

INSTITUTO TECNOLÓGICO DE CHIHUAHUA
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

***“ANÁLISIS DE TEJIDOS EN MUESTRAS
HISTOLÓGICAS USANDO REDES NEURONALES
CONVOLUCIONALES”***

TESIS

QUE PARA OBTENER EL GRADO DE

***MAESTRO EN CIENCIAS EN INGENIERÍA
ELECTRÓNICA***

PRESENTA:

ING. IVAN ESTEBAN ACOSTA LARA

**DIRECTOR DE LA TESIS:
DR. JUAN ALBERTO RAMÍREZ QUINTANA**

**CODIRECTOR DE LA TESIS:
DR. MARIO IGNACIO CHACÓN MURGUÍA**

CHIHUAHUA, CHIH., SEPTIMBRE 2021



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO





Chihuahua, Chih. 1 de septiembre de 2021

M.C. ROGELIO ENRIQUE BARAY ARANA
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
PRESENTE

En cumplimiento con los requerimientos para la obtención del grado de Maestro en Ciencias en Ingeniería Electrónica, le notificamos que el documento de tesis del alumno C. IVÁN ESTEBAN ACOSTA LARA, titulado "*Análisis de tejidos en muestras histológicas usando redes neuronales convolucionales*" dirigido por Dr. Juan Alberto Ramírez Quintana ha sido aprobado y aceptado para su impresión.

Por lo anterior, proponemos le sea concedida la autorización de impresión correspondiente.

Agradeciendo la atención a la presente, quedamos de usted:

ATENTAMENTE

Excelencia en Educación Tecnológica®
"La Técnica por el Engrandecimiento de México"

Juan Ramírez
DR. JUAN ALBERTO RAMÍREZ QUINTANA
DIRECTOR DE TESIS

Miguel
DR. MARIO IGNACIO CHACÓN MURGUÍA
MIEMBRO DEL COMITÉ TUTORIAL

Alma Delia Corral Sáenz
MTRA. ALMA DELIA CORRAL SÁENZ
MIEMBRO DEL COMITÉ TUTORIAL

Carlos Arzate Q.
MTRO. CARLOS ARZATE QUINTANA
MIEMBRO DEL COMITÉ TUTORIAL





Chihuahua, Chih. 6 de septiembre de 2021

**C. IVÁN ESTEBAN ACOSTA LARA
PRESENTE**

En cumplimiento con los requerimientos para la obtención del grado de Maestro en Ciencias en Ingeniería Electrónica y a propuesta de su Comité Tutorial, la División de Estudios de Posgrado e Investigación le concede la autorización para imprimir la tesis titulada *"Análisis de tejidos en muestras histológicas usando redes neuronales convolucionales"* dirigida por Dr. Juan Alberto Ramírez Quintana con el siguiente contenido de capítulos:

I Antecedentes

II Análisis de los modelos en el estado del arte de los métodos de procesamiento de imágenes de histología

III Redes neuronales convolucionales para el análisis de imágenes de histología

IV Preprocesamiento de datos y redes propuestas

V Resultados

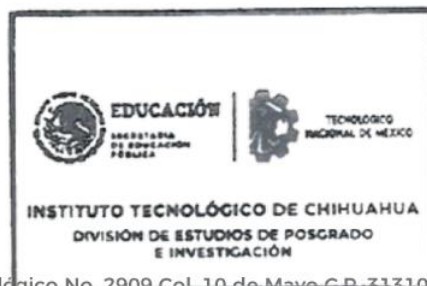
VI Conclusiones y trabajo a futuro

ATENTAMENTE

Excelencia en Educación Tecnológica®

"La Técnica por el Engrandecimiento de México"

M.C. ROGELIO ENRIQUE BARAY ARANA
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN



Av. Tecnológico No. 2909 Col. 10 de Mayo C.P. 31310, Chihuahua, Chih.,

Tel. (614) 201 2000, (614)413 5187, Ext. 2157 e-mail: dir_chihuahua@tecnm.mx

tecnm.mx | itchihuahua.edu.mx





EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Chihuahua

CARTA CESIÓN DE DERECHOS

En la ciudad de Chihuahua el día 6 de septiembre de 2021 10 de septiembre de 2021, el que suscribe C. IVÁN ESTEBAN ACOSTA LARA con número de control G19060581, de la Maestría en Ciencias en Ingeniería Electrónica, adscrita a la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Juan Alberto Ramírez Quintana y cede los derechos del trabajo titulado "Análisis de tejidos en muestras histológicas usando redes neuronales convolucionales", al Tecnológico Nacional de México y/o Instituto Tecnológico de Chihuahua para su difusión, divulgación, transmisión, reproducción, así como su digitalización con fines académicos y de investigación.

Ivan E. Acosta

C. IVÁN ESTEBAN ACOSTA LARA



INSTITUTO TECNOLÓGICO
DE CHIHUAHUA



Av. Tecnológico No. 2909 Col. 10 de Mayo C.P. 51310, Chihuahua, Chih.

Tel. (614) 201 2000. (614) 413 5187, Ext. 2157 e-mail: dir_chihuahua@tecnm.mx

tecnm.mx | itchihuahua.edu.mx





EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Chihuahua

DECLARACIÓN DE ORIGINALIDAD

En la ciudad de Chihuahua el día 6 de septiembre de 2021 10 de septiembre de 2021, el que suscribe C. IVÁN ESTEBAN ACOSTA LARA de la Maestría en Ciencias en Ingeniería Electrónica, con número de control G19060581, adscrito a la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua, manifiesta que es autor intelectual de la tesis titulada "Análisis de tejidos en muestras histológicas usando redes neuronales convolucionales" bajo la dirección Dr. Juan Alberto Ramírez Quintana: que el contenido es original y que las fuentes de información consultadas para su fundamentación están debidamente citadas y referenciadas.

Ivan E. Acosta

C. IVÁN ESTEBAN ACOSTA LARA



Av. Tecnológico No. 2909 Col. 10 de Mayo C.P. 31310, Chihuahua, Chih.,

Tel. (614) 201 2000, (614) 413 5187, Ext. 2157 e-mail. dir_chihuahua@tecnm.mx

tecnm.mx | itchihuahua.edu.mx



Chihuahua, Chihuahua a 7 de septiembre de 2021

Dra. María Elena Álvarez-Buylla Roces

Director de CONACYT.

Presente

Por este conducto aprovecho la ocasión para saludarlo e informarle que a la fecha he obtenido el grado de Maestría en Ciencias en Ingeniería Electrónica en la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua. Motivo por el cual agradezco todo el apoyo brindado por esta Institución que usted representa, el otorgamiento de esta beca permitió dedicarme de tiempo completo a la realización de mis estudios de Posgrado y de esta manera lograr el cumplimiento del objetivo principal del convenio establecido.

Sin otro particular por el momento, me es grato quedar de usted como su seguro servidor, no sin antes reiterar mi agradecimiento. Muchas Gracias.

Atentamente

Ivan E. Acosta

Ivan Esteban Acosta Lara

Exbecario CONACYT

Agradecimientos

Agradezco al Tecnológico Nacional de México por el financiamiento del proyecto “Clasificación de parásitos en muestras de copro mediante aprendizaje de máquina”, con clave 10071.21-P. Agradezco al CONACYT por el apoyo económico brindado durante mis estudios. Agradezco al personal de la DEPI por su apoyo. Agradezco a mi director de tesis el Dr. Juan A. Ramírez por su asesoramiento, paciencia y dedicación. Agradezco a mi codirector de tesis, Dr. Mario I. Chacón y al Comité Tutorial, M.C. Alma Corral y Dr. Carlos Arzate por su apoyo en este trabajo de tesis. Agradezco a mis compañeros y amigos de generación Javier Álvarez y Alonso Reyes por compartir sus conocimientos. Agradezco a mis compañeros de laboratorio, Alejandro Camacho, David Hernández, Abimael Guzmán, Gerardo Esquivel, Luis Madrid, José Macías, y Eduardo Rivas, por el ambiente y convivencia. Agradezco a mis papás y hermanos por su apoyo incondicional. Agradezco a mi compañero Oscar Soto por platicarme acerca de la maestría y animarme a estudiarla. Agradezco a mis papás y hermanos por su apoyo incondicional. Y agradezco a todas aquellas personas que indirectamente ayudaron a realizar este trabajo de tesis.

RESUMEN

ANÁLISIS DE TEJIDOS EN MUESTRAS HISTOLÓGICAS USANDO REDES NEURONALES CONVOLUCIONALES

Ing. Ivan Esteban Acosta Lara

Maestría en Ciencias en Ingeniería Electrónica

División de Estudios de Posgrado e Investigación

Instituto Tecnológico de Chihuahua

Chihuahua, Chih. 2021

Director de tesis: Dr. Juan Alberto Ramírez Quintana

Codirector de tesis: Dr. Mario Ignacio Chacón Murguía

El análisis histológico se refiere al estudio de muestras de cortes de tejido humano para estudiar la microestructura de los materiales biológicos por medio de la visualización por microscopio. Estas visualizaciones pueden realizarse por medio de imágenes digitales para su posterior análisis por medio de procesadores especializados para catalogar diferentes patologías. Por ello, en la literatura existen diversos métodos de procesamiento para la clasificación de imágenes de histología, tales como con redes neuronales artificiales, reconocimiento de patrones, aprendizaje profundo, etc.

Entre las muchas aplicaciones del procesamiento de imágenes de histología, se encuentra en análisis de tejido mamario, que se refiere a la clasificación de tumores malignos y benignos. Sin embargo, estos métodos se centran en métricas de desempeño, sin considerar el costo computacional. Por ello, el objetivo de esta tesis es proponer un modelo de red neuronal artificial profunda de bajo costo computacional que clasifique imágenes de cortes histológicos de mama y las catalogue en tumores benignos y malignos. La red propuesta se denomina

LightHist y se basa en la arquitectura MobileNet-v2, pero con convoluciones separables en profundidad en paralelo, con el fin de extraer la mayor cantidad de características útiles para la clasificación. Los resultados demostraron que LightHist tiene buen desempeño utilizando las métricas de evaluación y una menor cantidad de parámetros en comparación con las redes existentes en la literatura. De esta manera, a partir de LightHist se puede realizar análisis histológico de tejido mamario en circuitos integrados de función específica (ASIC, por sus siglas en inglés) y en unidades de procesamiento gráfico embebidas.

Esta tesis forma parte del proyecto general “Algoritmos de visión artificial para realizar análisis de imágenes digitales para diagnóstico clínico parasitológico, urológico, hematológico y en muestras histológicas de patología” que realiza el Instituto Tecnológico de Chihuahua en conjunto con la Facultad de Medicina y Ciencias Biomédicas de la Universidad Autónoma de Chihuahua, y del proyecto “Clasificación de parásitos en muestras de copro mediante aprendizaje de máquina”, con clave 10071.21-P del Tecnológico Nacional de México.

CONTENIDO

LISTA DE FIGURAS	xii
LISTA DE TABLAS	xiv
CAPÍTULO 1 ANTECEDENTES	1
1.1 Histología.....	1
1.1.1 Método de tinción con hematoxilina y eosina	2
1.1.2 Análisis histológicos.....	3
1.2 Redes Neuronales Artificiales.....	4
1.3 Aprendizaje automático	7
1.3.1 Tipos de aprendizaje	8
1.4 Aprendizaje profundo	8
1.4.1 Redes neuronales convolucionales	9
1.4.2 Capas que componen las CNN	9
1.4.3 Entrenamiento.....	14
1.5 Estado del arte.....	14
1.5.1 Aprendizaje automático	15
1.5.2 Aprendizaje profundo	16
CAPÍTULO 2 ANÁLISIS DE LOS MODELOS EN EL ESTADO DEL ARTE DE LOS MÉTODOS DE PROCESAMIENTO DE IMÁGENES DE HISTOLOGÍA	19
2.1 Modelos utilizados en el estado del arte para clasificación de imágenes histológicas 19	
2.2 Bases de datos	21
2.3 Implementación de algoritmos.....	23
2.4 Métricas de evaluación	24
2.5 Conclusiones del análisis realizado	26
CAPÍTULO 3 REDES NEURONALES CONVOLUCIONALES PARA EL ANÁLISIS DE IMÁGENES DE HISTOLOGÍA	28
3.1 ResNet.....	31

3.2	Redes Inception.....	32
3.3	MobileNet-v2.....	39
3.4	Xception.....	42
CAPÍTULO 4 PREPROCESAMIENTO DE DATOS Y REDES PROPUESTAS		45
4.1	Preprocesamiento.....	45
4.1.1	Normalización de tinción.....	46
4.1.2	Normalización de imágenes.....	47
4.1.3	Aumentación de datos	47
4.2	Redes propuestas.....	54
CAPÍTULO 5 RESULTADOS		58
5.1	Resultados de desempeño de las redes propuestas	58
5.2	Comparación de la red propuesta en comparación con el estado del arte.....	60
CAPÍTULO 6 CONCLUSIONES Y TRABAJO A FUTURO		65
GLOSARIO		67
REFERENCIAS		69

LISTA DE FIGURAS

Figura 1.1 Esquema general de una neurona biológica.	5
Figura 1.2 Esquema general del perceptrón.	6
Figura 1.3 Ejemplo de red neuronal.	7
Figura 1.4 Arquitectura de un neocognitrón.....	10
Figura 1.5 LeNet-5: una de las primeras redes neuronales convolucionales [15].	11
Figura 1.6 Red neuronal convolucional simple.	11
Figura 1.7 Aplanamiento (<i>flatten</i>), transformación de la matriz de características en un vector	13
Figura 2.1 Modelos dedicados a trabajar con imágenes de histología en artículos del periodo 2015-2019.	20
Figura 2.2 Lenguajes utilizaos en los algoritmos de clasificación de imágenes de histología.	23
Figura 2.3 Métodos que realizan una aumentación de datos antes del entrenamiento.	24
Figura 2.4 Validación cruzada de 5-carpetas.....	26
Figura 3.1 Gráfico de bolas que informa la precisión contra la complejidad computacional El tamaño de cada círculo corresponde a la complejidad de la red[61].	29
Figura 3.2 Capas y funciones en los modelos de red.....	30
Figura 3.3 Módulos y bloques en los modelos de red [62].	30
Figura 3.4 Módulos repetidos en la red [62].	31
Figura 3.5 Bloques de la arquitectura de ResNet50 [62].	32
Figura 3.6 Bloques de la arquitectura de ResNet50 [62].	33
Figura 3.7 Módulo Inception simple.[64].	34
Figura 3.8 Arquitectura de la red Inception-v3 [62].	34
Figura 3.9 Módulos de reducción y tallo.	35
Figura 3.10 Módulos Inception-V3 [62].	37
Figura 3.11 Arquitectura Inception-ResNet.[62].	38
Figura 3.12 Módulos de tallo y reducciones para la InceptionResNet [62].	38
Figura 3.13 Módulos <i>Residual Inception</i> [62].	39
Figura 3.14 Convoluciones separables [68].	40
Figura 3.15 Ejemplo de costo de una convolución estándar y una separable en profundidad [67].	41
Figura 3.16 Arquitectura de MobileNet-v2 [62].	42
Figura 3.17 Arquitectura de Xception [62].	43
Figura 3.18 Bloques de la red Xception [62].	44
Figura 4.1 Imágenes tomadas de la base de datos BreakHist.	46

Figura 4.2 Preprocesamiento de imágenes de histología, donde se muestran: a) Imagen original, b) Normalización de intensidad (X_{norm}), c) Normalización de tinción (OD), d) Normalización de tinción e intensidad (OD y X_{norm})	47
Figura 4.3 Imagen de muestra a la cual se le aplicaran las transformaciones.	48
Figura 4.4 Visualización del rango de corte	49
Figura 4.5 Ejemplo de rango de corte.....	49
Figura 4.6 Ejemplos de rango de acercamiento. A la izquierda, una imagen con acercamiento. En el centro una imagen con alejamiento en el eje horizontal. A la derecha una imagen con alejamiento en el eje vertical.	49
Figura 4.7 Ejemplos de desplazamientos horizontales. A la izquierda, un desplazamiento a la derecha. En el centro un desplazamiento a la izquierda. A la derecha un desplazamiento a la derecha menor.....	50
Figura 4.8 Ejemplos de desplazamientos verticales.	51
Figura 4.9 Ejemplos de cambio de brillo. A la izquierda aumento de brillo. En el centro la imagen original. A la derecha disminución de brillo.....	51
Figura 4.10 Ejemplo de volteo de imágenes. A la izquierda volteo horizontal. En el centro la imagen original. A la derecha volteo vertical.	52
Figura 4.11 Ejemplo de rotaciones de imágenes.	52
Figura 4.12 Ejemplo del modo de relleno reflect. a) Relleno reflejo. b) Imagen con corrimiento a la derecha.	53
Figura 4.13 Arquitectura base de las redes LightHist.	54
Figura 4.14 Arquitectura del bloque Inverted Residual Block de Light-hist.	56
Figura 4.15 Arquitectura del bloque Inverted Residual Block de Light-hist-v2.	56
Figura 4.16 Arquitectura del bloque Inverted Residual Block de Light-hist-v3.	57

LISTA DE TABLAS

Tabla 2.1 Distribución por factor de magnificación y clase.....	22
Tabla 4.1 Parámetros de aumentación de datos.....	53
Tabla 5.1 Resultados de las redes propuestas con las imágenes de pruebas a 70 épocas.....	58
Tabla 5.2 Métricas de las redes propuestas con las imágenes de pruebas a 70 épocas	59
Tabla 5.3 Resultados de las redes propuestas con las imágenes de pruebas a 1000 épocas.....	59
Tabla 5.4 Métricas de las redes propuestas con las imágenes de pruebas a 1000 épocas.	60
Tabla 5.5 Resultados LightHist-v2 usando validación cruzada a 70 épocas.....	60
Tabla 5.6 Comparación de la red propuesto con los existentes en el estado del arte.	61
Tabla 5.7 Resultados de Xception usando validación cruzada a 70 épocas.	63
Tabla 5.8 Resultados de MobileNet-v2 usando validación cruzada a 70 épocas.	63

CAPÍTULO 1

ANTECEDENTES

El análisis de tejidos mediante procesamiento inteligente de imágenes de histología es un área interdisciplinaria que requiere conocimientos de diagnóstico médico y aprendizaje automático. Por ello, este capítulo presenta conceptos elementales de ambas disciplinas y se organiza de la siguiente manera: la sección 1.1 describe el concepto de histología, así como el proceso para obtener imágenes digitales de muestras de tejido humano. La sección 1.2 describe las redes neuronales artificiales y su utilidad para la clasificación de imágenes de histología como herramienta del diagnóstico médico.

1.1 Histología

La histología es el estudio científico de las estructuras microscópicas de la biología molecular, las funciones celulares, los tejidos y los órganos del cuerpo [1]. Para el análisis de tejidos por medio de la histología, es necesario prepararlos adecuadamente para su estudio mediante técnicas de tinción, que consisten en aplicar un colorante a un tejido para su análisis. La tinción con hematoxilina y eosina con fijación en formalina es la más utilizada, ya que pone en evidencia las características estructurales del tejido a estudiar. No obstante, este procedimiento tiene pérdida visual de elementos como elastina, fibras reticulares, membranas basales y lípidos. En estos casos, se utilizan otros métodos de tinción que puedan realzar sus características si se desea observarlas [1].

