsia.

En trabajos liderados tanto por la Universidad de Stanford como por Google en su rama de investigación en ciencias de la salud, se han enfocado en estos retos que puedan ser benéficos para que el médico pueda tomar una decisión, tal como se presenta en [117] en cáncer de hígado por citar un ejemplo, donde el algoritmo empleado sea capaz de tener una buena respuesta con datos diferentes a los cuales se utilizaron para entrenar el algoritmo que arrojó buenos resultados.

De igual forma, otra opción sería el realizar la segmentación de manera automática de los senos en la termografía para que la red neuronal de un resultado enfocándose netamente a los senos y sin que se tengan otros elementos que puedan generarle confusión.

Por último, es de resaltar que no solamente en este trabajo, sino en general para aquellos que trabajan con alguna forma de inteligencia artificial, queda pendiente el poder resolver los problemas de tener un conjunto de datos pequeño, así como la robustez para que este pueda tener una buena generalización independientemente del conjunto de datos que se le pueda proporcionar.

## 6.2. Publicaciones

Las siguientes obras científicas fueron publicadas durante la obtención del grado:

Juan Carlos Torres-Galván, Edgar Guevara, Eleazar Samuel Kolosovas-Machuca, Antonio Ocegüera-Villanueva, Jorge L. Flores and F.J. González, “*Deep Convolutional Neural Networks for classifying breast cancer using infrared thermography*”. in Quantitative InfraRed Thermography Journal. DOI: 10.1080/17686733.2021.1918514

Juan Carlos Torres-Galván, Edgar Guevara, F.J. González, “*Comparison of Deep Learning Architectures for Pre-Screening of Breast Cancer Thermograms*”. in 2019 Photonics North (PN). 2019 Photonics North (PN). Vol. CFP1909V-ART. Mayo de 2019, págs. 1-2. doi:10.1109/PN.2019.8819587

Edgar Guevara, Juan Carlos Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, “*Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools*” Biomed. Opt. Express, BOE, vol. 9, no. 10, pp. 4998–5010, Oct. 2018, doi: 10.1364/BOE.9.004998.

Edgar Guevara, Juan Carlos Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, “*Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools: reply to comment,*” Biomed. Opt. Express, BOE, vol. 10, no. 9, pp. 4492–4495, Sep. 2019, doi: 10.1364/BOE.10.004492

Edgar Guevara, Juan Carlos Torres-Galván, M. G. R. Elías, C. Luévano-Contreras, and F. J. González, “*Non-invasive in vivo Raman spectroscopy of the skin for diabetes screening,*” in 2017 Photonics North (PN), Jun. 2017, pp. 1–2, doi: 10.1109/PN.2017.8090556.

## 6.3. Participaciones en Congresos y Simposiums

Durante la obtención del grado se tuvieron varias participaciones en congresos y simposiums, los cuales se detallan a continuación:

“*Comparison of Deep Learning Architectures for Pre-Screening of Breast Cancer Thermograms*”, Photonics North 2018, Junio 2018, Montreal, Canadá.

“*Análisis Automático de Termogramas de Cáncer de Seno*”, Concurso de Exhibición de Carteles de Proyectos de Investigación de Estudiantes de Posgrado de la UASLP, Octubre 2017, San Luis Potosí, México.

“*Análisis Automático de Termogramas de Cáncer de Seno*”, Concurso de Exhibición de Carteles de Proyectos de Investigación de Estudiantes de Posgrado de la UASLP, Octubre 2018, San Luis Potosí, México.

“*Análisis de Termogramas de Cáncer de Seno Mediante el Uso de Aprendizaje Profundo*”, Seminario Laboratorio Nacional CIACYT, Septiembre 2018, San Luis Potosí, México.

“*Clasificación de Termografías de Cáncer de Seno Utilizando Transfer Learning*”, 1er Foro de Investigadores Cátedras CONACYT de la Universidad Autónoma de San Luis Potosí, Junio 2019, San Luis Potosí, México.

# Apéndice A

## Códigos Utilizados

Los códigos utilizados para la elaboración de los diferentes análisis pueden encontrarse en github. En los siguientes párrafos se indicará el link y una breve descripción sobre el uso dado a cada uno de los códigos.

El aumento de datos de forma sintética se aprecia en <https://github.com/TorresGalvanJC/PhDThesis/blob/master/AugmentedImages.m>. Originalmente se había realizado las transformaciones de inclinación, pero al ver que esto no aplicaba se canceló tanto este como sus diferentes combinaciones.

La búsqueda de los parámetros óptimos requirió de varias iteraciones, disponibles en <https://github.com/TorresGalvanJC/PhDThesis/tree/master/GridSearch>, están separadas en función del enfoque utilizado. Por su parte, los códigos mediante los cuales se crearon los mapeos de activación de clase están dados en el link <https://github.com/TorresGalvanJC/PhDThesis/tree/master/ZhouCAMVGG16>.

La gráfica en donde se aprecia la comparativa de las curvas ROC de ambos enfoques utilizados contra el empleado en otros estudios de termografía se encuentra en [https://github.com/TorresGalvanJC/PhDThesis/blob/master/ROC\\_Thermography\\_Comparison.m](https://github.com/TorresGalvanJC/PhDThesis/blob/master/ROC_Thermography_Comparison.m).

El redondeo de las matrices de temperatura utilizadas se encuentra en <https://github.com/TorresGalvanJC/PhDThesis/blob/master/RoundTemperatures.m> para su posterior evaluación en <https://github.com/TorresGalvanJC/PhDThesis/blob/master/KolosovasSensibilidadTermica.m> donde se muestran las evaluaciones tanto para clases desbalanceadas como balanceadas, a las cuales se le agregó un 1 al inicio para diferenciarlo.

El código <https://github.com/TorresGalvanJC/PhDThesis/blob/master/Test3RandomGroups.m> muestra la comparativa al crear tres grupos aleatorios para ser clasificados. Así como para realizar las clasificaciones determinando si el termograma tiene lesiones benignas, malignas o es del grupo control para cada arquitectura se encuentra en <https://github.com/TorresGalvanJC/PhDThesis/tree/master/BenignMalignantNormal>.

# Bibliografía

- [1] K.L. Moore, A.F. Dalley y A.M.R. Agur. *Anatomía con Orientación Clínica*. Lippincott Williams & Wilkins, 2013. Cap. 1, págs. 151-153. ISBN: 9788415684770.
- [2] “Rook’s textbook of dermatology”. En: ed. por Arthur Rook y Tony Burns. 7th ed. Malden, Mass: Blackwell Science, 2004. Cap. 67, pág. 67.1. ISBN: 978-0-632-06429-8.
- [3] Satish G. Kandlikar y col. “Infrared imaging technology for breast cancer detection – Current status, protocols and new directions”. En: *International Journal of Heat and Mass Transfer* 108, Part B (mayo de 2017), págs. 2303-2320. ISSN: 0017-9310. DOI: [10.1016/j.ijheatmasstransfer.2017.01.086](https://doi.org/10.1016/j.ijheatmasstransfer.2017.01.086). URL: <http://www.sciencedirect.com/science/article/pii/S0017931016336031>.
- [4] J Ferlay y col. *Global Cancer Observatory: Cancer Today*. en. Lyon, France: International Agency for Research on Cancer. URL: <http://gco.iarc.fr/today>.
- [5] María del Rocío Castrezana Campos y María del Rocío Castrezana Campos. “Geografía del cáncer de mama en México”. En: *Investigaciones geográficas* 93 (ago. de 2017). Publisher: Instituto de Geografía, UNAM, págs. 0-0. ISSN: 0188-4611. DOI: [10.14350/rig.56879](https://doi.org/10.14350/rig.56879). URL: [http://www.scielo.org.mx/scielo.php?script=sci\\_abstract&pid=S0188-46112017000200010&lng=es&nrm=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0188-46112017000200010&lng=es&nrm=iso&tlng=es).
- [6] Freddie Bray y col. “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. en. En: *CA: A Cancer Journal for Clinicians* 68.6 (2018), págs. 394-424. ISSN: 1542-4863. DOI: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492). URL: <https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21492>.
- [7] J. Ferlay y col. “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods”. en. En: *International Journal of Cancer* 144.8 (2019), págs. 1941-1953. ISSN: 1097-0215. DOI: [10.1002/ijc.31937](https://doi.org/10.1002/ijc.31937). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31937>.
- [8] Marina Milosevic, Dragan Jankovic y Aleksandar Peulic. “Thermography based breast cancer detection using texture features and minimum variance quantization”. En: *EXCLI J* 13 (nov. de 2014), págs. 1204-1215. ISSN: 1611-2156. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4464488/>.
- [9] U. Raghavendra y col. “Computer-aided diagnosis for the identification of breast cancer using thermogram images: A comprehensive review”. En: *Infrared Physics & Technology* 102 (nov. de 2019), pág. 103041. ISSN: 1350-4495. DOI: [10.1016/j.infrared.2019.103041](https://doi.org/10.1016/j.infrared.2019.103041). URL: <http://www.sciencedirect.com/science/article/pii/S1350449519304608>.
- [10] *Tipos de cáncer de seno*. URL: <https://www.cancer.org/es/cancer/cancer-de-seno/compreension-de-un-diagnostico-de-cancer-de-seno/tipos-de-cancer-de-seno.html>.
- [11] Md. Shajedur Rahman Shawon, Mikael Eriksson y Jingmei Li. “Body size in early life and risk of breast cancer”. En: *Breast Cancer Research* 19.1 (jul. de 2017), pág. 84. ISSN: 1465-542X. DOI: [10.1186/s13058-017-0875-9](https://doi.org/10.1186/s13058-017-0875-9). URL: <https://doi.org/10.1186/s13058-017-0875-9>.