Una aplicación de la histología es la patología, que es el estudio científico de las estructuras de los órganos y función del cuerpo ante la presencia de una enfermedad. Esta se divide en patología general, que estudia los principios generales de las enfermedades, y patología sistemática que se enfoca en el estudio de enfermedades en órganos y sistemas [2]. Ambas, recurren a la histología para encontrar enfermedades y padecimientos que se pueden medir con cambios estructurales observados a simple vista y/o detectados mediante microscopía (lo que se denomina como histopatología) [3].

Para el análisis de tejidos de histología e histopatología, es común el uso de imágenes digitales de tejido celular que se preparan mediante métodos de tinción basados en hematoxilina y eosina, los cuales se describen a continuación.

1.1.1 Método de tinción con hematoxilina y eosina

La hematoxilina y eosina (HE) son colorantes de uso frecuente en histología ya que muestran adecuadamente las características estructurales generales de los tejidos, aunque se pierde la composición química de los elementos celulares porque muchos componentes se pierden durante la preparación de la muestra.

El primer paso en el proceso de tinción de HE es la preparación de una muestra de tejido de un órgano, para conservar su estructura. Para ello, se hace la fijación que se obtiene mediante una sustancia o mezcla de sustancias químicas que conservan permanentemente la estructura del tejido para los siguientes pasos de la tinción. Otras funciones de la fijación son:

- Abolir el metabolismo celular.
- Impedir la degradación enzimática de las células y tejidos por la autólisis (autodigestión).
- Destruir microorganismos patógenos tales como bacterias, hongos, virus y endurecer el tejido como resultado de la formación de enlaces cruzados o de la desnaturalización de moléculas proteicas.

El fijador más comúnmente usado es la formalina, que es una solución acuosa con una concentración del 37% de formaldehído y diluida con otras sustancias químicas y amortiguadores. El formaldehído conserva la estructura general de la célula y de los componentes extracelulares al reaccionar con los grupos amino de las proteínas. Además, las proteínas mantienen su capacidad de reaccionar con anticuerpos específicos con el formaldehído.

En el segundo paso, la muestra se dispone para su inclusión en parafina con el fin de permitir su corte, que por lo general a la muestra fijada se le realiza un corte en el rango de 5 μm a 15 μm para obtener una muestra delgada que pueda ser atravesada por la luz del microscopio. Para ello, la muestra se lava y se deshidrata en una serie de soluciones alcohólicas de concentración

creciente hasta alcanzar alcohol al 100%. La muestra es aclarada con solventes como el xileno o tolueno que se puede mezclar tanto en alcohol como en parafinas. Cuando la parafina fundida se ha enfriado y endurecido, se reduce el tamaño de bloque del tejido para que pueda ser colocado en una máquina cortadora especial conocida como el micrótomo. Los cortes obtenidos se montan sobre portaobjetos de vidrio utilizando un medio de montaje.

En el último paso, se tiñe la muestra para su posterior análisis histológico. Para colorear o teñir los cortes histológicos, la parafina debe disolverse y extraerse con xileno o tolueno, y los tejidos deben rehidratarse mediante el uso de una serie de soluciones de alcohol de concentración decreciente. Después, el tejido se tiñe sobre el portaobjetos con hematoxilina en agua. Debido a que la eosina (colorante de contraste) es más soluble en alcohol que en agua, se vuelve a deshidratar la muestra a través de una serie de soluciones alcohólicas de concentración creciente y después se tiñe con eosina en alcohol [1].

Una vez que se obtiene la muestra del tejido en la laminilla, esta lista para ser analizada con el uso de algún microscopio.

1.1.2 Análisis histológicos

Para realizar el análisis de una muestra de tejido se usa comúnmente un microscopio óptico, que consta de múltiples lentes para amplificar una imagen y ver los detalles estructurales de los elementos de los tejidos. Existen distintos tipos de microscopios, entre los que destacan:

- Microscopía de contraste de fase.
- Microscopía de campo oscuro.
- Microscopía de fluorescencia.
- Microscopía de barrido confocal.
- Microscopía de luz ultravioleta.
- Microscopía electrónica.

La imagen óptica obtenida por el microscopio se puede transformar en una imagen digital mediante el acoplamiento de una cámara digital al microscopio, y así poder salvar la imagen para su posterior análisis o evidencia del mismo. Otra alternativa es la microscopía virtual que integra la microscopía óptica convencional con la tecnología digital.

A partir de la microscopía auxiliada con tecnología digital, los preparados histológicos se exploran utilizando sistemas de adquisición de imágenes ópticas con enfoque automático, para crear archivos digitales de dos dimensiones [1]. El análisis suele ser realizado por un experto en la materia ya que las imágenes varían de un tejido a otro, de la luz incidente, de un corte a otro, del aumento, de la resolución etc. Su análisis suele ser de vital importancia ya que estos se realizan por la sospecha de alguna anomalía.

Entre las aplicaciones más importantes del estudio de la histología, se encuentra la histología mamaria que consiste en el análisis de tejidos de la mama femenina para conocer la variación estructural y funcional de sus células. A partir de este análisis se pueden diagnosticar varios trastornos de la mama, como la enfermedad mamaria benigna o patologías más serias, como el cáncer de mama [4]. Una forma reciente de análisis de imágenes de histología mamaria consiste en aplicar inteligencia artificial con modelos de aprendizaje de máquina. Es decir, procesar imágenes histológicas con redes neuronales artificiales para segmentar o interpretar la información contenida en estas imágenes. Por el enfoque de esta tesis, la siguiente sección describirá estas redes.

1.2 Redes neuronales artificiales

Una red neural artificial (ANN, por sus siglas en inglés) es un modelo computacional inspirado en la estructura de las capas corticales del cerebro humano [5]. Estos modelos imitan los mecanismos de aprendizaje, constan de unidades que imitan el comportamiento de una neurona biológica y se conectan con otras unidades para realizar cálculos y devolver una salida. En la Figura 1.1 se muestra un esquema simplificado de las neuronas más comunes donde: el cuerpo central se le denomina soma y contiene el núcleo celular, su prolongación se le llama axón y a la ramificación terminal se le denomina dendritas. A la conexión entre una neurona y otra se le denomina sinapsis, las cuales son intercambios eléctricos entre una neurona y otra.

El modelo biológico de una neurona se comporta de la siguiente manera: recibe estímulos eléctricos de neuronas predecesoras por medio de las dendritas, para luego procesar dichos estímulos en el soma. Estos estímulos trabajan bajo un umbral que, al ser cruzado, se produce una señal positiva que será enviado a las siguientes neuronas sucesoras.

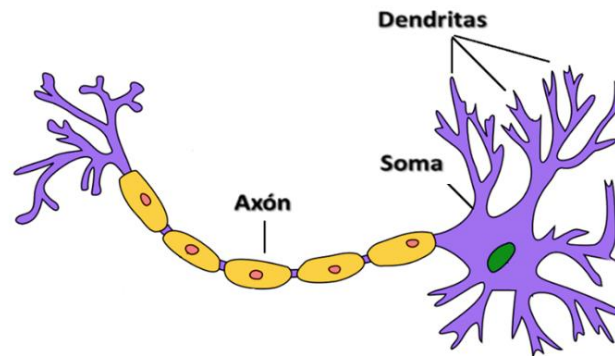


Figura 1.1 Esquema general de una neurona biológica.

El psiquiatra y neuroanatomista Warren McCulloch y el matemático Walter Pitt propusieron en 1943 el primer modelo matemático de una neurona artificial, creado con el fin de llevar a cabo tareas simples. Este modelo acepta entradas binarias y produce una salida binaria basada en un umbral que puede ajustarse, lo que resulta útil para problemas de clasificación. Al tomar varias entradas, la función toma una decisión basada en la suma de las entradas binarias. Si el valor de la suma excede el umbral, la salida es 1, de lo contrario es 0. Este modelo fue retomado por Rosenblatt en 1958, quien creó un modelo llamado perceptrón, que se muestra en la Figura 1.2 y la estructura fundamental de la neurona artificial moderna [6]. Siguiendo el esquema biológico, las entradas del perceptrón actúan como las dendritas o señales de entrada hacia la neurona. Los pesos actúan como la sinapsis ponderando que tanto se conectan las entradas con el soma. La función de activación realiza el papel del soma que decide en qué momento disparar esa señal eléctrica, y la salida modela al axón, que al activarse envía información hacia las neuronas con las cuales se encuentra conectada.

Este modelo funciona como un discriminante lineal siguiendo la siguiente ecuación:

$$g(x) = w^T v \quad (1.1)$$

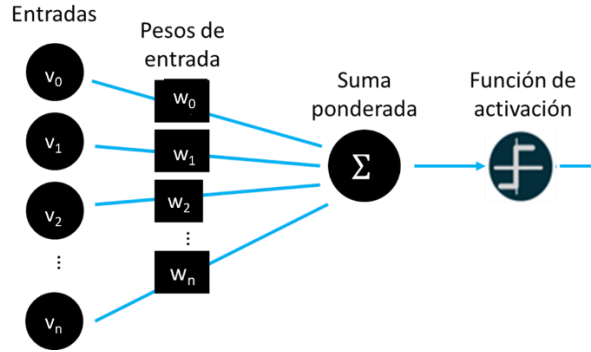


Figura 1.2 Esquema general del perceptrón.

Donde v es el matriz de entrada, w es la matriz pesos (ponderación de cada entrada), el cual tiene un *bias* que es un valor de sesgo que le permite cambiar la función de activación que mueve al hiperplano del origen. Para que la neurona genere una decisión, existe un hiperplano donde $w^{*T}v = 0$, haciendo dos clases linealmente separables [7], tal que:

$$w^{*T}v < 0 \quad \forall v \in \mathbb{R}1 \quad (1.2)$$

$$w^{*T}v > 0 \quad \forall v \in \mathbb{R}2$$

Estas neuronas artificiales se conectan entre sí para formar una red de comunicación que puede realizar cálculos altamente complejos y se le denomina red neuronal artificial [5]. Estas redes se pueden describir como un gráfico dirigido cuyos nodos corresponden a neuronas y los bordes corresponden a enlaces entre ellos. Cada neurona recibe como entrada una suma ponderada de las salidas de las neuronas conectadas a sus bordes. La Figura 1.3 muestra un ejemplo de una red que se construye a partir de la agregación de perceptrones de distintos tipos en estructuras complejas, esta imagen tiene un estructura sencilla que se conforma de tres capas, una capa de entrada, una capa intermedia u oculta y una capa de salida [5].

En 1969 Minsky y Papert publicaron en [8] las limitaciones del perceptrón, lo que generó dudas sobre el campo de las RNA. Sin embargo, en 1985 se publicó el algoritmo de propagación hacia atrás (*backpropagation*, por su traducción al inglés) de Geoffrey Hinton et al. y se vuelve a revitalizar el campo, resurgiendo una nueva era para el aprendizaje automático [6].

Cabe mencionar que el perceptrón no es el único modelo de red neuronal artificial, ya que existen diferentes redes que varían en cómo se organizan las neuronas, tipo y número de capas y como aprenden, Ejemplos de redes son: el perceptrón multicapa (MLP), redes neuronales directas, redes neuronales convolucionales, redes neuronales recurrentes, redes neuronales de base radial, máquinas de Boltzmann restringidas, etc. Además, existen variantes de cada una de ellas, creando una gama muy amplia para diferentes aplicaciones [9].

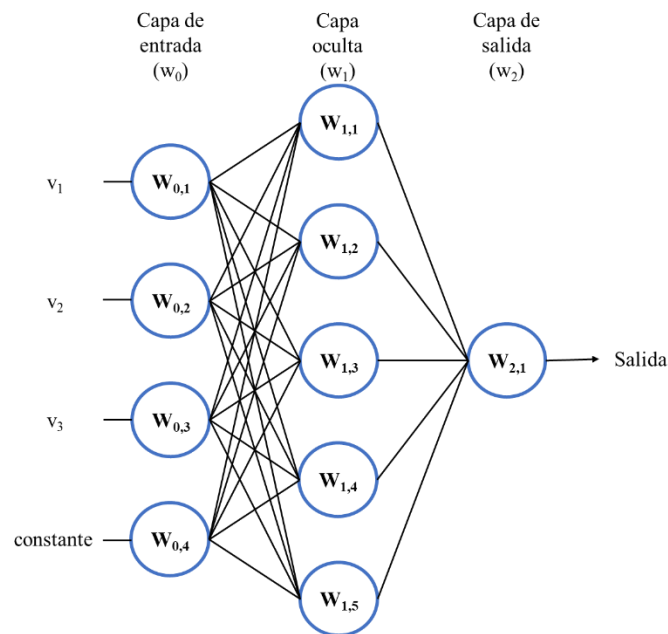


Figura 1.3 Ejemplo de red neuronal.

1.3 Aprendizaje automático

El aprendizaje automático, también conocido como aprendizaje máquina (ML, por sus siglas en inglés) es una forma de inteligencia artificial que permite que un sistema aprenda de los datos, en lugar de la programación explícita [10].

Con el objetivo de aplicar ML para encontrar la solución de una problemática, se deben considerar dos aspectos: la complejidad del problema y su necesidad de adaptabilidad [5]. La complejidad se refiere a que para resolver una problemática, se requiere analizar una cantidad de datos y variables que para un humano sería una tarea muy ardua o carente de sentido.

Adaptabilidad se refiere a cambiar las propiedades del algoritmo a la necesidad del problema. El ML representa una ventaja sobre la programación explícita, ya que el algoritmo se puede adaptar a la incertidumbre del manejo de los datos sin necesidad de cambiar su estructura [5].

1.3.1 Tipos de aprendizaje

Shai Shalev-Shwartz y Shai Ben-David presentan en [5], cuatro formas de clasificar los paradigmas de aprendizaje de un modelo de aprendizaje automático:

1. Supervisado contra no supervisado: se basa en la información que se utiliza para entrenar al clasificador durante el proceso de aprendizaje. En el caso del supervisado, el algoritmo aprende con los datos de entrada y su información de salida, de manera que la respuesta del clasificador se compara con la salida. El no supervisado se basa en analizar la composición estadística del conjunto de los datos de entrenamiento durante el proceso del aprendizaje. También existe el aprendizaje reforzado y se basa en que un agente define el estado del clasificador con base en refuerzos positivos o negativos que recibe al realizar una tarea.
2. Aprendizaje activo contra pasivo: El aprendizaje activo se basa en una interacción durante el entrenamiento entre el algoritmo y el entorno. Por ejemplo, haciendo preguntas o realizando experimentos. El pasivo se basa en que la máquina aprende solo observando la información proporcionada por el programador.
3. Ayuda de los datos de entrada: este se basa en el aprendizaje continuo, donde la máquina siempre aprende con los datos de entrada.
4. Protocolo de aprendizaje en línea contra aprendizaje por lotes: en este caso, el aprendizaje en línea se refiere a que la máquina aprende de forma iterativa cuando recibe datos y aprendizaje por lotes se basa en que la máquina aprende con una gran cantidad de datos previos.

1.4 Aprendizaje profundo

Aprendizaje profundo (DL, por sus siglas en inglés) tiene varias definiciones estrechamente relacionadas entre sí. Por ello, en el libro *Deep Learning Methods and Applications* [11] de Deng e Yu, se recopila una serie de definiciones para este concepto, las cuales conllevan a lo

mismo: un área de aprendizaje automático que se refiere a un conjunto de algoritmos que aprenden en múltiples niveles de abstracción que definen patrones para dar sentido a los datos.

Existen varios modelos de redes profundas, como redes neuronales convolucionales (CNN, por sus siglas en inglés), redes neuronales recurrentes, máquinas de Boltzmann, redes bayesianas, entre otras. De estos modelos, las redes neuronales convolucionales son las más utilizadas en el procesamiento de imágenes debido a su capacidad para extraer características abstractas [12].

En esta tesis se hará énfasis en las CNN ya que son los modelos más exitosos en el procesamiento de imágenes histológicas. A continuación, se describen estas redes.

1.4.1 Redes neuronales convolucionales

Las redes neuronales convolucionales son redes profundas utilizadas para la clasificación de imágenes y detección de objetos. La motivación básica para la red neuronal convolucional deriva de la comprensión de Hubel y Wiesel sobre el funcionamiento de la corteza visual del gato, donde porciones específicas del campo visual parecían excitar neuronas particulares [13]. La primera arquitectura basada en esta inspiración biológica fue el neocognitrón (Figura 1.4), una red neuronal jerárquica de varias capas de reconocimiento de patrones visuales a través de aprendizaje [13]. Posteriormente, se propuso la arquitectura LeNet-5, que consiste en dos conjuntos de capas convolucionales y de promediado, una capa convolucional totalmente conectada seguida por dos capas totalmente conectadas y al final un clasificador softmax [14]. Una ilustración de la arquitectura de LeNet-5 se muestra en Figura 1.5 y en los primeros años fue utilizada por varios bancos para reconocer números escritos a mano en cheques [9].

1.4.2 Capas que componen las CNN

Como un modelo general, las CNN constan de dos etapas: extracción de características y clasificación, como se muestra en la Figura 1.6. La primera etapa está compuesta por las capas convolucionales, *ReLU* y agrupamiento (*pooling*), mientras que la etapa de clasificación consta de las capas totalmente conectadas y la función softmax. En la red neuronal convolucional, los parámetros se organizan en conjuntos de unidades estructurales tridimensionales, conocidas

como filtros o núcleos. Suelen ser cuadrados de tamaño f , por ejemplo $f=3$ que representa un tamaño de 3×3 [9]. A continuación, se presentan detalles de cada una de las etapas de la red.

La entrada en general se puede representar matemáticamente como un tensor con las siguientes dimensiones:

$$ent = (n_H, n_{Wi}, n_C) \quad (1.3)$$

$$k = (f, f, n_C) \quad (1.4)$$

donde n_H es el tamaño en altura, n_{Wi} es el tamaño en ancho, n_C es el numero de canales y k es el tamaño del kernel.