- [12] Paolo Bruzzi y col. "Estimating the population attributable risk for multiple risk factors using case-control data". En: *American journal of epidemiology* 122.5 (1985), págs. 904-914.
- [13] Peh Joo Ho y col. "Incidence of breast cancer attributable to breast density, modifiable and non-modifiable breast cancer risk factors in Singapore". En: *Scientific Reports* 10.1 (2020), págs. 1-11.
- [14] *Breast Cancer Facts & Figures | American Cancer Society*. Library Catalog: [www.cancer.org](http://www.cancer.org). URL: <https://www.cancer.org/research/cancer-facts-statistics/breast-cancer-facts-figures.html>.
- [15] *Breast Self-Exam*. National Breast Cancer Foundation. Library Catalog: [www.nationalbreastcancer.org](http://www.nationalbreastcancer.org). URL: <https://www.nationalbreastcancer.org/breast-self-exam> (visitado 06-05-2020).
- [16] Irene Ma y col. "Clinical and self breast examination remain important in the era of modern screening". En: *Annals of surgical oncology* 19.5 (2012), págs. 1484-1490.
- [17] Anuradha Godavarty y col. "Optical imaging for breast cancer prescreening". En: *Breast Cancer: Targets and Therapy* 7 (2015), pág. 193.
- [18] Deborah A Kennedy, Tanya Lee y Dugald Seely. "A comparative review of thermography as a breast cancer screening technique". En: *Integrative cancer therapies* 8.1 (2009), págs. 9-16.
- [19] L. Aibar y col. "Clasificación radiológica y manejo de las lesiones mamarias". En: *Clin Invest Ginecol Obstet* (), págs. 141-149. ISSN: 0210-573X. DOI: [10.1016/j.gine.2010.10.016](https://doi.org/10.1016/j.gine.2010.10.016). URL: <http://www.elsevier.es/es-revista-clinica-e-investigacion-ginecologia-obstetricia-7-articulo-clasificacion-radiologica-manejo-las-lesiones-S0210573X10001231>.
- [20] E. Y.-K. Ng y col. "Computerized detection of breast cancer with artificial intelligence and thermograms". En: *Journal of Medical Engineering & Technology* 26.4 (ene. de 2002), págs. 152-157. ISSN: 0309-1902. DOI: [10.1080/03091900210146941](https://doi.org/10.1080/03091900210146941). URL: <http://dx.doi.org/10.1080/03091900210146941>.
- [21] Mohamed Abdel-Nasser, Antonio Moreno y Domenech Puig. "Breast Cancer Detection in Thermal Infrared Images Using Representation Learning and Texture Analysis Methods". En: *Electronics* 8.1 (ene. de 2019), pág. 100. DOI: [10.3390/electronics8010100](https://doi.org/10.3390/electronics8010100). URL: <https://www.mdpi.com/2079-9292/8/1/100>.
- [22] Thomas M. Kolb, Jacob Lichy y Jeffrey H. Newhouse. "Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations". En: *Radiology* 225.1 (2002). PMID: 12355001, págs. 165-175. DOI: [10.1148/radiol.2251011667](https://doi.org/10.1148/radiol.2251011667). eprint: <https://doi.org/10.1148/radiol.2251011667>. URL: <https://doi.org/10.1148/radiol.2251011667>.
- [23] C.K. Kuhl, W. Kuhn y H. Schild. "Management of women at high risk for breast cancer: New imaging beyond mammography". En: *The Breast* 14.6 (2005). 9th International Conference on Primary Therapy of Early Breast Cancer, St Gallen, Switzerland, págs. 480-486. ISSN: 0960-9776. DOI: <https://doi.org/10.1016/j.breast.2005.08.005>. URL: <http://www.sciencedirect.com/science/article/pii/S096097760500192X>.
- [24] Karla Kerlikowske y col. "Effect of Age, Breast Density, and Family History on the Sensitivity of First Screening Mammography". en. En: *JAMA* 276.1 (jul. de 1996), págs. 33-38. ISSN: 0098-7484. DOI: [10.1001/jama.1996.03540010035027](https://doi.org/10.1001/jama.1996.03540010035027). URL: <https://jamanetwork.com/journals/jama/fullarticle/404660>.
- [25] Jennifer S. Drukteinis y col. "Beyond Mammography: New Frontiers in Breast Cancer Screening". en. En: *The American Journal of Medicine* 126.6 (jun. de 2013), págs. 472-479. ISSN: 00029343. DOI: [10.1016/j.amjmed.2012.11.025](https://doi.org/10.1016/j.amjmed.2012.11.025). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0002934313000910>.