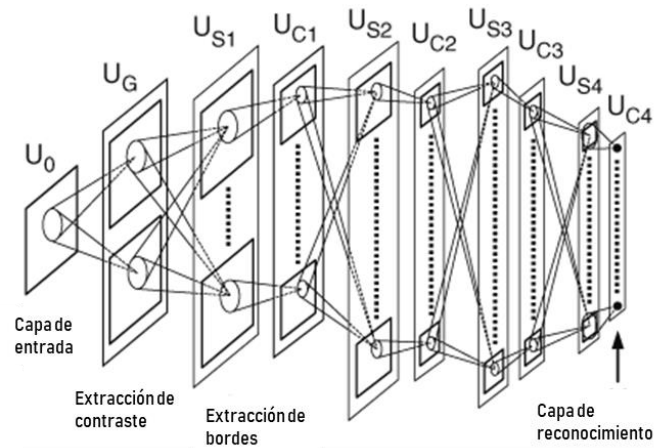


Figura 1.4 Arquitectura de un neocognitrón

Las capas de convolución extraen características abstractas mediante la convolución de la entrada a la red o salida de la capa anterior, mediante lo siguiente:

$$conv(ent, k)_{x,y} = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} \sum_{f=1}^{n_C} k_{i,j,k} ent_{x+i-1,y+j-1,f} \quad (1.5)$$

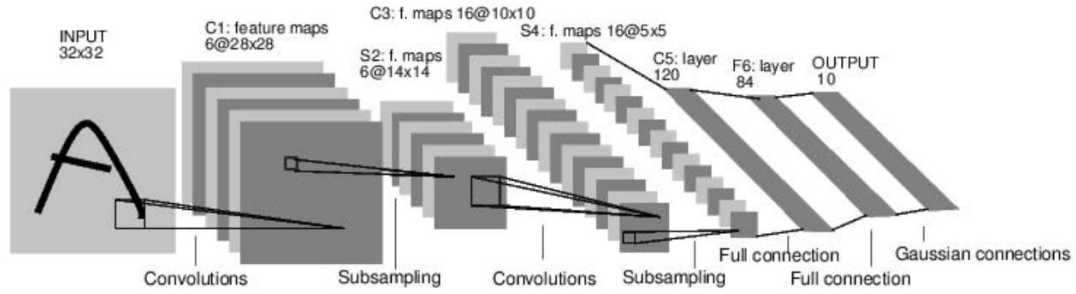


Figura 1.5 LeNet-5: una de las primeras redes neuronales convolucionales [15].

donde *conv* es el resultado de la convolución [9]. En la primera capa convolucional, la entrada suele ser una imagen en el formato de color rojo, verde y azul (RGB, por sus siglas en inglés) que equivale a una matriz de tres dimensiones donde $n_c = 3$ y la convolución se realizará tantas veces según la profundidad de la matriz. Para el resto de las capas, *ent* es un mapa de características resultado de capas anteriores el cual será entrada para las siguientes convoluciones.

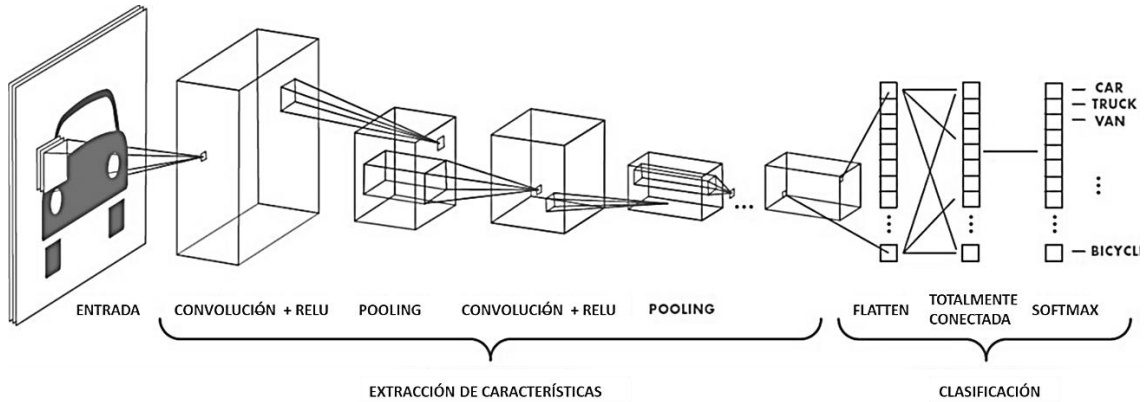


Figura 1.6 Red neuronal convolucional simple.

La salida de cada capa convolucional es un mapa de características con un tamaño dado por:

$$\dim(\text{conv}(\text{ent}, k)) = \left(\left\lceil \frac{n_H + 2p - k}{s} + 1 \right\rceil, \left\lceil \frac{n_W + 2p - k}{s} + 1 \right\rceil \right) \quad (1.6)$$

donde p representa el relleno añadido y s el tamaño del salto.

Después de las capas convolucionales, se tienen capas que son funciones de activación para mapa de características. La función activación más popular en las CNN es la Unidad Lineal

Rectificada(ReLU, por sus siglas en inglés) y actúa como un rectificador que genera una salida con las mismas dimensiones que la entrada, pero cambia los valores negativos a cero [9], como se observa en:

$$ReLU(valor) = \max(0, valor) = \begin{cases} 0, & \text{para } valor < 0 \\ valor, & \text{para } valor > 0 \end{cases} \quad (1.7)$$

donde valor corresponde cada elemento del mapa de características.

Posterior a la ReLU, se tienen las capas agrupación o *pooling*, que se encarga de reducir las dimensiones del mapa de características. El enfoque conocido como *max-pooling* extrae el valor máximo de pequeñas regiones (ventanas) de la salida de la ReLU. El resultando es un nuevo mapa de características con una dimensión más pequeña, pero con la misma profundidad. Esta capa está dada por:

$$max - pooling = \max_{x,y=1}^{h,w} y_{x,y,k} \quad (1.8)$$

donde h y w representan el alto y el ancho de la ventana a evaluar.

También se existe un enfoque conocido como *average-pooling*, que trabaja de una manera similar a *max-pooling*, pero regresa el promedio de los valores de la ventana [9]. El *average-pooling* está dado por:

$$average - pooling = \frac{1}{hw} \sum_{x,y=1}^{h,w} y_{x,y,k} \quad (1.9)$$

El tamaño resultante después de realizar el *pooling* se define por la siguiente ecuación:

$$\dim(pooling(imagen)) = \left(\left\lfloor \frac{n_H + 2p - f}{s} + 1 \right\rfloor, \left\lfloor \frac{n_W + 2p - f}{s} + 1 \right\rfloor, n_C \right) \quad (1.10)$$

Dentro de la etapa de extracción de características, es posible tener capas repetidas de convolución variando el tamaño de sus *kernels*, así como repetir sus capas de activación y *pooling*. Esto con el fin de extraer el mayor número de características útiles para la clasificación.

La salida de la última capa de extracción de características se transforma en un vector, como se muestra en la Figura 1.7. Esta salida sirve ahora como entrada a la primera capa de clasificación, que es la totalmente conectada.

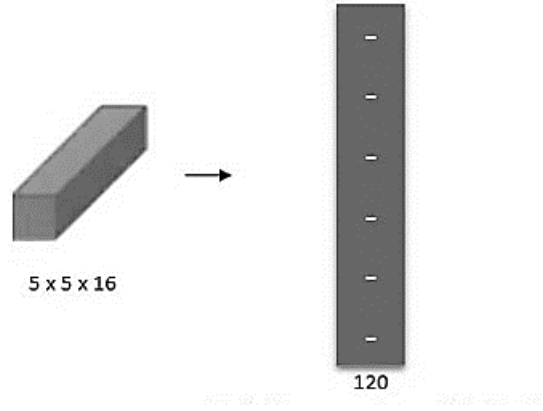


Figura 1.7 Aplanamiento (*flatten*), transformación de la matriz de características en un vector

La etapa de clasificación de la CNN cuenta con al menos una capa totalmente conectada, que es una red neuronal multicapa *feed-forward* que efectúa la clasificación con base en las características extraídas en capas anteriores. En esta capa todos los nodos están conectados con la capa precedente. [9]. Esta capa está dada por:

$$t_j = (w \cdot ent_{n_H n_{W_i} n_C}) + b \quad (1.11)$$

donde t_j corresponde a un vector de salida de la capa totalmente conectada y b al bias.

Por último, después de la capa totalmente conectada se tiene una función de activación softmax (S) que genera una distribución de probabilidad a partir de la salida de la capa totalmente conectada. Esta función consiste en probabilidades proporcionales a los exponenciales de los números de entrada dados por:

$$S(t)_j = \frac{e^{y_j}}{\sum_{k=1}^N (e^{y_k})} \quad (1.12)$$

donde y es la salida de las capas ocultas, N es el número de clases de nuestra red y S las probabilidades de salida de cada clase.

En general, esta es la estructura de una CNN, pero de acuerdo con la literatura se reportan variantes con distinto número de capas, cantidad, acomodo y flujo de datos, etc, que no se reportan aquí por no ser el objeto de estudio.

1.4.3 Entrenamiento

El proceso de entrenamiento es el *backpropagation* y consiste en encontrar los valores adecuados de los coeficientes de los kernels y las neuronas. Este proceso inicia con propagación hacia adelante, donde se inicializan aleatoriamente los pesos, los sesgos y los filtros. Después, se continúa con el proceso de propagación hacia atrás, donde el modelo intentará actualizar los parámetros para que la predicción general sea más precisa. El algoritmo de *backpropagation* utiliza la regla de la cadena de cálculo, que encuentra el gradiente del error en función de la suma de los productos del gradiente local en varias rutas desde el nodo hasta la salida. El algoritmo de *backpropagation* consta de dos fases principales, que son:

Fase de propagación: en esta etapa, la entrada alimenta a la red neuronal. La red realizará una serie de cálculos utilizando el conjunto actual de pesos. La salida final esperada se puede comparar con la salida del ejemplo de entrenamiento y se calcula la diferencia de la función de pérdida con respecto a la salida deseada[9].

Fase de propagación hacia atrás: a partir de la diferencia calculada en la fase de propagación, se realiza la regla de la cadena de cálculo para aprender el gradiente de la función de pérdida con respecto a diferentes pesos. Estos gradientes se utilizan para actualizar los pesos. Dado que estos gradientes se aprenden en la dirección inversa desde el nodo de salida, este proceso de aprendizaje se denomina fase inversa [9].

1.5 Estado del arte

Existen numerosos artículos publicados en el estado del arte que resuelven la problemática de la clasificación de imágenes de histología utilizando IA con métodos de aprendizaje automático y profundo. De esta manera, existe un gran acervo de distintos métodos de

procesamiento de imágenes en histología para detectar cáncer, parásitos en la sangre y en los intestinos.

En los trabajos reportados más antiguos, era común el uso de máquinas de vectores de soporte (SVM, por sus siglas en inglés), que funciona relativamente bien cuando hay un claro margen de separación entre clases, pero era necesario que un médico experto auxiliara en el desarrollo del esquema de extracción de características. En años recientes, se ha utilizado aprendizaje profundo con CNNs para procesamiento de imágenes de histología, ya que han mostrado mejores resultados debido a que las CNNs encuentran características morfológicas primarias como colores o bordes, para luego seguir con características más complejas como formas y objetos que identifican por medio de patrones. Sin embargo, tienen la limitante de necesitar un gran conjunto de datos para el entrenamiento de esta red.

A continuación, se presentarán algunos trabajos relacionados con SVM y aprendizaje profundo aplicados al análisis de imágenes en histología.

1.5.1 Aprendizaje automático

Uno de los modelos más citados para la clasificación de imágenes de histología son las funciones discriminantes lineales [16], que es un algoritmo de aprendizaje supervisado empleado para clasificación binaria o regresión. Por ejemplo, Ai Ping Qu et al. proponen en [17] un análisis de imágenes para detección de cáncer de mama mediante SVM, detectando nidos de tumores de carcinoma ductal invasivo de mama. El resultado mostró que el método propuesto logró un 87.1% de precisión en clasificación y un 80.2% de precisión en la segmentación de estroma tumoral basada en píxeles con su base de datos.

Siew Chin Neoh et al. proponen en [2] un sistema inteligente de soporte de decisiones para el diagnóstico de leucemia linfoblástica aguda a partir de imágenes de sangre microscópicas. Comprende áreas recortadas de células normales y blásticas obtenidas de muestras de sangre periférica de pacientes con leucemia y de individuos sanos. Aquí, la segmentación de cuencas hidrográficas controlada por marcadores se aplica primero para identificar y segmentar los glóbulos blancos basados en imágenes microscópicas. Luego, el algoritmo de agrupamiento basado en medidas discriminantes se utiliza para realizar la

segmentación del núcleo y el citoplasma. Posteriormente, se extraen un total de 80 características que representan la forma, la textura, el color y la información estadística del núcleo y las subimágenes del citoplasma. Se utilizan varios clasificadores que incluyen el MLP, SVM y los conjuntos con diversos métodos de combinación de ponderación para reconocer las células sanas y blásticas.

1.5.2 Aprendizaje profundo

Existen numerosos modelos de CNN basados en aprendizaje profundo para la clasificación de imágenes de histología, uno de ellos es la Red Neuronal Residual (ResNet) que proponen Kaiming He et al. en [18]. Por ejemplo, Gandomkar et al. propone en [19] una algoritmo para clasificar los portaobjetos digitales de mama teñidos con hematoxilina-eosina como benignos o cancerosos. Este método propuesto se divide en dos etapas. En la primera etapa, se entrenó una red ResNet para clasificar los parches de las imágenes como benignos o malignos. En la siguiente etapa, las imágenes clasificadas como malignas se subdividieron en cuatro subcategorías de cáncer. Finalmente, el diagnóstico para cada paciente se realizó combinando los resultados de las imágenes procesadas de ResNet's en diferentes factores de aumento utilizando un árbol de metadecisiones. Se logró un 96.25% de exactitud en clasificaciones correctas en 8 clases distintas.

Yun Jian et al. proponen en [20] clasificar imágenes histopatológicas de mama para la detección de cáncer mediante una red que incluye capas convolucionales, el módulo SE-ResNet, que se basa en la combinación de un bloque de exprimir y excitar (SE, por sus siglas en ingles) y el bloque de ResNet. Este modelo de red fue nombrado *Breast Cancer Histopathology Image Classification Network* (BHCNet). En este trabajo se utilizó el conjunto de datos BreakHis [21] y se logró una exactitud del 90.66% y del 93.74% para multclasificación y detección binaria respectivamente.

Otro modelo de red profunda para la clasificación de imágenes es la arquitectura convolucional profunda Inception [22]. Esta red se introdujo con el nombre de GoogleNet, que más tarde se modificó de varias maneras. Primero con la introducción de la normalización por lotes (Inception-V2), luego con factorizaciones en la tercera iteración de la red a las que se hará

referencia como Inception-v3. La red está formada por bloques de construcción simétricos y asimétricos que incluyen convoluciones, reducción de promedio, reducción máxima, concatenaciones y capas completamente conectadas. La normalización por lotes se usa con frecuencia en toda la red y se aplica a las entradas de activación. GoogleNet ha sido utilizada en varios artículos para inspección y clasificación de imágenes histopatológicas.

Utilizando Inception, Justin Ker et al. proponen en [23] automatizar el proceso de clasificación de imágenes histológicas de cerebro, que fue previamente entrenado en un subconjunto de imágenes de la base de datos de Imagenet, cumpliendo cuatro tareas:

1. Clasificar las imágenes de histología en cerebro normal o cerebro con alto grado de glioma (HGG).
2. Clasificar las imágenes de histología en cerebro normal, bajo grado de glioma o HGG.
3. Clasificar los cortes histológicos de mama en normal o carcinoma-in-situ (CIS) y repitiendo esto aplicando transferencia de aprendizaje para mejorar el desempeño de la red.
4. Optimizando la clasificación binaria de CIS y cerebro con HGG.

Justin Ker et al. modificaron en [23] la GoogleNet en sus últimas cuatro capas remplazándolas con capa de agrupamiento promedio global, 4 capas totalmente conectadas y 2 o 3 neuronas softmax. Los resultados de sus experimentos de los experimentos de clasificación coinciden o superan los de publicaciones similares.

Alom et al. proponen en [24] una red para mejorar la detección de cáncer de mama que combina la red Inception-v4, ResNet, y red neuronal convolucional recurrente. Esta red se llama Red Neuronal Recurrente Residual Convolucional Inception. De acuerdo con los experimentos, esta red presentó una precisión de 99.05% para la clasificación binaria y 98.59% clasificación multiclase de imágenes cáncer de mama en el conjunto de datos de 2015 Breast Cancer Classification Challenge.

Rui Yan et al. proponen en [25] integrar las ventajas de las redes neurales convolucionales, redes recurrentes y las correlaciones espaciales para la clasificación de imágenes histopatológicas de cáncer de mama. En este trabajo se utiliza la base de datos BACH que contiene 3771 imágenes patológicas de alta resolución de cáncer de mama tenidas con hematoxilina y eosina. El método utilizado logra una exactitud del 91.3% en la categorización de 4 clases.

Abdullah Al Nahid et al. recopilan en [12] distintos métodos de clasificación de imágenes de cáncer de mama. Mencionan que en el estado del arte las técnicas de aprendizaje profundo en especial las CNN han tenido mejor capacidad de extracción de características en la clasificación avanzada de imágenes histológicas, con el único inconveniente de requerir un gran número de muestras para su aprendizaje.

CAPÍTULO 2

ANÁLISIS DE LOS MODELOS EN EL ESTADO DEL ARTE DE LOS MÉTODOS DE PROCESAMIENTO DE IMÁGENES DE HISTOLOGÍA

Para proponer la CNN de procesamiento de imágenes de histología de esta tesis, fue necesario realizar un estado del arte acerca de los métodos de ML y DL dedicados a trabajar con imágenes de histología. Por ello, en este capítulo se presentan las estadísticas de este análisis del estado del arte. El análisis se hizo con base en la información que presentan los artículos, que incluye la red propuesta o utilizada, la base de datos y su origen, aumentación de datos, procesamiento previo, métricas utilizadas y su desempeño.

De esta manera, este capítulo se organiza de la siguiente forma: la sección 2.1 presenta las redes más utilizadas en el estado del arte, mientras que la sección 2.2 las bases de datos. La sección 2.3 describe la forma de implementación de algoritmos. La sección 2.4 muestra las métricas de evaluación y finalmente, la sección 2.5 presenta la conclusión de este análisis.

2.1 Modelos utilizados en el estado del arte para clasificación de imágenes histológicas

En el análisis del estado del arte realizado en esta tesis, se reportaron diversas redes para extracción de características y clasificación. De los más de 40 artículos analizados, se presentaron los siguientes modelos:

- SVM [17], [21], [26]–[29].
- k vecinos más cercanos (KNN, por sus siglas en inglés) [26].
- Algoritmos bayesianos (*NaiveBayes*) [26].
- Árbol de decisión [26], [30].
- Long-Short-Term-Memory (LSTM, por sus siglas en inglés) con redes neuronales recurrentes (RNN, por sus siglas en inglés) [31]–[33].
- Red neuronal convolucional profunda (DCNN, por sus siglas en inglés) [24], [34]–[36].
- Redes neuronales probabilísticas [37].
- Canal-inception-residual Unet (RIC-Unet) [38].
- Red de creencia profunda (DBN, por sus siglas en inglés) [39].

- Análisis de núcleos adaptativos multiescala (MANA, por sus siglas en inglés) [40].
- CNN tradicionales [31], [33], [41]–[51].
- Otras CNN con variantes para mejorar su clasificación, tales como AlexNet, Xception, Inception-v3, Inception-v4, ResNet e InceptionResNet-v2 [19], [20], [23]–[25], [29] [32], [34], [52]–[55].

Cabe mencionar que en algunos artículos se utiliza más de un modelo.

La Figura 2.1 muestra el porcentaje de trabajos analizados en la literatura que utilizan cada uno de las redes mencionadas. De acuerdo con esta gráfica, las CNN tradicionales son las redes más utilizadas. Estas redes se proponen con diferentes hiperparámetros tales como número de capas, tamaño, cantidad de filtros, parámetros de aprendizaje, etc. Después de estas CNNs, se encuentran ResNet, SVM y otras redes profundas.