- [26] Wendie A. Berg y col. “Combined Screening With Ultrasound and Mammography vs Mammography Alone in Women at Elevated Risk of Breast Cancer”. En: *JAMA* 299.18 (14 de mayo de 2008). Publisher: American Medical Association, págs. 2151-2163. ISSN: 0098-7484. DOI: [10.1001/jama.299.18.2151](https://doi.org/10.1001/jama.299.18.2151). URL: <https://jamanetwork.com/journals/jama/fullarticle/181896>.
- [27] Anouk Pijpe y col. “Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (GENE-RAD-RISK)”. En: *BMJ* 345 (2012). DOI: [10.1136/bmj.e5660](https://doi.org/10.1136/bmj.e5660). eprint: <https://www.bmj.com/content/345/bmj.e5660.full.pdf>. URL: <https://www.bmj.com/content/345/bmj.e5660>.
- [28] Kevin M. Kelly y col. “Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts”. En: *Eur Radiol* 20.3 (1 de mar. de 2010), págs. 734-742. ISSN: 1432-1084. DOI: [10.1007/s00330-009-1588-y](https://doi.org/10.1007/s00330-009-1588-y). URL: <https://doi.org/10.1007/s00330-009-1588-y>.
- [29] Christiane K. Kuhl y col. “Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer”. En: *J. Clin. Oncol.* 23.33 (20 de nov. de 2005), págs. 8469-8476. ISSN: 0732-183X. DOI: [10.1200/JCO.2004.00.4960](https://doi.org/10.1200/JCO.2004.00.4960).
- [30] Shao-Pow Lin y Jeffrey J. Brown. “MR contrast agents: Physical and pharmacologic basics”. En: *J. Magn. Reson. Imaging* 25.5 (mayo de 2007), págs. 884-899. ISSN: 10531807, 15222586. DOI: [10.1002/jmri.20955](https://doi.org/10.1002/jmri.20955). URL: <http://doi.wiley.com/10.1002/jmri.20955>.
- [31] Ingolf Griebisch y col. “Cost-effectiveness of screening with contrast enhanced magnetic resonance imaging vs X-ray mammography of women at a high familial risk of breast cancer”. En: *British journal of cancer* 95.7 (2006), págs. 801-810.
- [32] *Grado de un tumor*. cvgFactSheet. URL: <https://www.cancer.gov/espanol/cancer/diagnostico-estadificacion/pronostico/hoja-informativa-grado-tumor>.
- [33] Lidia Torres Aja. “Metástasis cutánea de carcinoma mamario: cáncer de mama en coraza. Revisión de la literatura y presentación de un caso”. es. En: *Revista Finlay* 2.3 (sep. de 2012), págs. 215-220. ISSN: 2221-2434. URL: <http://www.revfinlay.sld.cu/index.php/finlay/article/view/121>.
- [34] *Etapas del cáncer de seno*. URL: <https://www.cancer.org/es/cancer/cancer-de-seno/comprencion-de-un-diagnostico-de-cancer-de-seno/etapas-del-cancer-de-seno.html>.
- [35] Carl August Wunderlich. *On the temperature in diseases: a manual of medical thermometry*. Vol. 49. New Sydenham Society, 1871.
- [36] OP-TEC. “Infrared Systems for homeland Security”. En: SPIE. OP-TEC, 2008. ISBN: 1-57837-486-3.
- [37] OP-TEC. *Basics of Spectroscopy : Photonics-Enabled Technologies*. OP-TEC, ene. de 2008. ISBN: 978-1-57837-501-1.
- [38] *Espectro electromagnético*. URL: <https://iie.fing.edu.uy/proyectos/esopo/eem/>.
- [39] James M Palmer y Barbara G Grant. “The art of radiometry”. En: SPIE. 2009.
- [40] Norio Ookawa. “Infrared camera and infrared camera system”. Pat. europea 1051031A1. Mitsubishi Electric Corp. Library Catalog: Google Patents. 8 de nov. de 2000. URL: <https://patents.google.com/patent/EP1051031A1/en> (visitado 22-06-2020).
- [41] B. B. Lahiri y col. “Medical applications of infrared thermography: A review”. En: *Infrared Physics & Technology* 55.4 (jul. de 2012), págs. 221-235. ISSN: 1350-4495. DOI: [10.1016/j.infrared.2012.03.007](https://doi.org/10.1016/j.infrared.2012.03.007). URL: <http://www.sciencedirect.com/science/article/pii/S1350449512000308>.