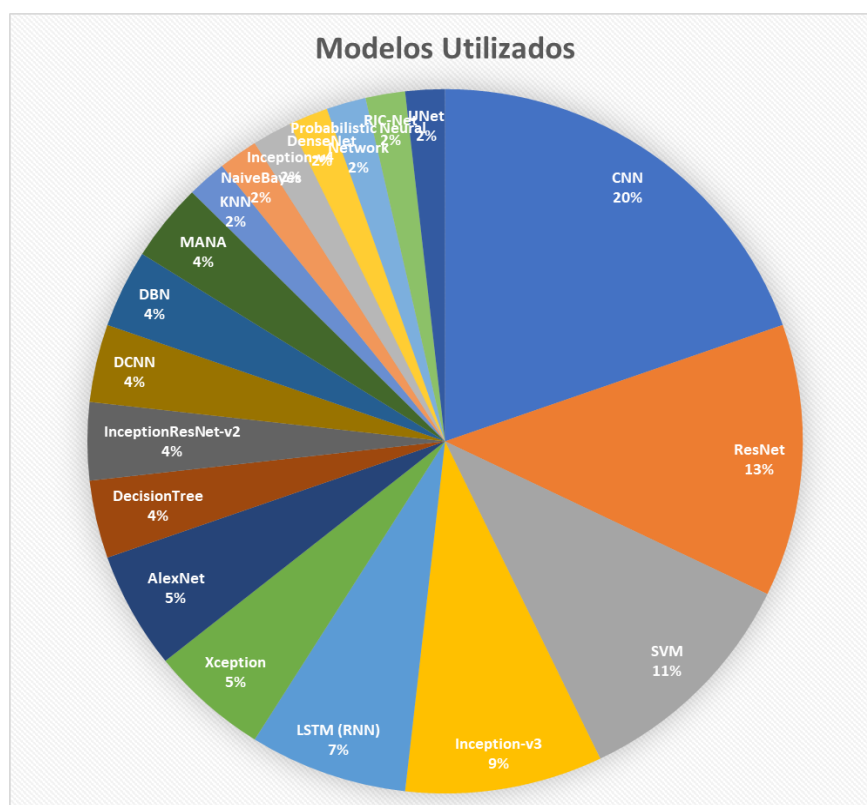


Figura 2.1 Modelos dedicados a trabajar con imágenes de histología en artículos del periodo 2015-2019.

2.2 Bases de datos

Las bases de datos utilizadas en los artículos de procesamiento de histología son privadas en su mayoría. Por ejemplo, en [23] reportan Brain histology, una base de datos de imágenes histológicas de tejido cerebral obtenida en Departamento de Patología del Hospital Tan Tock Seng, Singapur. El objetivo de esta base de datos fue entrenar una CNN que ayude a los patólogos en la clasificación e inspección de las diapositivas de histología para agilizar la atención médica. Como esta, existen múltiples bases de datos que se utilizan para analizar problemáticas muy específicas con imágenes de histología, pero cuentan con acceso restringido por lo cual no pueden ser analizadas.

Por otro lado, existen pocas bases de datos públicas, que son muy populares en la literatura y se pueden obtener bajo solicitud. La más utilizada de estas bases de datos es BreakHis publicada en 2016 por Fabio A. Spanhol et al. en [21], y puede ser adquirida llenando un formulario en la página <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>. Esta base de datos se compone de 9,109 imágenes microscópicas de tejido tumoral de seno teñidas con HE, recolectadas de 82 pacientes a través de estudios clínicos desde enero de 2014 hasta diciembre de 2014 en el Laboratorio P&D en Paraná, Brasil usando diferentes factores de aumento (40X, 100X, 200X y 400X). El procedimiento de preparación utilizado para obtener las imágenes es un proceso de parafina estándar para preservar la estructura original del tejido y la composición molecular observada bajo un microscopio óptico. El proceso de preparación completo incluye pasos como fijación, deshidratación, limpieza, infiltración, incrustación y recorte. Para montar la muestra de tejido en un portaobjetos de vidrio, se realizan cortes en secciones de 3 μm con un micrótomo. Una vez que se obtiene la muestra, el patólogo identifica el área del tumor en cada portaobjetos mediante el análisis visual de la sección de tejido bajo un microscopio. El diagnóstico final en cada caso lo realiza un patólogo experimentado y se confirma mediante pruebas complementarias como el análisis inmunohistoquímico. A partir de esto, se logró obtener un conjunto de muestras que contiene 2,480 muestras benignas divididas en cuatro tipos de tumores:

- Adenosis

- Fibroadenoma,
- Tumor phyllodes
- Adenoma tubular

También contiene 5,429 muestras malignas divididas en cuatro tipos de tumores que representan cáncer de mama:

- Carcinoma ductal,
- Carcinoma lobulillar,
- Carcinoma mucinoso
- Carcinoma papilar.

Las imágenes tienen las siguientes características: tamaño de 700x460 píxeles, composición de color RGB, profundidad de 8 bits en cada canal, formato PNG y se distribuyen como se muestra en Tabla 2.1.

Tabla 2.1 Distribución por factor de magnificación y clase.

Factor de aumento	Benigno	Maligno	Total
40x	625	1370	1995
100x	644	1437	2081
200x	623	1390	2013
400x	588	1232	1820
Total	2480	5429	7909
Número de pacientes	24	58	82

El propósito de esta base de datos es proporcionar imágenes médicas necesarias para llevar a cabo experimentos fiables para demostrar la eficacia de los métodos propuestos. De esta manera se puede estandarizar las pruebas para desarrollar algoritmos y evitar que se realicen experimentos utilizando datos seleccionados por investigadores.

2.3 Implementación de algoritmos

Tres aspectos importantes en el desarrollo de algoritmos para clasificación de imágenes de histología son el lenguaje de programación, preprocesamiento de datos y métricas de evaluación. Lenguaje de programación se refiere a la serie de instrucciones que bajo una sintaxis le proporcionan a una computadora las órdenes para su actuar [56]. Preprocesamiento significa aplicar técnicas de preparación de datos, como el aumento de datos, la normalización de tinción y las estrategias de normalización de imágenes para optimizar el proceso de entrenamiento. Las métricas de evaluación se verán más adelante en este capítulo.

En referente al lenguaje de programación más utilizado para clasificación de imágenes de histología, 43% de los trabajos utilizaron Python, 32% utilizaron Matlab y 25% no mencionan el lenguaje, tal como se muestra en Figura 2.2. Cabe mencionar que el lenguaje más utilizado para los modelos de aprendizaje profundo es Python ya que este cuenta con librerías especializadas y es de código libre [57].



Figura 2.2 Lenguajes utilizaos en los algoritmos de clasificación de imágenes de histología.

En cuanto a preprocesamiento, pocos trabajos recurren a la aumentación de datos, tal como lo muestra la Figura 2.3, que consiste en un aumento de nuestro conjunto de datos mediante perturbaciones a las imágenes originales (esto se ve más a detalle en el capítulo 4). No obstante, los pocos los artículos que utilizan aumentación, reportan mejor desempeño a comparación de

aquellos que no utilizan estas técnicas basado en el mismo modelo y misma base de datos. Por ejemplo, en [58] se reporta la comparación de desempeño al utilizar aumentación de datos contra no utilizar aumentación de datos en cinco modelos de DCNN utilizados en el estado del arte para la detección de melanoma en imágenes dermatoscópicas de lesiones cutáneas. En precisión, reporta en las redes un incremento de 7% en AlexNet, 7.2% en ResNet50, 6.2% en VGG16, 3.8% en VGG19 y 45% en Xception, en comparación de las mismas redes sin utilizar técnicas de aumentación de datos.

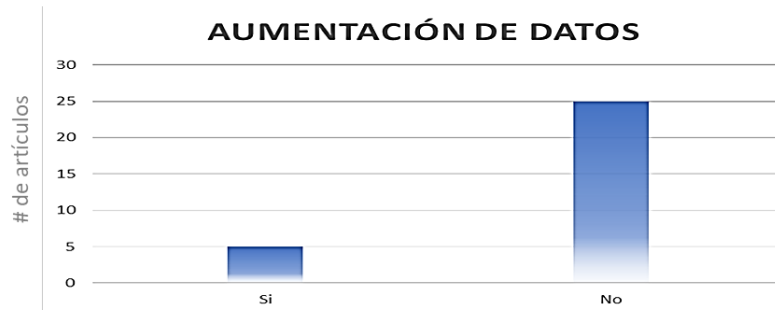


Figura 2.3 Métodos que realizan una aumentación de datos antes del entrenamiento.

2.4 Métricas de evaluación

El entrenamiento de las redes neuronales para la clasificación de imágenes histológicas suele ser supervisado. Por ello, los resultados para la predicción de tejido mamario se miden con las evaluaciones de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN). En el caso de estudio, los TP representan las imágenes con tumores benignos bien clasificados, FP representan las imágenes con tumores benignos mal clasificados, TN representan las imágenes con tumores malignos bien clasificados y FN representan las imágenes con tumores malignos mal clasificados [59]. Las métricas más comunes para medir el desempeño son exactitud, precisión, exhaustividad y puntuación-F1 que se calculan de la siguiente manera:

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2.1)$$

$$Precisión = \frac{TP}{TP + FP} \times 100 \quad (2.2)$$

$$Exhaustividad = \frac{TP}{TP + FN} \times 100 \quad (2.3)$$

$$Puntuación - F1 = 2 \times \frac{Recall \times Precisión}{Recall + Precisión} \quad (2.4)$$

Donde la exactitud [59] (*accuracy*, por su traducción al inglés) dada por la ecuación 2.2, es la métrica más utilizada para medir un modelo en ML ya que muestra los casos que se clasificaron correctamente. Sin embargo, la exactitud no distingue entre los tumores malignos que se clasificaron como benignos y los tumores benignos que se clasificaron como malignos. Por esta razón, se utilizan las métricas de precisión (2.2), exhaustividad (2.3) y puntuación F1 (2.4). La precisión [59] (*precision*, por su traducción al inglés) es la métrica que nos indica cuales tumores que se clasificaron como benignos son realmente benignos. La exhaustividad [59] (*recall*, por su traducción al inglés) es la fracción de tumores benignos que identificó el sistema. Finalmente, la puntuación F1 [5] (*F1-score*, por su traducción al inglés) es la media armónica de la precisión y la exhaustividad.

Para evaluar que estas métricas midan correctamente la capacidad de aprendizaje de los modelos, se utiliza la validación. Este método (también llamado remuestreo) es una estrategia que puede estimar el poder predictivo del modelo cuando solo se aplican datos con los que el modelo no ha sido entrenado (en este caso imágenes). El ejemplo más sencillo consiste en repartir aleatoriamente los datos disponibles en dos grupos, uno se emplea para entrenar y otro para evaluar. Aunque esta es la opción más simple, tiene dos problemas principales:

1. La estimación del error es muy variable, depende directamente con cuales imágenes se utilizan como conjunto de entrenamiento y cuales se utilizan como conjunto de validación (problema de varianza).
2. Al excluir algunas de las imágenes disponibles como datos de entrenamiento, hay menos información que se puede utilizar para entrenar la red, lo que reduce su capacidad de aprendizaje.

Una forma para evitar estos problemas de la validación y medir la capacidad de aprendizaje del sistema es la validación cruzada, que es un método para evaluar la capacidad de

generalización de modelos predictivos y prevenir el sobreajuste. Una de sus variantes es la k -carpetas (en este trabajo se sustituye la constante k conocida en la literatura por la h , ya que k se utiliza en este documento para definir los filtros), y consiste en dividir aleatoriamente la base de datos en h grupos de aproximadamente el mismo tamaño. Entre estos, $h-1$ grupos son usados para entrenar la red y un grupo usado para validación. En cada iteración, este proceso se repite h veces utilizando diferentes grupos como validación. Este proceso es evaluado h veces y se utiliza el promedio para la evaluación final de la red[60], por lo tanto el tiempo de entrenamiento incrementa h veces. Por lo general, se recomienda que h entre 5 y 10 (El valor más alto de h conduce a una red menos sesgada pero una gran varianza podría conducir a un sobreajuste), en la Figura 2.4 se muestra un ejemplo de una base de datos dividida en 5 carpetas y que rol juega cada carpeta durante el entrenamiento. La explicación del proceso de validación cruzada se puede ver en [60].

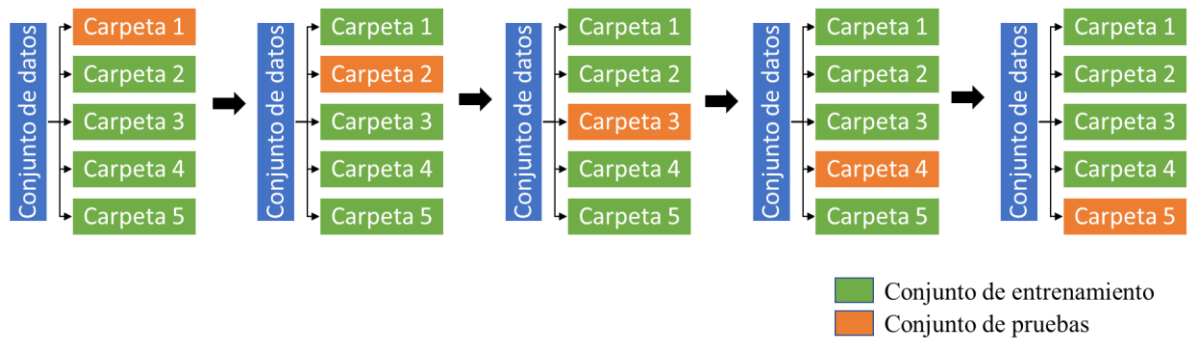


Figura 2.4 Validación cruzada de 5-carpetas.

2.5 Conclusiones del análisis realizado

Del análisis de artículos que trabajan con imágenes histológicas y que fueron publicados en la literatura científica en el periodo del 2015 al 2019, se encontró que los algoritmos de aprendizaje profundo fueron los más utilizados, alcanzando un 62% del total de artículos. Las CNN más utilizadas fueron las tradicionales con diseño único en cuanto al número y tipo de capas. No obstante, en este trabajo se van a descartar como base para el desarrollo de los métodos a proponer, ya que los trabajos que proponen estas redes carecen de la información necesaria para ser replicadas. Seguido de estas, las CNN especializadas como las redes ResNet, Inception-v3, Xception y InceptionResNet-v2m, quienes fueron analizadas ya que son

utilizadas en varios artículos que trabajan con imágenes de histología y en otras tareas de clasificación. Además, la arquitectura de estas redes especializadas fue publicada a detalle y su código se encontró público en Internet. Otro aspecto a considerar en estas redes es que reportan buen desempeño en diversas tareas de clasificación. Por estas razones, estas redes servirán como base para crear una CNN de clasificación propia.

Se seleccionó Python para trabajar en esta tesis, ya que este lenguaje es más propicio para los modelos de aprendizaje profundo por la portabilidad que les brinda. Esta portabilidad les permite correr en diferentes plataformas, sistemas embebidos y sistemas operativos. Además, Python es de código abierto y está disponible en Internet para su descarga, lo que lo hace más flexible a cambios.

CAPÍTULO 3

REDES NEURONALES CONVOLUCIONALES PARA EL ANÁLISIS DE IMÁGENES DE HISTOLOGÍA

En el capítulo anterior se pudo ver que las CNN especializadas han cobrado popularidad en el estudio de imágenes de histología de diversos tejidos. Por ello, en este capítulo se presenta un análisis que parte del trabajo de Bianco et al. quienes presentan en [61] una revisión de CNN especializadas que fueron propuestas en el estado del arte para el reconocimiento de imágenes. En este trabajo se comparan las redes usando múltiples índices de rendimiento, como la precisión del reconocimiento, la complejidad de la red (número de capas), la complejidad computacional (número de operaciones de punto flotante por segundo), el uso de memoria y el tiempo de inferencia. Este análisis se puede ilustrar en la Figura 3.1, donde se observa que las redes más utilizadas en histología son ResNet, Inception-v3, Xception e InceptionResNet-v2. Estas redes presentan un alto desempeño en cuanto a precisión y se encuentran en la media de número de operaciones de punto flotante por segundo (FLOPS, del inglés *floating point operations per second*).

Por ello, en este capítulo se presenta un análisis de las redes ResNet, Inception y Xception. Además, también se añade un análisis de la red de MobileNet-V2, ya que es una arquitectura para sistemas embebidos que realiza menos operaciones de punto flotante que las mencionadas de Bianco et al. Esta red fue incluida en este análisis ya que se busca una red de procesamiento de imágenes de histología con baja cantidad de operaciones que sea compatible con dispositivos embebidos y móviles.

El análisis presentado en este capítulo se basa en explicar una serie de diagramas que representan las redes, por lo que es necesario conocer la simbología presentada en la Figura 3.3, donde se muestran las capas que componen una CNN con diferentes colores. Como se observa en esta figura, las capas se representan de la siguiente manera:

- Bloques rojos se representan las capas convolucionales dada por la ecuación 1.5.

- Bloques grises simbolizan las capas de agrupación (*pooling*), dadas por la ecuación 1.8 para *max-pooling* (abreviada como max-pool dentro del bloque) y la ecuación 1.9 para *average-pooling* (abreviada como avg-pool dentro del bloque).
- Bloques morados representan operaciones de unión.
- Capas densas dadas por 1.11 se simbolizan en azul.
- Funciones de activación como la softmax dada por la ecuación 1.12 con un círculo blanco y la letra inicial de la función a la que corresponde.

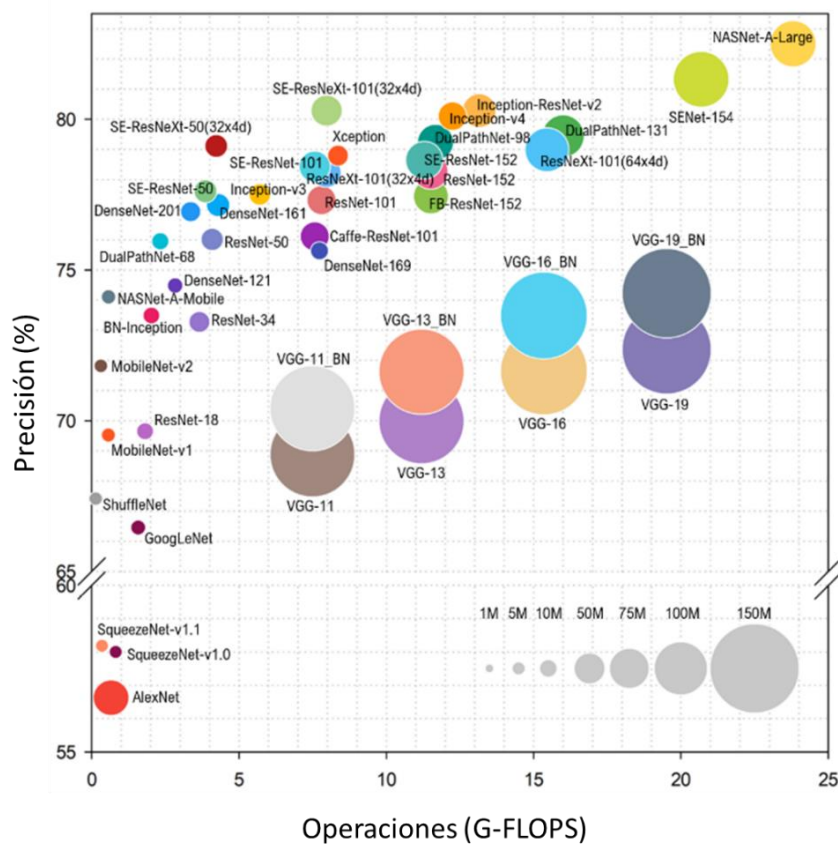


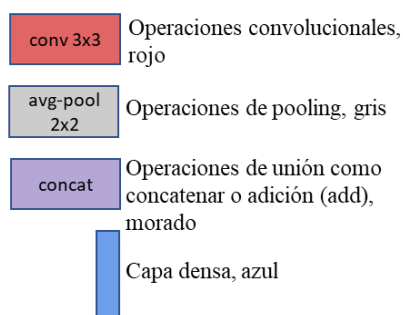
Figura 3.1 Gráfico de bolas que informa la precisión contra la complejidad computacional. El tamaño de cada círculo corresponde a la complejidad de la red [61].