- [42] F. J. González. “Thermal simulation of breast tumors”. En: *Revista mexicana de física* 53.4 (ago. de 2007), págs. 323-326. ISSN: 0035-001X. URL: [http://www.scielo.org.mx/scielo.php?script=sci\\_abstract&pid=S0035-001X2007000400014&lng=es&nrm=iso&tlng=en](http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0035-001X2007000400014&lng=es&nrm=iso&tlng=en).
- [43] *Thermography fundamentals*. Library Catalog: www.fluke.com. URL: <https://www.fluke.com/en-us/learn/best-practices/measurement-basics/thermography> (visitado 22-06-2020).
- [44] Francisco J. González. “Theoretical and clinical aspects of the use of thermography in non-invasive medical diagnosis”. En: *Biomedical Spectroscopy and Imaging* 5.4 (oct. de 2016), págs. 347-358. ISSN: 22128808, 22128794. DOI: [10.3233/BSI-160152](https://doi.org/10.3233/BSI-160152). URL: <http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/BSI-160152>.
- [45] J. R. Keyserlingk y col. “Functional infrared imaging of the breast”. En: *IEEE Engineering in Medicine and Biology Magazine* 19.3 (mayo de 2000), págs. 30-41. ISSN: 0739-5175. DOI: [10.1109/51.844378](https://doi.org/10.1109/51.844378).
- [46] M. Gautherie y C. M. Gros. “Breast thermography and cancer risk prediction”. eng. En: *Cancer* 45.1 (ene. de 1980), págs. 51-56. ISSN: 0008-543X.
- [47] Thomas D. Vreugdenburg y col. “A systematic review of elastography, electrical impedance scanning, and digital infrared thermography for breast cancer screening and diagnosis”. En: *Breast Cancer Research and Treatment* 3 (2013), pág. 665. ISSN: 0167-6806. DOI: [10.1007/s10549-012-2393-x](https://doi.org/10.1007/s10549-012-2393-x).
- [48] Muffazzal Rassiwalá y col. “Evaluation of digital infra-red thermal imaging as an adjunctive screening method for breast carcinoma: A pilot study”. En: *International Journal of Surgery* 12.12 (2014), págs. 1439-1443.
- [49] Y. R. Parisky y col. “Efficacy of Computerized Infrared Imaging Analysis to Evaluate Mammographically Suspicious Lesions”. En: *American Journal of Roentgenology* 180.1 (ene. de 2003), págs. 263-269. ISSN: 0361-803X. DOI: [10.2214/ajr.180.1.1800263](https://doi.org/10.2214/ajr.180.1.1800263). URL: <https://www.ajronline.org/doi/10.2214/ajr.180.1.1800263>.
- [50] Xianwu Tang y col. “Morphological measurement of localized temperature increase amplitudes in breast infrared thermograms and its clinical application”. En: *Biomedical Signal Processing and Control* 3.4 (oct. de 2008), págs. 312-318. ISSN: 1746-8094. DOI: [10.1016/j.bspc.2008.04.001](https://doi.org/10.1016/j.bspc.2008.04.001). URL: <http://www.sciencedirect.com/science/article/pii/S1746809408000347>.
- [51] Nimmi Arora y col. “Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer”. eng. En: *Am. J. Surg.* 196.4 (oct. de 2008), págs. 523-526. ISSN: 1879-1883. DOI: [10.1016/j.amjsurg.2008.06.015](https://doi.org/10.1016/j.amjsurg.2008.06.015).
- [52] M. Kontos, R. Wilson e I. Fentiman. “Digital infrared thermal imaging (DITI) of breast lesions: sensitivity and specificity of detection of primary breast cancers”. En: *Clinical Radiology* 66.6 (jun. de 2011), págs. 536-539. ISSN: 0009-9260. DOI: [10.1016/j.crad.2011.01.009](https://doi.org/10.1016/j.crad.2011.01.009). URL: <http://www.sciencedirect.com/science/article/pii/S0009926011000699>.
- [53] Kurt Ammer. “The Glamorgan Protocol for recording and evaluation of thermal images of the human body”. En: *Thermology international* 18.4 (2008), págs. 125-144.
- [54] Marco Antonio Garduño-Ramón y col. “Supportive Noninvasive Tool for the Diagnosis of Breast Cancer Using a Thermographic Camera as Sensor”. En: *Sensors (Basel, Switzerland)* 17.3 (mar. de 2017). ISSN: 1424-8220. DOI: [10.3390/s17030497](https://doi.org/10.3390/s17030497). URL: <http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=28273793&lang=es&site=eds-live>.
- [55] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. “Deep learning”. en. En: *Nature* 521.7553 (mayo de 2015), págs. 436-444. ISSN: 0028-0836. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539). URL: <https://www.nature.com/nature/journal/v521/n7553/full/nature14539.html#access>.
- [56] Andrew Ng. *Machine Learning*. 2012. URL: <https://www.coursera.org/learn/machine-learning>.