También, en pequeños círculos amarillos se representan las funciones de activación dada por la ecuación 1.7, mientras que los círculos blancos simbolizan otras funciones como la ecuación 1.12.

Dentro de las arquitecturas, existen conjuntos de capas y operaciones que se repiten, y se reúnen dentro de un recuadro (amarillo, verde o naranja) en forma de módulos como se muestra en la Figura 3.3. Para indicar las veces que cada módulo se repite, este se encierra entre corchetes y se le asigna un multiplicador como se muestra en la Figura 3.4.

Para revisar estas redes, el resto de este capítulo se organiza de la siguiente manera. En la sección 3.1 se ve la red ResNet, la sección 3.2 Inception, la sección 3.3 Xception y en la sección 3.4 MobileNet.

Capas



Funciones de activación



Otras funciones

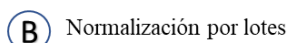


Figura 3.2 Capas y funciones en los modelos de red.

Módulos / Bloques

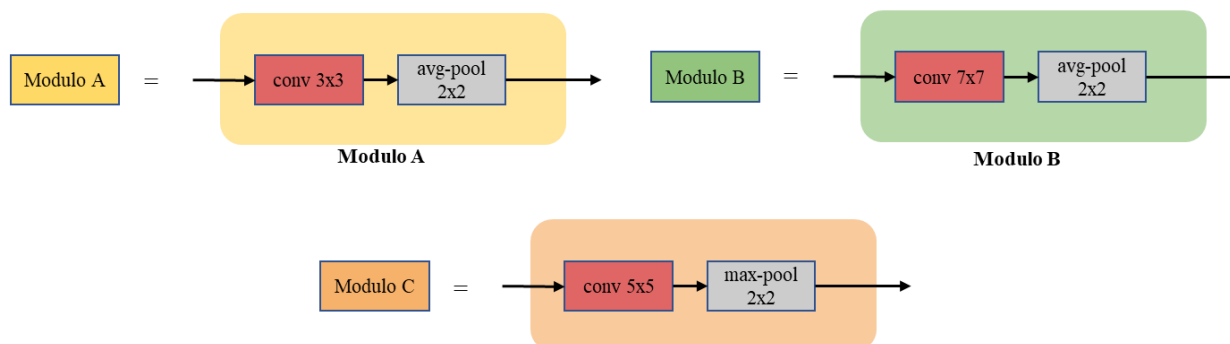


Figura 3.3 Módulos y bloques en los modelos de red [62].

Capas o Módulos / Bloques repetidos

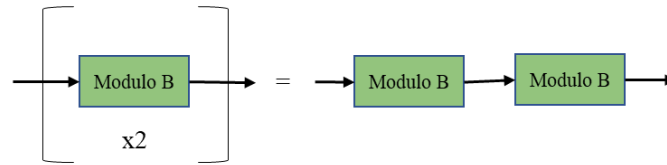


Figura 3.4 Módulos repetidos en la red [62].

3.1 ResNet

Una de las redes más populares en la literatura es la Red neuronal residual, que es una red neuronal convolucional con conexiones residuales [18]. Una de sus variantes propuestas fue la ResNet-50, que es una red de 50 capas, representada por la Figura 3.3, y consta de:

Una imagen de entrada representada matemáticamente como un tensor con dimensiones basadas en la ecuación 1.3. Una capa de convolución con un tamaño de $7 \times 7 \times 64$. A continuación, una capa de *max-pooling*. Las siguientes capas están son tres bloques compuestos por convoluciones, la primera convolución con un kernel $1 \times 1 \times 64$ canales, la siguiente de con un kernel $3 \times 3 \times 64$ y la última con un kernel $1 \times 1 \times 256$ canales. El primer bloque de “convolución” añade una convolución en paralelo a las tres anteriores cuyo kernel es de $1 \times 1 \times 256$ y la salida la añade a la salida de la tercera convolución y los otros dos bloques “identidad” añaden una conexión residual entre la entrada del bloque y la añade a la salida de la tercera convolución. A continuación, se tiene 4 bloques compuestos por una serie de convoluciones, la primera con un kernel de $1 \times 1 \times 128$, la segunda con un kernel de $3 \times 3 \times 128$ y la última con un kernel de $1 \times 1 \times 512$. Al igual que los bloques anteriores, el primero corresponde a un bloque que añade una convolución en paralelo a las tres anteriores cuyo kernel es de $1 \times 1 \times 512$ y los otros tres añaden conexión residual al igual que los anteriores. Después se añaden 6 bloques con el mismo patrón que los anteriores con kernel de tamaño $1 \times 1 \times 256$ en la primera convolución, $3 \times 3 \times 256$ para la segunda y $1 \times 1 \times 1024$ para la tercera. El primer bloque corresponde al “convolución” que añade la convolución en paralelo con kernel de tamaño $1 \times 1 \times 1024$ y cinco de “identidad” con conexión residual. A continuación, se tiene 4 bloques compuestos por una serie de convoluciones, la primera con un kernel de $1 \times 1 \times 512$, la segunda de $3 \times 3 \times 512$ y la tercera

de $1 \times 1 \times 2048$. El primer bloque corresponde al “convolución” que añade la convolución en paralelo con kernel de tamaño $1 \times 1 \times 2048$ y dos de “identidad” con conexión residual.

Una vez terminada la extracción de características por las capas convolucionales, se tiene una capa de reducción de características, una capa de agrupación promedio global y una capa de clasificación de una capa completamente conectada y al final una función softmax (ecuaciones 1.11 y (1.12) [63].

En la descripción de la red anterior, no se detallan las funciones de activación ni las que cuentan con normalización por lotes ya que es básicamente agregar un paso adicional entre la neurona y la función de activación para normalizar la activación de salida. Esto se ilustra en las Figura 3.6 y Figura 3.6.

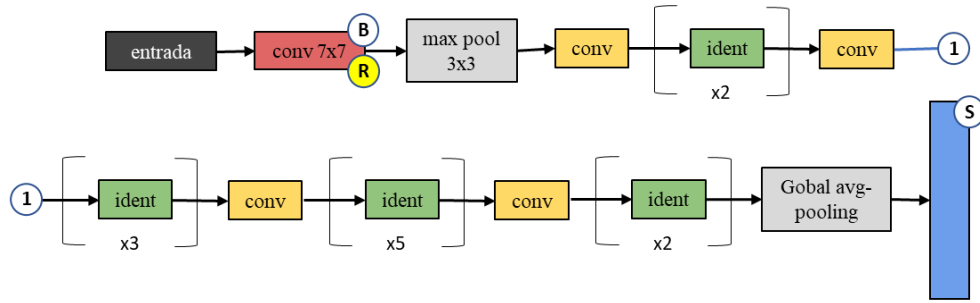


Figura 3.5 Bloques de la arquitectura de ResNet50 [62]

3.2 Redes Inception

Otra red popular debido a su bajo costo computacional es la red neuronal convolucional Inception [22] publicada en 2015. Esta utiliza convolucionales que se realizan de manera paralela con diferentes tamaños de kernel, por ejemplo de 1×1 , 3×3 y 5×5 , para después concatenar la salida de las convoluciones en un solo mapa de características. Para entender el beneficio de las redes Inception, hay que conocer las redes convolucionales definidas por la ecuación 1.5, como el ejemplo de la Figura 1.6, donde se utiliza la operación de *pooling* dada por la ecuación 1.8 o 1.9, para reducir el tamaño del mapa de características. No obstante, esto genera un costo computacional tan grande como la matriz en la que se aplique la convolución. Por ejemplo, si tenemos una matriz tamaño $d \times d$ con k filtros y se desea reducir a $\frac{d}{2} \times \frac{d}{2}$ con $2k$

filtros, se necesita la convolución con el paso de uno con $2k$ filtros y aplicar una capa *pooling* adicional, con lo que se necesitan $2d^2k^2$ operaciones.

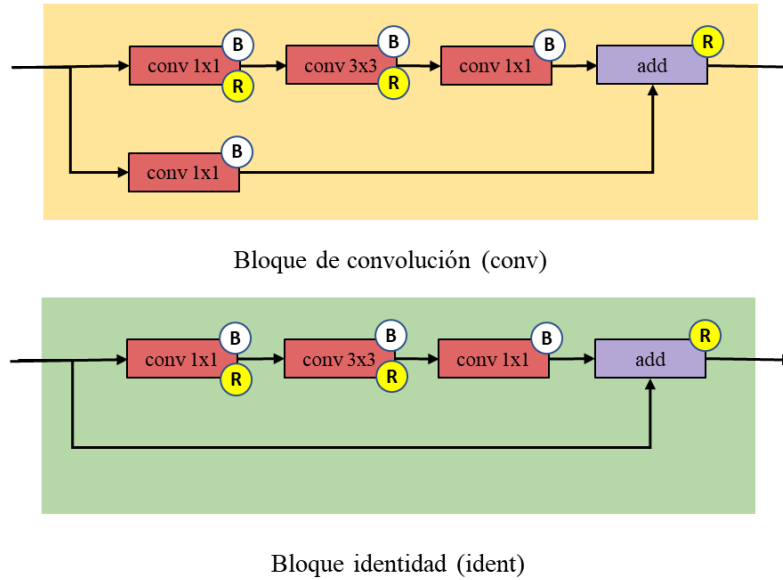


Figura 3.6 Bloques de la arquitectura de ResNet50 [62].

Para reducir el costo computacional, el módulo Inception plantea hacer dos bloques en paralelo con el mismo tamaño paso. El primero con una capa de *max-pooling* y el segundo con la capa de convolución en paralelo. Aunado a esto, se añaden convoluciones de 1×1 antes de las convoluciones 3×3 y 5×5 para reducir los costos computacionales tal como se muestra en el módulo Inception simple de la Figura 3.7.

La tercera versión de esta red (Inception-v3) fue lanzada en 2016 en [65] y hasta finales del 2020 ha sido citada por más de 12443 artículos entre los que se encuentran algunos dedicados a la clasificación de imágenes de tejidos de mama y cerebro como se muestra en [52]. La arquitectura la red Inception-v3 se muestra Figura 3.8.

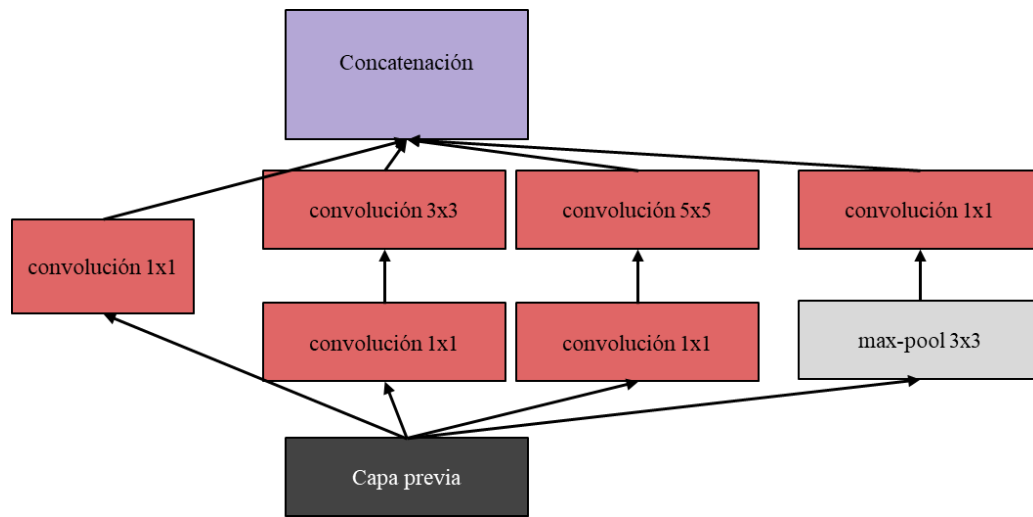


Figura 3.7 Módulo Inception simple.[64]

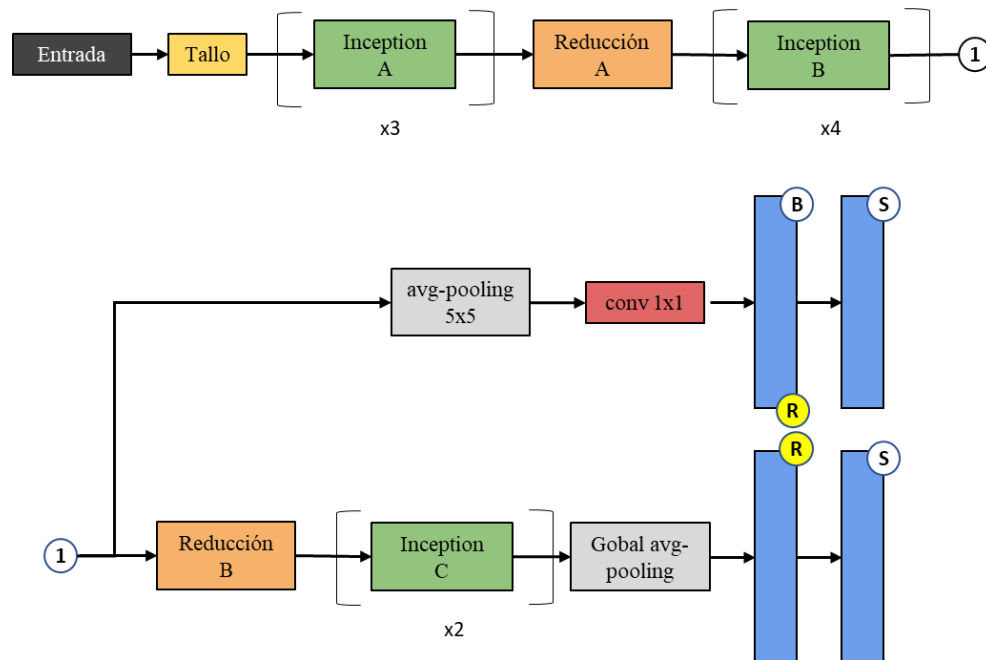


Figura 3.8 Arquitectura de la red Inception-v3 [62].

De esta manera, la red Inception-v3 de 42 capas consta de:

- Una imagen de entrada representada matemáticamente como un tensor con dimensiones basadas en la ecuación 1.3.

- El bloque del tallo, compuesto por una secuencia de 3 capas convolucionales de tamaño 3×3 seguido de una capa de agrupación tipo *max-pooling*, dos capas convolucionales de 1×1 y 3×3 respectivamente y finaliza con otra capa de agrupación *max-pooling*. Este bloque se ilustra en la Figura 3.9 .
- El bloque *Inception A*, una de las variantes del módulo Inception simple. En este bloque los autores factorizan la convolución de 5×5 en dos operaciones de convolución de 3×3 para aumentar la velocidad de cálculo. Este cambio se debe a que la convolución 5×5 tiene un costo computacional 2.78 mayor a la convolución 3×3 reduciendo más de 28% el número de operaciones. Este bloque se ilustra en la Figura 3.10.

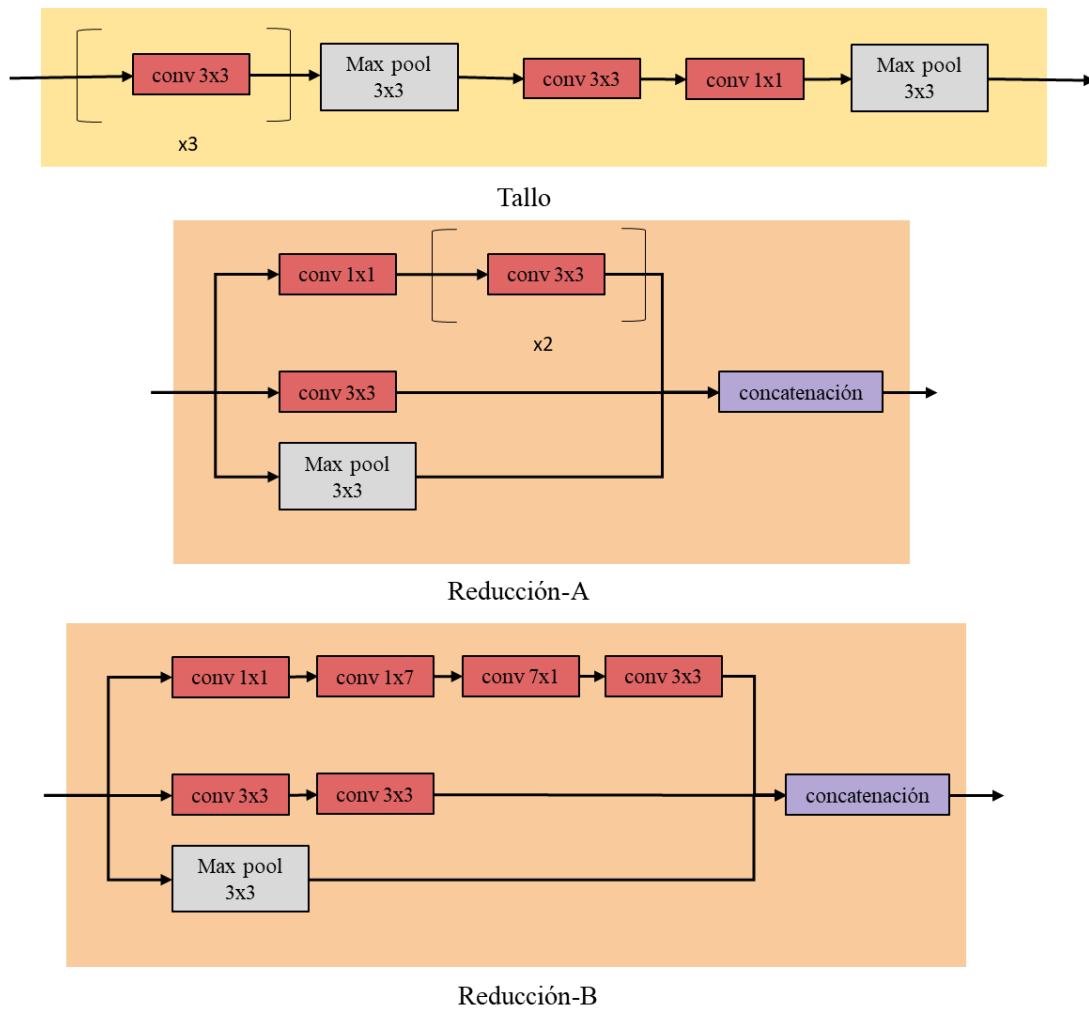


Figura 3.9 Módulos de reducción y tallo.