- [57] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [58] Arthur L. Samuel. "Some Studies in Machine Learning Using the Game of Checkers". En: *IBM Journal of Research and Development* 3 (1959), págs. 210-229. DOI: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210).
- [59] Andrew Ng. "Machine learning yearning". En: URL: [http://www.mlyearning.org/\(96\)](http://www.mlyearning.org/(96)) (2017).
- [60] A. K. Jain, M. N. Murty y P. J. Flynn. "Data Clustering: A Review". En: *ACM Comput. Surv.* 31.3 (sep. de 1999), págs. 264-323. ISSN: 0360-0300. DOI: [10.1145/331499.331504](https://doi.org/10.1145/331499.331504). URL: <http://doi.acm.org/10.1145/331499.331504>.
- [61] Longlong Jing y Yingli Tian. "Self-supervised visual feature learning with deep neural networks: A survey". En: *arXiv preprint arXiv:1902.06162* (2019).
- [62] Sotiris B Kotsiantis, Ioannis D Zaharakis y Panayiotis E Pintelas. "Machine learning: a review of classification and combining techniques". En: *Artificial Intelligence Review* 26.3 (2006), págs. 159-190.
- [63] Stanley I Grossman. *Álgebra lineal*. McGraw Hill Educación, 2008.
- [64] Ron Larson y col. *Cálculo*. McGraw-Hill, 2006.
- [65] *Deep Learning vs Machine Learning*. Library Catalog: explore.mathworks.com. URL: <https://explore.mathworks.com/machine-learning-vs-deep-learning>.
- [66] J. V. Tu. "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes". eng. En: *J Clin Epidemiol* 49.11 (nov. de 1996), págs. 1225-1231. ISSN: 0895-4356.
- [67] Rich Caruana, Steve Lawrence y C Lee Giles. "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping". En: *Advances in neural information processing systems*. 2001, págs. 402-408.
- [68] SAGAR SHARMA. *Activation Functions in Neural Networks*. en. Feb. de 2019. URL: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- [69] Bing Xu y col. "Empirical evaluation of rectified activations in convolutional network". En: *arXiv preprint arXiv:1505.00853* (2015).
- [70] Dan Becker. *Introduction to Deep Learning in Python*. URL: <https://www.datacamp.com/courses/introduction-to-deep-learning-in-python>.
- [71] George Musser y col. *Does autism arise because the brain is continually surprised?* Science | AAAS. Library Catalog: www.sciencemag.org. 9 de mar. de 2018. URL: <https://www.sciencemag.org/news/2018/03/does-autism-arise-because-brain-continually-surprised>.
- [72] Nitish Srivastava y col. "Dropout: a simple way to prevent neural networks from overfitting." En: *Journal of Machine Learning Research* 15.1 (2014), págs. 1929-1958. URL: <http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf>.
- [73] Feng Ning y col. "Toward automatic phenotyping of developing embryos from videos". En: *IEEE Transactions on Image Processing* 14.9 (2005), págs. 1360-1371.
- [74] Srinivas C Turaga y col. "Convolutional networks can learn to generate affinity graphs for image segmentation". En: *Neural computation* 22.2 (2010), págs. 511-538.
- [75] *4. Major Architectures of Deep Networks - Deep Learning [Book]*. Library Catalog: www.oreilly.com. URL: <https://www.oreilly.com/library/view/deep-learning/9781491924570/ch04.html> (visitado 27-03-2020).
- [76] *List of Deep Learning Layers - MATLAB & Simulink*. URL: <https://www.mathworks.com/help/deeplearning/ug/list-of-deep-learning-layers.html> (visitado 27-03-2020).