- El bloque *Inception-B* factoriza una convolución con un tamaño de filtro de $n \times n$ en una combinación de convoluciones $1 \times n$ y $n \times 1$, donde $n=7$. La idea de este módulo es reducir las dimensiones de los kernels y el costo computacional, pero sin mucha o ninguna pérdida de representación del mapa de características. Este bloque se ilustra en la Figura 3.10.
- El bloque *Inception-C* donde el banco de filtros del módulo se hace más ancho en lugar de más profundo para eliminar los cuellos de botella. Por el contrario, si el módulo se profundiza, el tamaño del mapa de características se reduciría provocando pérdida de información. Este bloque se ilustra en la Figura 3.10.
- Entre los módulos *Inception A* e *Inception-B* y de este último con *Inception-C*, existen dos módulos de reducción que buscan reducir el tamaño del mapa de características mediante bancos de filtros. Estos módulos se llaman *Reduction-A* y *Reduction-B* respectivamente y se ilustran en la Figura 3.9 .
- Después de los bloques Inception, se encuentra una capa de agrupación de promedio global seguido de una capa completamente conectada y al final una función softmax (ecuaciones 1.11 y 1.12).

En el 2017 Szegedy et al. publicó en [66] Inception-ResNet, de la que se proponen 2 versiones: la primera “InceptionResNet-v1” con el mismo coste computacional que la Inception-v3 y la versión dos “Inception-Resnet-v2” (Figura 3.11) con mas en el número de filtros.

La red Inception-ResNet añade un nuevo bloque *Inception-A* (Figura 3.12) y modifica los módulos Inception y los convierte en módulos de bloques *Residual Inception* (Figura 3.13) que consta de un banco de convoluciones en paralelo con una conexión residual que adiciona a la salida del bloque la entrada del mismo.

Así la red Inception-ResNet-v2 mostrado en la Figura 3.11 consta de: una imagen de entrada representada matemáticamente como un tensor con dimensiones basadas en la ecuación 1.3.

- El bloque tallo está definido con anterioridad en la arquitectura de la red Inception-V3 y representado en la Figura 3.9.

- El bloque *Inception-A* presentado en [66] que se diferencia del utilizado en la Inception-v3 por el cambio de tamaño de uno de las capas convolucionales que cambia de 3×3 a 5×5 y es representado en la Figura 3.12.

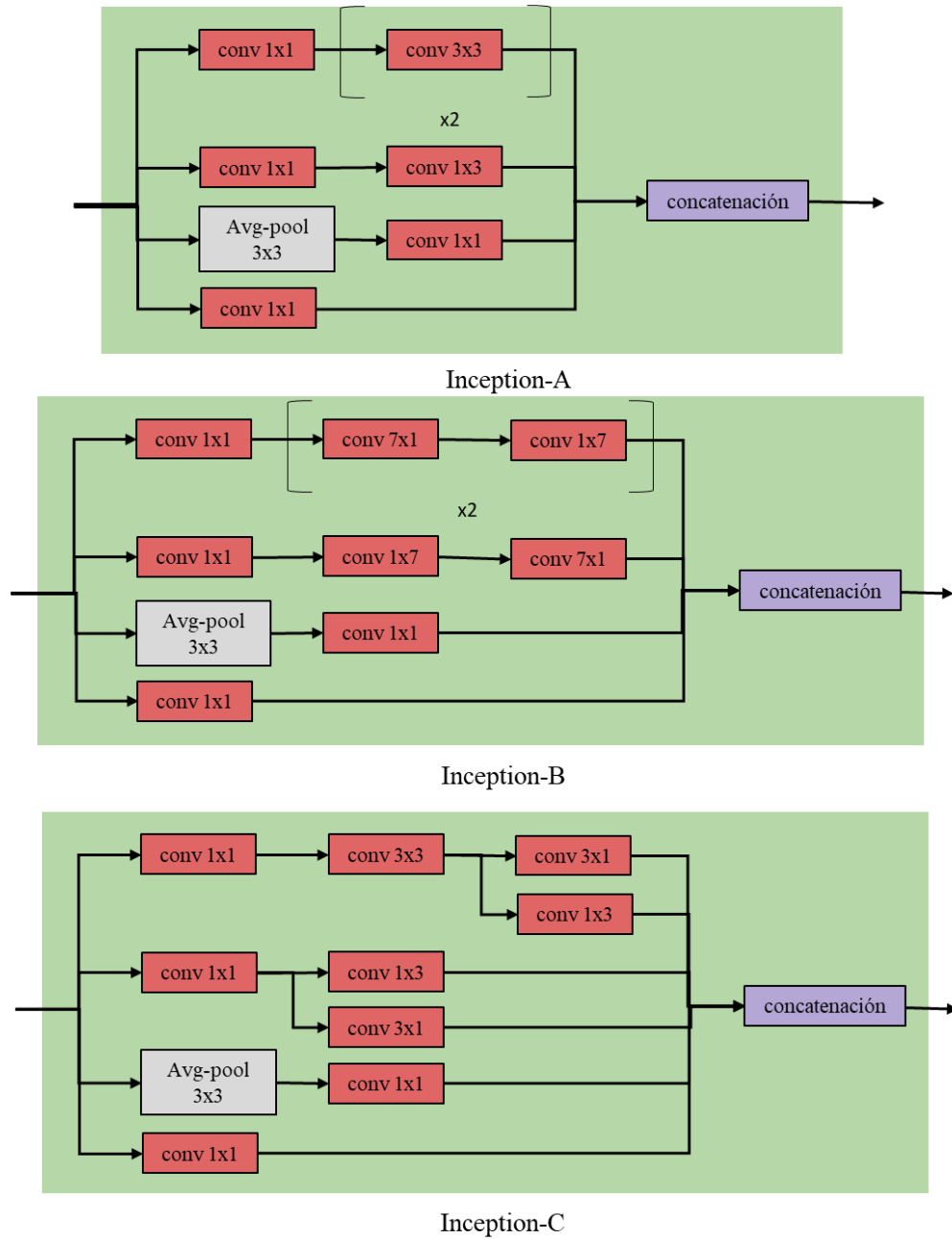


Figura 3.10 Módulos Inception-V3 [62].

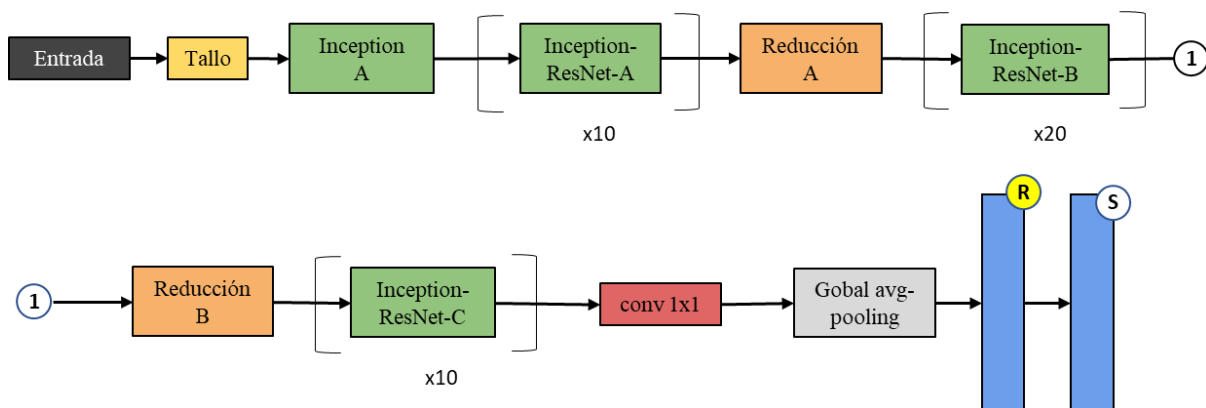


Figura 3.11 Arquitectura Inception-ResNet.[62].

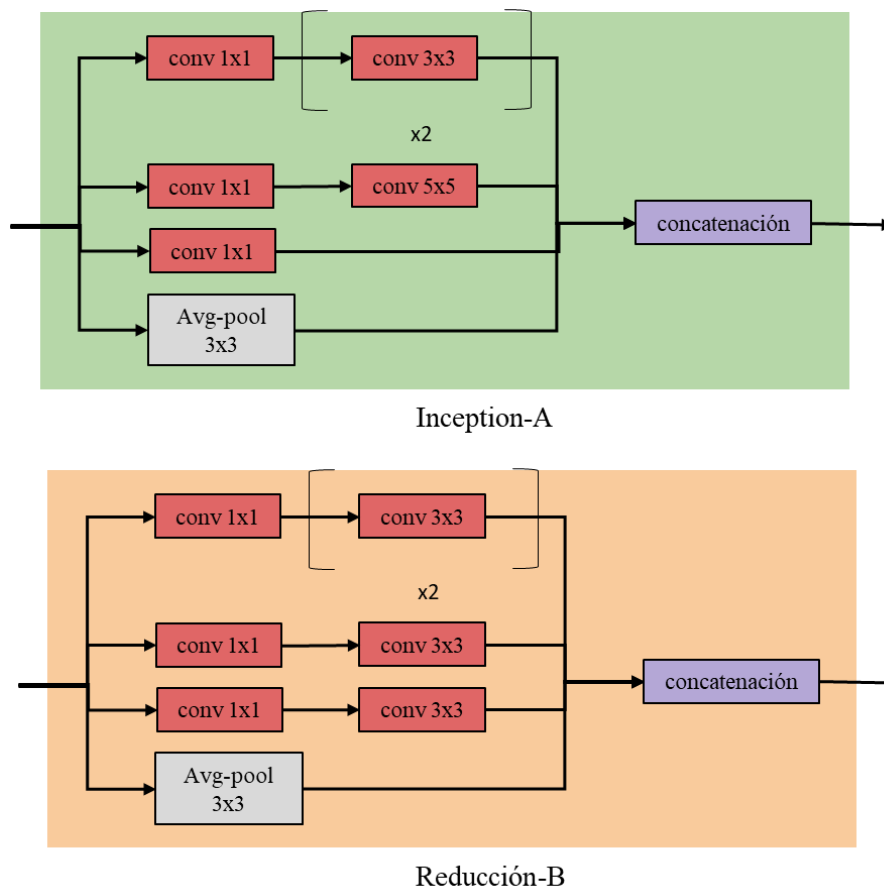


Figura 3.12 Módulos de tallo y reducciones para la InceptionResNet [62].

- Continúan los bloques *Inception-ResNet-A*, *Inception-ResNet-B* e *Inception-ResNet-C* representados en la Figura 4.12 que se intercalan con los bloques de reducción, *Reduction-*

A, igual al presentado en la Figura 3.10 y el *Reduction-B* que se presenta en la Figura 3.12

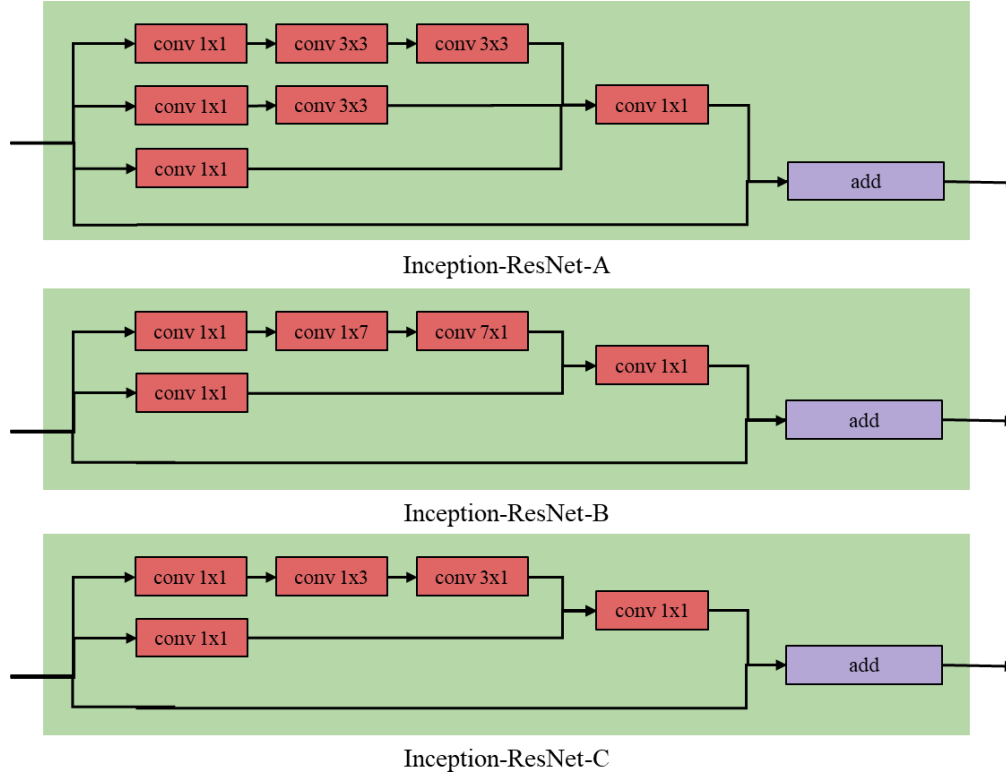


Figura 3.13 Módulos *Residual Inception* [62].

- Después de los bloques Inception, se añade una capa de convolución de reducción de 1×1 para seguir con un *avg-pooling* global dado por la ecuación 1.9 pero con un tamaño de kernel tal que $k=h$. Una vez terminada la sección de extracción de características, se continua con una capa completamente conectada y una función softmax (ecuaciones 1.11 y 1.12).

3.3 MobileNet-v2

En 2017 Andrew G. Howard publica en [67] MobileNet, una red que se basa en una arquitectura optimizada que utiliza convoluciones separables en profundidad para construir redes profundas de bajo costo computacional. En esta red, el trabajo de la capa convolucional se divide en dos subtarefas: primero, una capa convolucional profunda que filtra la entrada,

seguida de una capa convolucional puntual de 1×1 que combina estos valores filtrados para crear nuevas características. Esta hace lo mismo que una convolución normal, pero más rápido.

Las convoluciones separables se ilustran por la Figura 3.14, donde se puede observar que se componen de una convolución estándar en una en profundidad y una convolución puntual. Una convolución estándar filtra y combina entradas en un nuevo conjunto de salidas en una sola operación. La convolución separable en profundidad divide la convolución en dos operaciones donde los kernels tendrán el tamaño de $1 \times k$ y $k \times 1$. Esta factorización tiene el efecto de reducir drásticamente el cálculo y el tamaño de la red[67].

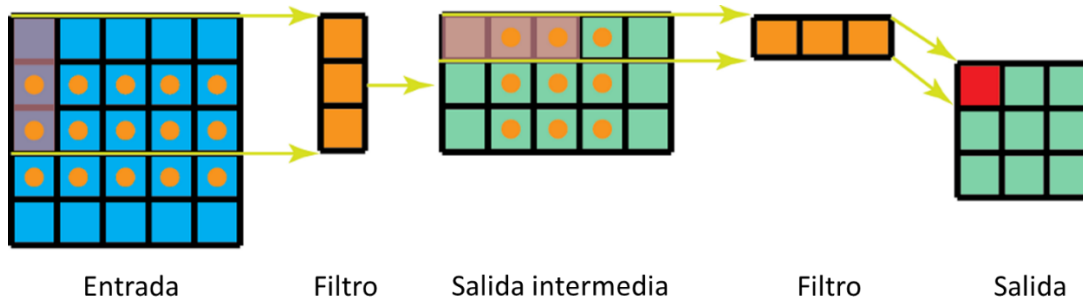


Figura 3.14 Convoluciones separables [68].

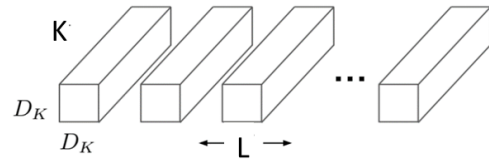
La convolución estándar tiene el costo de cálculo definido por:

$$CC_{std} = D_k \times D_k \times L \times K \times D_F \times D_F \quad (3.1)$$

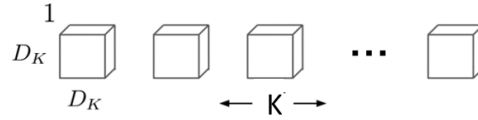
Si se analiza la Figura 3.15, el costo computacional de la convolución separable en profundidad es definido por:

$$CC_{sep} = D_k \times D_k \times K \times D_F \times D_F + L \times K \times D_F \times D_F \quad (3.2)$$

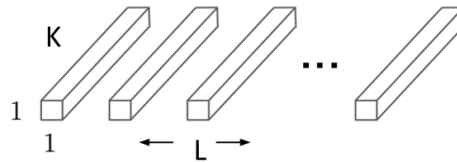
Donde D_K es la dimensión espacial del núcleo (se asume que es cuadrado), K es el número de canales de entrada, L es el número de canales de salida y D_F es el ancho y la altura espaciales de un mapa de características de entrada cuadrado [67].



a) Filtros de una convolución estándar



b) Filtros de una convolución en profundidad



c) filtros de convolución llamados convolución puntual en el contexto de convoluciones separables en profundidad

Figura 3.15 Ejemplo de costo de una convolución estándar y una separable en profundidad [67].

En MobileNet la función de activación utilizada es ReLU6, donde 6 es un límite superior que limita los valores de la función de activación de 0 a 6, la elección fue arbitraria y obtuvo buen desempeño. Los autores de [69] creen que incorporando un límite superior anima a su red a aprender antes las características abstractas [70] y está definida por:

$$ReLU6(valor) = \min(\max(0, valor), 6) = \begin{cases} 0, & \text{para } valor < 0 \\ valor, & \text{para } 0 < valor < 6 \end{cases} \quad (3.3)$$

Dos años después, Mark Sandler et al. proponen en [71] la segunda versión de esta red denominada MobileNet-v2, donde reduce el número de canales a comparación de la primer versión. Esta proyecta datos con un gran número de canales en un tensor con un número mucho menor de dimensiones. Este tipo de capa también se denomina capa de cuello de botella porque reduce la cantidad de datos que fluyen a través de la red.

Otra diferencia de MobileNet con MobileNet-v2 es la conexión residual que existe para ayudar al flujo del gradiente en la red. Esta conexión residual se aplica en el bloque *Inverted Residual Block*, que consta de una convolución con kernel de 1x1 seguida de una convolución

separable de 3×3 . Le sigue una capa convolucional de 1×1 y al último se le agrega la conexión residual que viene del inicio del bloque. La arquitectura de la MobileNet-v2 se muestra en Figura 3.16.