- [77] Alex Krizhevsky, Ilya Sutskever y Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. En: *Commun. ACM* 60.6 (mayo de 2017), págs. 84-90. ISSN: 0001-0782. DOI: [10.1145/3065386](https://doi.org/10.1145/3065386). URL: <https://doi.org/10.1145/3065386>.
- [78] *ImageNet*. URL: <http://www.image-net.org/>.
- [79] Christian Szegedy y col. “Going deeper with convolutions”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, págs. 1-9.
- [80] Sanjeev Arora y col. “Provable bounds for learning some deep representations”. En: *International Conference on Machine Learning*. 2014, págs. 584-592.
- [81] *Places: A 10 million Image Database for Scene Recognition*. URL: <http://places2.csail.mit.edu/>.
- [82] Bolei Zhou y col. “Places: An Image Database for Deep Scene Understanding”. En: *arXiv:1610.02055 [cs]* (6 de oct. de 2016). arXiv: [1610.02055](https://arxiv.org/abs/1610.02055). URL: <http://arxiv.org/abs/1610.02055>.
- [83] Christian Szegedy y col. “Rethinking the inception architecture for computer vision”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 2818-2826.
- [84] Santosh Giri y Basanta Joshi. “Transfer Learning Based Image Visualization Using CNN”. En: *International Journal of Artificial Intelligence and Application* 10.4 (2019), págs. 47-55.
- [85] Karen Simonyan y Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. En: *arXiv preprint arXiv:1409.1556* (2014).
- [86] *Review: VGGNet — 1st Runner-Up (Image Classification), Winner (Localization) in ILSVRC 2014 | by Sik-Ho Tsang | Coinmonks | Medium*. URL: <https://medium.com/coinmonks/paper-review-of-vggnet-1st-runner-up-of-ilsvrc-2014-image-classification-d02355543a11> (visitado 25-11-2020).
- [87] Kaiming He y col. “Deep residual learning for image recognition”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778.
- [88] *7.6. Residual Networks (ResNet) — Dive into Deep Learning 0.14.3 documentation*. URL: [https://d21.ai/chapter\\_convolutional-modern/resnet.html](https://d21.ai/chapter_convolutional-modern/resnet.html) (visitado 25-08-2020).
- [89] *Train Deep Learning Network to Classify New Images*. URL: <https://www.mathworks.com/examples/neural-network/mw/nnet-ex74362329-transfer-learning-using-googlenet>.
- [90] S. J. Pan y Q. Yang. “A Survey on Transfer Learning”. En: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (oct. de 2010), págs. 1345-1359. ISSN: 1041-4347. DOI: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- [91] *Pretrained GoogLeNet convolutional neural network - MATLAB googlenet*. URL: <https://www.mathworks.com/help/deeplearning/ref/googlenet.html>.
- [92] Luke Taylor y Geoff Nitschke. “Improving Deep Learning using Generic Data Augmentation”. En: *arXiv:1708.06020 [cs, stat]* (ago. de 2017). arXiv: 1708.06020. URL: <http://arxiv.org/abs/1708.06020>.
- [93] Bofan Song y col. “Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning”. EN. En: *Biomed. Opt. Express, BOE* 9.11 (nov. de 2018), págs. 5318-5329. ISSN: 2156-7085. DOI: [10.1364/BOE.9.005318](https://doi.org/10.1364/BOE.9.005318). URL: <https://www.osapublishing.org/boe/abstract.cfm?uri=boe-9-11-5318>.
- [94] Luis Perez y Jason Wang. “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”. En: *arXiv:1712.04621 [cs]* (dic. de 2017). arXiv: 1712.04621. URL: <http://arxiv.org/abs/1712.04621>.
- [95] Ali Madani y col. “Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease”. En: *npj Digital Medicine* 1.1 (oct. de 2018), pág. 59. ISSN: 2398-6352. DOI: [10.1038/s41746-018-0065-x](https://doi.org/10.1038/s41746-018-0065-x). URL: <https://www.nature.com/articles/s41746-018-0065-x>.

- [96] Abdul Ghaaliq Lalkhen y Anthony McCluskey. “Clinical tests: sensitivity and specificity”. En: *Continuing Education in Anaesthesia Critical Care & Pain* 8.6 (dic. de 2008), págs. 221-223. ISSN: 1743-1816. DOI: [10.1093/bjaceaccp/mkn041](https://doi.org/10.1093/bjaceaccp/mkn041). URL: <http://www.sciencedirect.com/science/article/pii/S1743181617304572>.
- [97] Sirko Straube y Mario M. Krell. “How to evaluate an agent’s behavior to infrequent events?—Reliable performance estimation insensitive to class distribution”. English. En: *Front. Comput. Neurosci.* 8 (2014). ISSN: 1662-5188. DOI: [10.3389/fncom.2014.00043](https://doi.org/10.3389/fncom.2014.00043). URL: <https://www.frontiersin.org/articles/10.3389/fncom.2014.00043/full>.
- [98] M. Sirsat. *Data Science and Machine Learning : Confusion Matrix*. Data Science and Machine Learning. Library Catalog: Blogger. 29 de abr. de 2019. URL: <https://manisha-sirsat.blogspot.com/2019/04/confusion-matrix.html>.
- [99] Andre Esteva y col. “Dermatologist-level classification of skin cancer with deep neural networks”. en. En: *Nature* 542.7639 (feb. de 2017), págs. 115-118. ISSN: 0028-0836. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056). URL: <https://www.nature.com/nature/journal/v542/n7639/full/nature21056.html>.
- [100] Myura Nagendran y col. “Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies”. En: *bmj* 368 (2020).
- [101] L. F. Silva y col. “A New Database for Breast Research with Infrared Image”. En: *Journal of Medical Imaging and Health Informatics* 4.1 (mar. de 2014), págs. 92-100. DOI: [10.1166/jmih.2014.1226](https://doi.org/10.1166/jmih.2014.1226).
- [102] J. C. Torres-Galván, E. Guevara y F. J. González. “Comparison of Deep Learning Architectures for Pre-Screening of Breast Cancer Thermograms”. En: *2019 Photonics North (PN)*. 2019 Photonics North (PN). Vol. CFP1909V-ART. Mayo de 2019, págs. 1-2. DOI: [10.1109/PN.2019.8819587](https://doi.org/10.1109/PN.2019.8819587).
- [103] William Navidi. *Estadística para ingenieros*. 519.5 N325. McGraw Hill Interamericana, 2006. ISBN: 970-105-62-99.
- [104] R.E. Walpole. *Probabilidad Y Estadística Para Ingeniería Y Ciencias*. Pearson Educación, 2007. ISBN: 978-970-26-0936-0.
- [105] Murray Spiegel y Larry Stephens. *Estadística schaum*. México DF: McGraw Hill, 2009.
- [106] Anton S. Becker y col. “Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer”. ENGLISH. En: *Investigative Radiology* 52.7 (jul. de 2017), págs. 434-440. ISSN: 0020-9996. DOI: [10.1097/RLI.0000000000000358](https://doi.org/10.1097/RLI.0000000000000358). URL: <https://insights.ovid.com/pubmed?pmid=28212138> (visitado 29-10-2017).
- [107] Ramesh Omranipour y col. “Comparison of the Accuracy of Thermography and Mammography in the Detection of Breast Cancer”. eng. En: *Breast Care (Basel)* 11.4 (ago. de 2016), págs. 260-264. ISSN: 1661-3791. DOI: [10.1159/000448347](https://doi.org/10.1159/000448347).
- [108] Terry M. Button y col. “Dynamic infrared imaging for the detection of malignancy”. eng. En: *Phys Med Biol* 49.14 (jul. de 2004), págs. 3105-3116. ISSN: 0031-9155.
- [109] Rinaldo Roberto de Jesus Guirro y col. “Accuracy and Reliability of Infrared Thermography in Assessment of the Breasts of Women Affected by Cancer”. en. En: *Journal of Medical Systems* 41.5 (mayo de 2017). ISSN: 0148-5598, 1573-689X. DOI: [10.1007/s10916-017-0730-7](https://doi.org/10.1007/s10916-017-0730-7). URL: <http://link.springer.com/10.1007/s10916-017-0730-7> (visitado 25-06-2017).
- [110] B. Krawczyk y G. Schaefer. “Breast Thermogram Analysis Using Classifier Ensembles and Image Symmetry Features”. En: *IEEE Systems Journal* 8.3 (sep. de 2014), págs. 921-928. ISSN: 1932-8184. DOI: [10.1109/JSYST.2013.2283135](https://doi.org/10.1109/JSYST.2013.2283135).

- [111] Antony Morales-Cervantes y col. “An automated method for the evaluation of breast cancer using infrared thermography”. En: *EXCLI J* 17 (oct. de 2018), págs. 989-998. ISSN: 1611-2156. DOI: [10.17179/excli2018-1735](https://doi.org/10.17179/excli2018-1735). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6295635/> (visitado 14-01-2019).
- [112] Tamar Sella y col. “A novel functional infrared imaging system coupled with multiparametric computerised analysis for risk assessment of breast cancer”. eng. En: *Eur Radiol* 23.5 (mayo de 2013), págs. 1191-1198. ISSN: 1432-1084. DOI: [10.1007/s00330-012-2724-7](https://doi.org/10.1007/s00330-012-2724-7).
- [113] G. C. Wishart y col. “The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy”. En: *European Journal of Surgical Oncology (EJSO)* 36.6 (jun. de 2010), págs. 535-540. ISSN: 0748-7983. DOI: [10.1016/j.ejso.2010.04.003](https://doi.org/10.1016/j.ejso.2010.04.003). URL: <http://www.sciencedirect.com/science/article/pii/S0748798310000909> (visitado 14-01-2019).
- [114] Dezső Ribli y col. “Detecting and classifying lesions in mammograms with Deep Learning”. En: *Sci Rep* 8.1 (15 de mar. de 2018). Number: 1 Publisher: Nature Publishing Group, págs. 1-7. ISSN: 2045-2322. DOI: [10.1038/s41598-018-22437-z](https://doi.org/10.1038/s41598-018-22437-z). URL: <https://www.nature.com/articles/s41598-018-22437-z>.
- [115] Linda Nordling. “A fairer way forward for AI in health care”. en. En: *Nature* 573 (sep. de 2019), S103-S105. DOI: [10.1038/d41586-019-02872-2](https://doi.org/10.1038/d41586-019-02872-2). URL: <http://www.nature.com/articles/d41586-019-02872-2>.
- [116] Yun Liu y col. “How to read articles that use machine learning: Users’ Guides to the Medical Literature”. En: *Jama* 322.18 (2019), págs. 1806-1816.
- [117] Amirhossein Kiani y col. “Impact of a deep learning assistant on the histopathologic classification of liver cancer”. En: *NPJ digital medicine* 3.1 (2020), págs. 1-8.