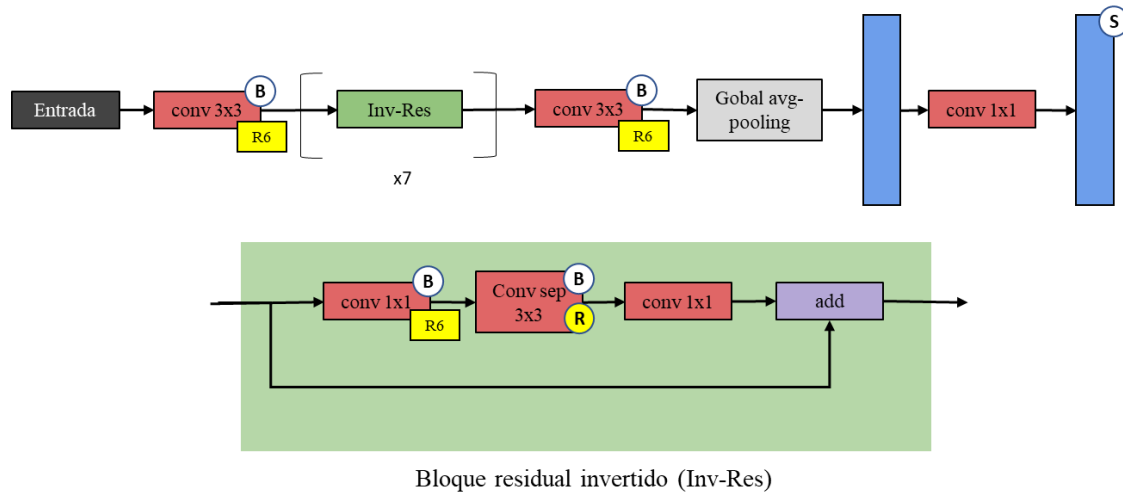


Figura 3.16 Arquitectura de MobileNet-v2 [62].

La arquitectura de MobileNet-v2 está conformada por:

- Una imagen de entrada representada por la ecuación 1.3.
- Una convolución regular de 3×3 .
- Siete bloques Inverted Residual Block.
- Una convolución regular de 1×1 .
- Una capa de agrupación promedio global, seguida de una capa de clasificación completamente conectada a una convolución equivalente 1×1 , y una capa de clasificación softmax.

3.4 Xception

Xception es una red neuronal convolucional cuya arquitectura se muestra en la Figura 3.17 y tiene 71 capas de profundidad. Xception es una adaptación de Inception, donde los módulos de Inception han sido reemplazados por convoluciones separables en profundidad como las usadas en MobileNet. Debido a que la hipótesis de esta arquitectura es una versión con mejor

rendimiento con el mismo número de parámetros que la hipótesis subyacente a la arquitectura Inception-v3, se le llamo Xception, que significa "*Extreme Inception*" [72].

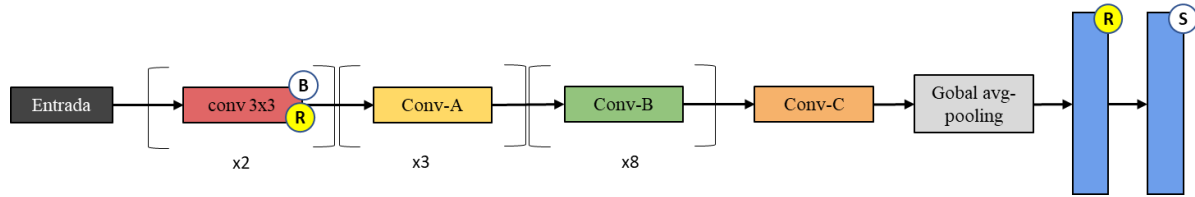
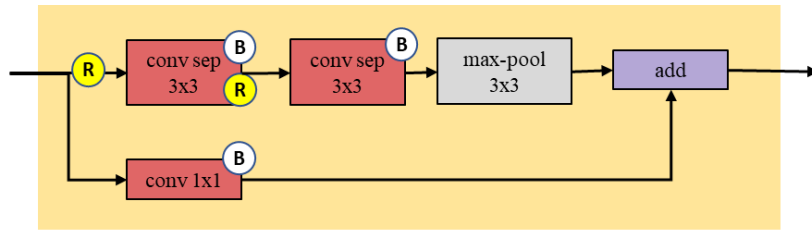


Figura 3.17 Arquitectura de Xception [62].

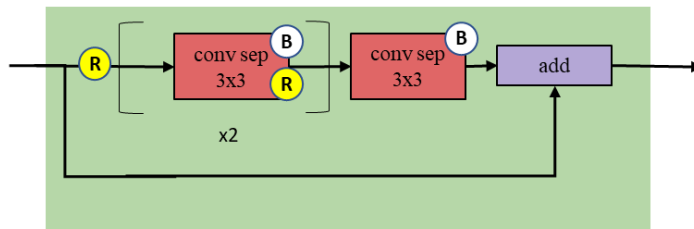
La red Xception consta de:

- Una imagen de entrada con dimensiones basadas en la ecuación 1.3.
- Dos capas consecutivas de convoluciones, la primera de $3 \times 3 \times 32$ y la segunda de $3 \times 3 \times 64$.
- El bloque Conv A (Figura 3.17 Arquitectura de Xception [62].) que es un bloque Inception simplificado, está conformado por dos capas convolucionales de $3 \times 3 \times 128$ seguidas de una capa de agrupación *max-pooling*. Además, el bloque cuenta con una conexión residual que le lleva una capa de convolución de $1 \times 1 \times 128$. Este bloque se repite tres veces cambiando el tamaño de canales en los kernels que cambian de 128 a 256 para el segundo bloque y 728 para el tercero.
- El bloque Conv B (Figura 3.17 Arquitectura de Xception [62].) que consta de tres capas convolucionales separables en profundidad en serie y a las cuales se les añade una conexión residual de principio a final del bloque. A estas convoluciones se les aplica un kernel de $3 \times 3 \times 728$ y se repite 8 veces.
- El bloque Conv C (Figura 3.17 Arquitectura de Xception [62].), similar a Conv A, está conformado por dos capas convolucionales de $3 \times 3 \times 728$ y $3 \times 3 \times 1024$ seguidas de una capa de agrupación *max-pooling*. Además, el bloque cuenta con una conexión residual que lleva una capa de convolución de $1 \times 1 \times 1024$. Sin embargo, a diferencia de Conv A, a este se le agregan dos capas convolucionales de $3 \times 3 \times 1526$ y $3 \times 3 \times 2048$.

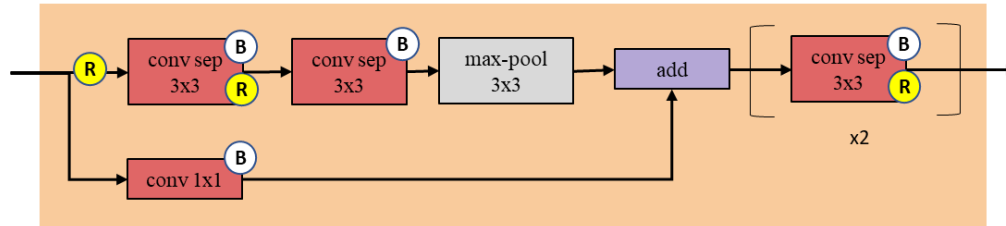
- Se finaliza la extracción de características continua una capa de agrupación de promedio global, una capa completamente conectada y finaliza con una función softmax (ecuaciones 1.11 a 1.12).



Conv-A



Conv-B



Conv-C

Figura 3.18 Bloques de la red Xception [62].

CAPÍTULO 4

PREPROCESAMIENTO DE DATOS Y REDES PROPUESTAS

En el capítulo anterior, se presentaron las redes neuronales profundas que son populares en el procesamiento de imágenes de histología y también MobileNet-v2. A partir de estas redes, en esta tesis se propusieron dos redes neuronales convolucionales para análisis de imágenes de histología de tejido mamario. La primera se denominó Light-hist, una red inspirada en MobileNet-v2 con un conjunto de convoluciones en paralelo aunadas con una conexión residual. La segunda se denominó Light-hist-v2, que es similar a la primera versión, pero se añade una convolución en paralelo de diferente tamaño de kernel para una mayor extracción de características. MobileNet fue seleccionada como base para el diseño de las redes a proponer debido a que la contribución de esta tesis es desarrollar clasificadores de imágenes de histología mamaria con baja cantidad de parámetros para que estas puedan ser procesadas por dispositivos móviles o dispositivos ASIC.

Para presentar estas redes, este capítulo se organiza de la siguiente manera: la sección 4.1 presenta el preprocesamiento de los datos utilizados para entrenar el modelo de CNN y en la sección 4.2 se describen las redes propuestas.

4.1 Preprocesamiento

La tinción más utilizada en los cortes de tejidos es la HE, cuyo método se menciona en el capítulo 1. Esta tinción se utiliza para diferenciar los núcleos teñidos con color púrpura, así como otras estructuras de tejido teñidas con color rosa y rojo para que los patólogos valoren la estructura general del tejido. No obstante, existe variabilidad entre las imágenes adquiridas debido a los diferentes protocolos de tinción, métodos de captura, iluminación y materias primas. La Figura 4.1 muestra como existe variabilidad aun siguiendo los mismos protocolos de tinción, solo variando la muestra de paciente a paciente. Por ello, el primer paso para preprocesar las imágenes es realizar una normalización del color.

Las bases de datos de imágenes de histología cuentan con pocas imágenes para que una red generalice, en comparación de otras bases de datos como Imagenet [73] que contiene decenas o cientos de miles de imágenes. Por ello es necesario incrementar la cantidad de imágenes de

la base de datos, además, estas imágenes presentan variaciones como en su tinción, brillo, etc. Una forma de solucionar este problema es aplicando técnicas normalización y aumentación de datos.

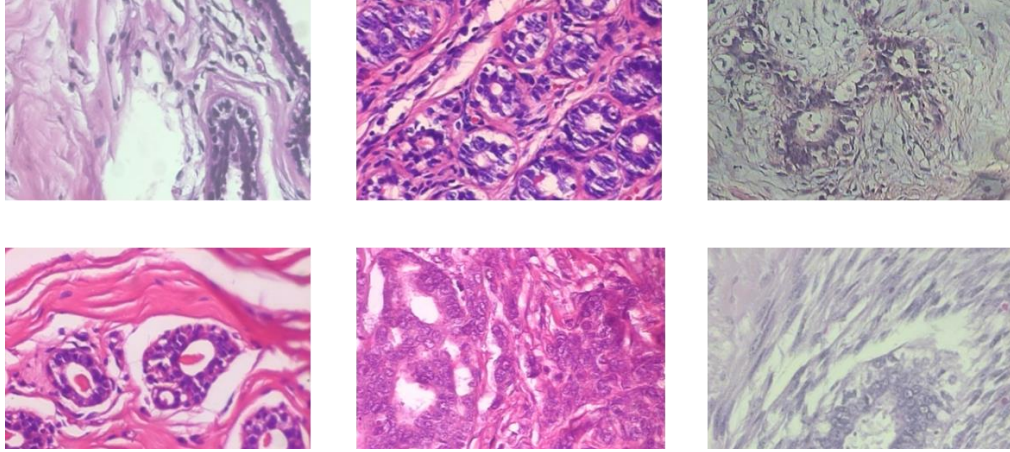


Figura 4.1 Imágenes tomadas de la base de datos BreakHist.

4.1.1 Normalización de tinción

La normalización de color de imágenes de histología tenidas con HE es el proceso de reducir la variación de color para obtener una mejor consistencia en la información de estas imágenes. Existen varios enfoques de normalización de color en imágenes de histología, por ejemplo los presentados por Macenko et al. [74], Reinhard et al. [75] y Vahadane et al. [76]. De los antes mencionados en este proyecto se utiliza el propuesto por Macenko, debido a que ha generado excelentes resultados en varios estudios como en [25], [34], [36], [40], [55], [77]–[81]. En estos trabajos se comparan los resultados de clasificación de CNN's con imágenes con normalización de tinción contra las mismas redes con imágenes sin normalización.

El método Macenko está basado en Descomposición de Valores Singulares (SVD por sus siglas en inglés). En este método, una función logarítmica se utiliza para transformar adaptativamente la concentración de color de la imagen histopatológica original en una imagen de densidad óptica como se indica en:

$$OD = -\log \frac{ent}{ent_0} \quad (4.1)$$

Donde OD es la imagen resultante después de normalizar la los colores de la tinción en una imagen e ent_0 es la intensidad de iluminación. Un ejemplo de este método se muestra en la Figura 4.2 inciso c.

4.1.2 Normalización de imágenes.

Otro paso del preprocesamiento es la normalización de la intensidad [82]. El propósito es obtener el mismo rango de valores para cada imagen de entrada antes entrenar al modelo CNN. Las imágenes de entrada se normalizan a la distribución normal estándar mediante la normalización mínima-máxima al rango de intensidad de [0, 1], que se calcula como:

$$Q_{norm} = \frac{q_{int} - q_{min}}{q_{max} - q_{min}} \quad (4.2)$$

donde Q_{norm} es una imagen resultante después de normalizar la intensidad de una imagen. p_{min} y p_{max} son los valores de intensidad mínima y máxima de la imagen de entrada en la ecuación 4.2.

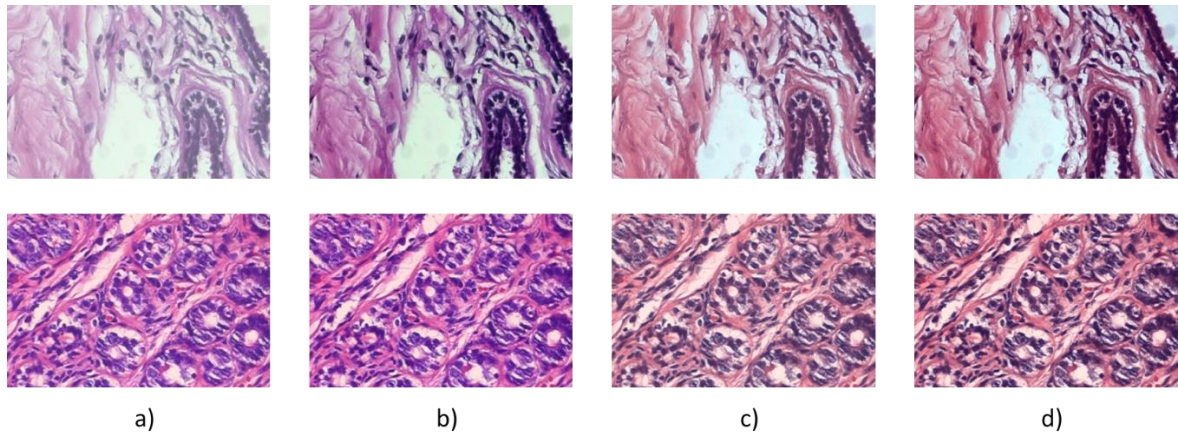


Figura 4.2 Preprocesamiento de imágenes de histología, donde se muestran: a) Imagen original, b) Normalización de intensidad (X_{norm}), c) Normalización de tinción (OD), d) Normalización de tinción e intensidad (OD y X_{norm})

4.1.3 Aumentación de datos

Debido al tamaño limitado del conjunto de datos de imágenes de histología para el entrenamiento, la CNN es propensa a sobre ajustar los pesos de la red, lo que conlleva a una

baja tasa de reconocimiento de imágenes de histología y resultados de diagnóstico insatisfactorios [83]. El método para sobrellevar esta problemática es la aumentación de datos, cuyo propósito es generar más datos basado en un conjunto de entrenamiento existente [58]. Basado en librerías ya definidas en keras con tensorflow, se utilizan diferentes técnicas de aumento de datos [77], como rango de corte, rango de zoom, volteo horizontal y vertical, entre otros creando así nuevas imágenes que serán utilizadas en el entrenamiento.

A continuación, se explican los parámetros y se muestra ejemplos de las transformaciones realizadas con dichos parámetros, estas son aplicadas a una imagen de muestra del conjunto de datos BreakHis, como la mostrada en la Figura 4.3.

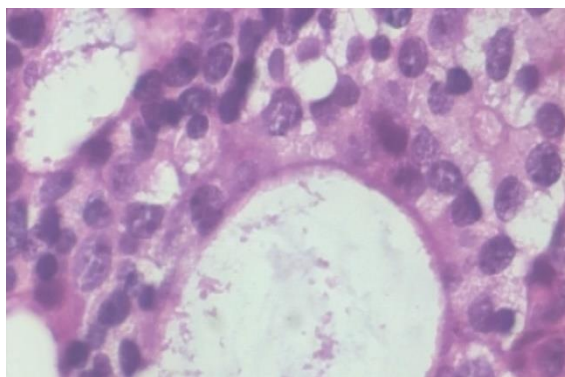


Figura 4.3 Imagen de muestra a la cual se le aplicaran las transformaciones.

Factor de cambio de escala (*rescale*): a pesar del nombre dado, no se aplica ningún cambio de escala. Su función es transformar los datos [84]. Para una imagen que contiene 3 componentes (rojo, verde y azul), y los pixeles tienen valores de 0 a 255, *rescale* es un factor de 1/255 para transformar cada valor del pixel de la siguiente manera:

$$[0, 255] \rightarrow [0, 1] \quad (4.3)$$

Rango de corte (shear_range): es la intensidad de corte (ángulo de corte en el sentido contrario a las agujas del reloj) [84]. En la Figura 4.5 se muestra un ejemplo del rango de corte, En la Figura 4.4 y Figura 4.5 se ilustra de una manera exagerada el efecto del corte.

Para realizar el rango de corte se aplica una matriz de transformación lineal como la siguiente a cada uno de los canales de la imagen:

$$imagen_{x,y} = \begin{bmatrix} 1 & Sh \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} \quad (4.4)$$

donde Sh es el rango en que la imagen se transforma.

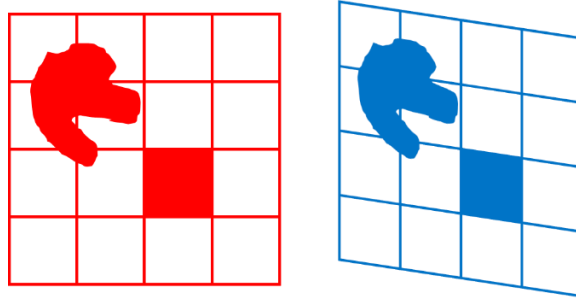


Figura 4.4 Visualización del rango de corte

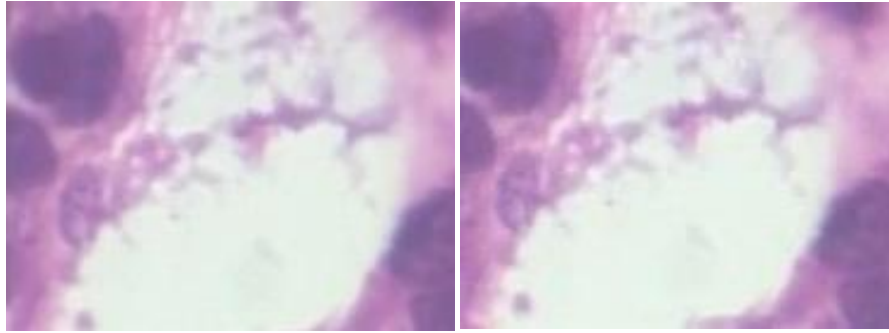


Figura 4.5 Ejemplo de rango de corte

Rango de acercamiento (zoom_range): rango para la transformación de escala aleatoria y se refiere a distintos acercamientos y alejamientos de la imagen original como se muestra en la Figura 4.6.

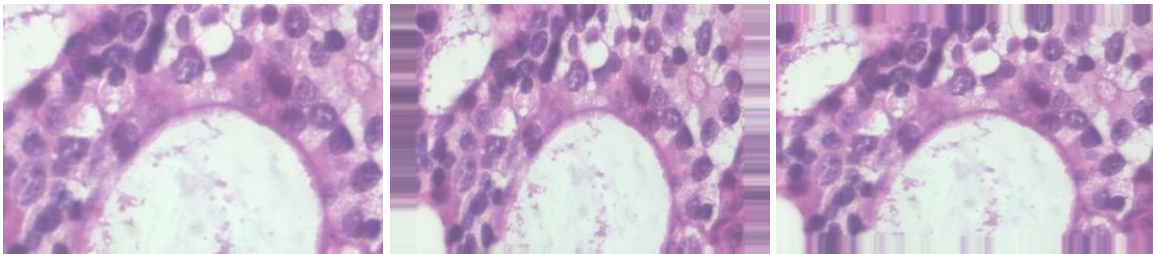


Figura 4.6 Ejemplos de rango de acercamiento. A la izquierda, una imagen con acercamiento. En el centro una imagen con alejamiento en el eje horizontal. A la derecha una imagen con alejamiento en el eje vertical.

Este cambio de escala está dado por:

$$Imagen_nueva(x, y) = Imagen(x/a_x, y/a_y) \quad (4.5)$$

Donde a_x es el aumento en el eje x y a_y es el aumento en el eje y. Si a_x y a_y son mayores que uno, entonces se tiene de un acercamiento y se debe aplicar una interpolación bilineal para rellenar los espacios vacíos, y si son menores que uno es una reducción y se debe aplicar un suavizado para eliminar ruido y bordes abruptos que se pueden presentar después de la reducción.

Rango de desplazamiento de ancho (width_shift_range): desplaza la imagen hacia la izquierda o hacia la derecha (desplazamientos horizontales). El valor positivo desplazará la imagen hacia el lado derecho y el valor negativo desplazará la imagen hacia el lado izquierdo, En la Figura 4.7 se muestran ejemplos de desplazamientos y se realiza con lo siguiente:

$$Imagen_nueva(x, y) = Imagen(x + xdesp, y) \quad (4.6)$$

donde $xdesp$ es el valor aleatorio de desplazamiento horizontal.

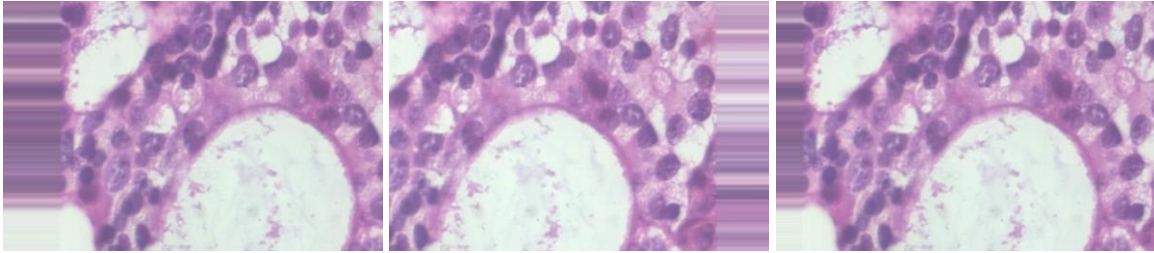


Figura 4.7 Ejemplos de desplazamientos horizontales. A la izquierda, un desplazamiento a la derecha. En el centro un desplazamiento a la izquierda. A la derecha un desplazamiento a la derecha menor.

Rango de desplazamiento de altura (height_shift_range): similar al anterior, pero este desplaza la imagen hacia arriba o hacia abajo. En la Figura 4.8 se muestran ejemplos de desplazamientos verticales y se realiza con lo siguiente:

$$Imagen_nueva(x, y) = Imagen(x, y + ydesp) \quad (4.7)$$

Donde $ydesp$ es el valor aleatorio de desplazamiento vertical.

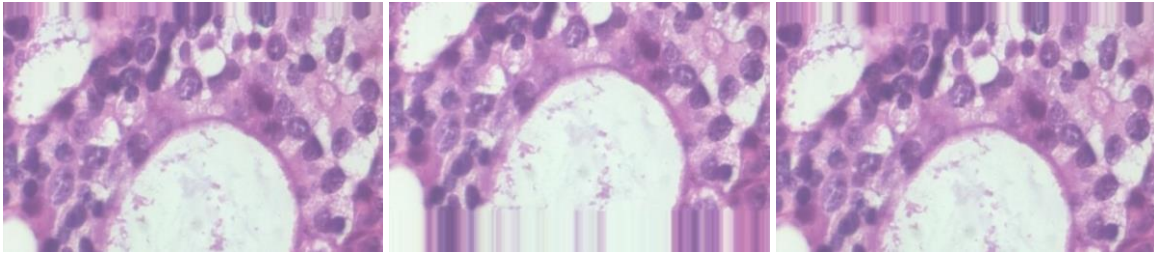


Figura 4.8 Ejemplos de desplazamientos verticales.

Rango de brillo (brightness_range): rango para elegir un valor de cambio de brillo, donde se incrementa o disminuye el valor de intensidad de cada pixel por igual en todos los canales de la figura, se ilustra un ejemplo en la Figura 4.9 y se realiza con lo siguiente:

$$Imagen_nueva(x, y) = Imagen(x, y) + brillo \quad (4.8)$$

Donde el *brillo* es un escalar que va de -255 a 255, que es la intensidad a sumar en toda la matriz.

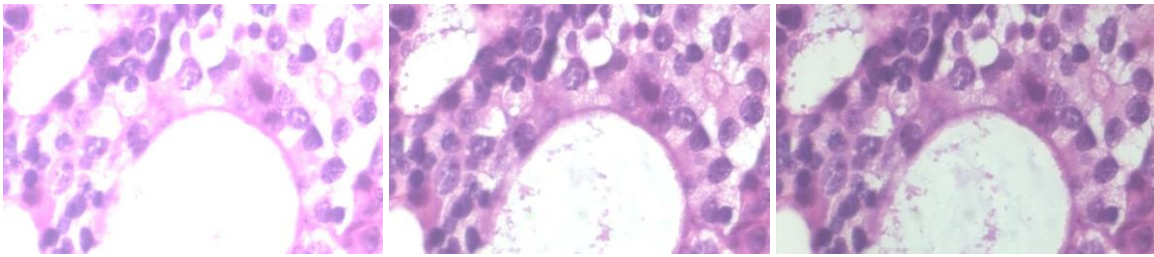


Figura 4.9 Ejemplos de cambio de brillo. A la izquierda aumento de brillo. En el centro la imagen original. A la derecha disminución de brillo.

Volteo horizontal (horizontal_flip) definido por la ecuación 4.9 voltea las entradas de forma aleatoria horizontalmente y *volteo vertical (vertical_flip)* y se define por la ecuación 4.10. Como se muestra en la Figura 4.10, se voltean las entradas verticalmente.

$$Imagen_nueva(x, y) = Imagen((x_{max} - x) + 1, y) \quad (4.9)$$

$$Imagen_nueva(x, y) = Imagen(x, (y_{max} - y) + 1) \quad (4.10)$$

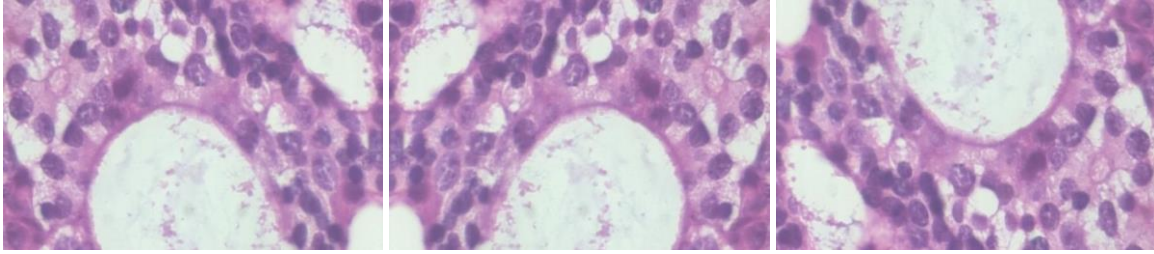


Figura 4.10 Ejemplo de volteo de imágenes. A la izquierda volteo horizontal. En el centro la imagen original. A la derecha volteo vertical.

Rotación (rotation_range): Realiza rotaciones aleatorias como se muestran en la Figura 4.11, estas se realizan con lo siguiente:

$$Imagen_nueva(x,y) = Imagen \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (4.11)$$

dónde θ es el ángulo de rotación.

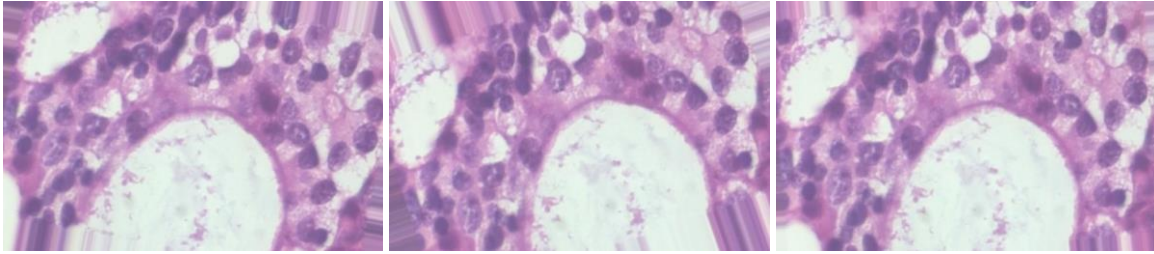


Figura 4.11 Ejemplo de rotaciones de imágenes.

Modo de relleno (fill_mode): Establece reglas para píxeles recién desplazados en el área de entrada. Al realizar corrimientos, rotaciones, acercamientos y cortes, se forman píxeles vacíos por lo que se tiene que generar un relleno para que sean de mayor provecho las imágenes. En el caso de la aumentación de datos utilizada en esta tesis, se utilizó el relleno de reflejo (*reflect*), que completa el corrimiento de la imagen con valores iguales a los de los píxeles más cercanos. Un ejemplo de este relleno es cuando los espacios vacíos generados un corrimiento dado por la ecuación 4.6 se le agrega un reflejo de la imagen dado por la ecuación 4.9, este ejemplo se muestra en la Figura 4.12.

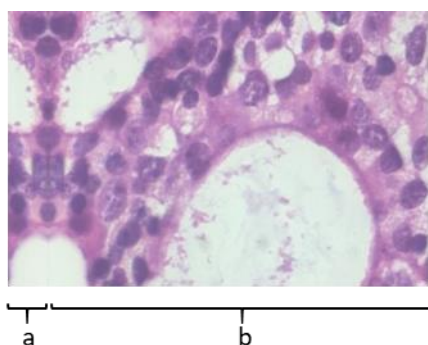


Figura 4.12 Ejemplo del modo de relleno reflect. a) Relleno reflejo. b) Imagen con corrimiento a la derecha.

Parámetros del aumento de datos. Para el caso de esta tesis, el aumento de datos se realizó en usando la función ImageDataGenerator disponible con la librería de Tensroflow usando lenguaje de programación Python y los parámetros de las transformaciones son los que se muestran en Tabla 4.1.

Tabla 4.1 Parámetros de aumentación de datos.

Parámetro	Valor
rescale	1/255
shear_range	0.2
zoom_range	0.2
width_shift_range	0.2
height_shift_range	0.2
brightness_range	[0.5,1.5]
horizontal_flip	True
vertical_flip	True
rotation_range	30
fill_mode	"reflect"

A continuación se describe el valor de cada parámetro:

- *Reescale*: normaliza los valores de cero a uno.
- *Shear range*, *zoom_range*, *width_shift_range* y *height_shift_range* se establecieron en 0.2 para agregar un 20% de rango, por ejemplo, un corrimiento hasta del 20% de la longitud de la imagen.
- *Brightness_range* se estableció para bajar la intensidad hasta un 50% y de 1.5 para aumentar la intensidad hasta en un 50%.
- *Horizontal_flip* y *vertical_flip* se colocan como true para que se realicen ambas operaciones.

- *Rotation_range* se estableció para rotar la imagen 30°.
- *Fill_mode* indica que se utilizó un parámetro en el modo de relleno.

4.2 Redes propuestas

Basados en la arquitectura de MobilNetV2 que se muestra en la Figura 3.16, se propusieron dos nuevas redes que buscan mejorar la eficiencia computacional en la clasificación de imágenes histológicas, pero sin incrementar significativamente el costo computacional de la red.

La primera red propuesta (Light-hist) se muestra en la Figura 4.14 y es muy similar a MobileNet, pero con una serie de convoluciones separables en profundidad en paralelo con el fin de generar una mayor cantidad de características [22], [65]. Estas convoluciones tienen kernels de tamaño de 1x1, 3x3 y 5x5. La decisión de incluir kernels de diferentes tamaños es tratar de incluir la mayor cantidad de características de las imágenes.

Los kernels de 1x1 tienen un campo receptivo más pequeño, por lo que analiza pocos pixeles. De esta manera, entre más crece el kernel, es mayor la cantidad de pixeles que analiza. Esto significa que las características extraídas por el kernel pequeño son locales y entre más crece el kernel, se van volviendo más globales. La necesidad de extraer diferentes tipos de características es porque las imágenes de histología contienen características abstractas de diferentes dimensiones que se extraen a partir de campos receptivos de diferentes tamaños. Además, estas imágenes tienen aumentos que van desde 40X hasta 400X, lo que causa que los patrones obtenidos en las imágenes también sean de diferente tamaño.

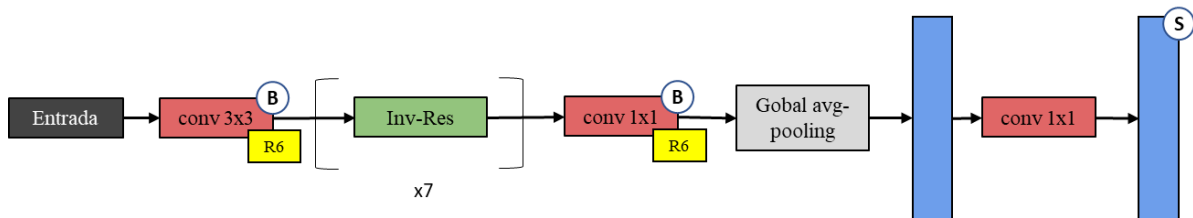


Figura 4.13 Arquitectura base de las redes LightHist.

De esta manera, la red LightHist está compuesta de la siguiente manera

- La entrada de la red es una imagen RGB, representada matemáticamente como un tensor con dimensiones basadas en la ecuación 1.3.
- La siguiente capa es una convolución con un kernel de 3x3 con 32 filtros dada por la ecuación 1.5, para la extracción de características y reducción de dimensiones. Esta capa tiene normalización por lotes que básicamente implica agregar un paso adicional entre la neurona y la función de activación para normalizar la salida, donde la media y la varianza de todo el conjunto de entrenamiento se utilizan para normalización de la salida. Se usa ReLU6 como función de activación definida por la ecuación 3.3, esto debido a que en los experimentos dieron como resultado que al usar esta función de activación las redes generalizan con mayor rapidez.
- Las siguientes capas corresponden a 7 bloques *Inverted Residual Block* que consiste en una convolución de 1x1 encargada de una reducción de dimensiones para que sea menos costoso computacionalmente. Generalmente se utilizan antes de convoluciones computacionalmente costosas. Seguida de esta capa convolucional, se encuentran convoluciones separables en profundidad, definidas por la ecuación 1.5, con la diferencia de que el kernel se divide en dos más pequeños para reducir el número de operaciones como se ve en el capítulo 3 en la red MobileNet-v2. Estas convoluciones se realizan con diferente tamaño de kernel, de tamaños 1x1, 3x3 y 5x5, ya que kernels de diferentes tamaños detectan entidades de diferente tamaño en el mapa de características de entrada. En paralelo a las convoluciones se agrega una capa de agrupación del tipo *average-pooling* seguido de una convolución de 1x1 para reducir el tamaño del mapa de características.
- Después se añade una convolución de 1x1 con normalización por lotes y ReLU6, para la reducción de características.
- Una capa de agrupación promedio global para minimizar el sobreajuste al reducir el número total de parámetros en la red. Esta capa reduce cada mapa de características $H \times W$ a un solo número simplemente tomando el promedio de todos los valores.
- Después se añade una capa de convolución de 1x1 para reducir las características.

- Al final, una capa de clasificación completamente conectada y una capa de clasificación softmax.

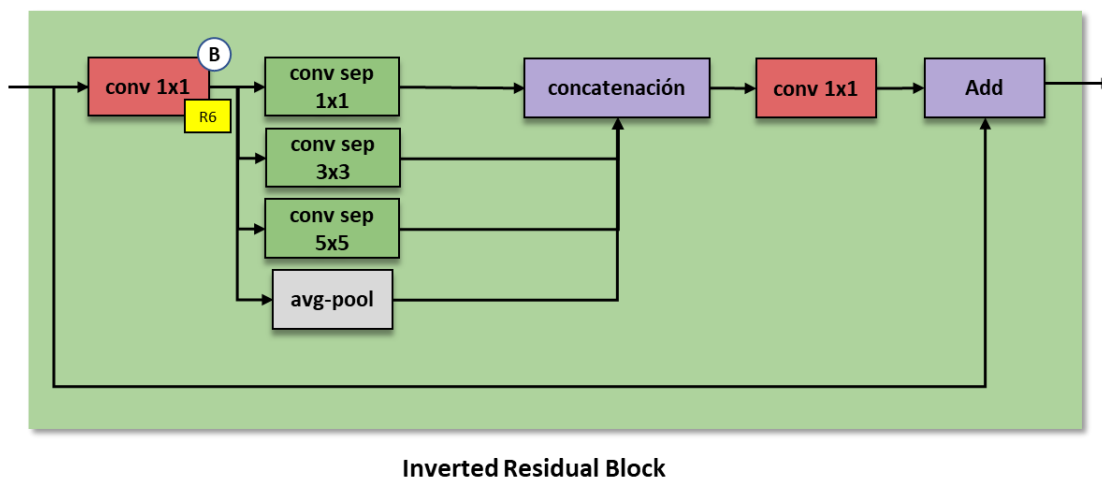


Figura 4.14 Arquitectura del bloque Inverted Residual Block de Light-hist.

La segunda red propuesta denominada Light-hist -v2 se observa en la Figura 4.15 y es básicamente la red Light-hist, pero se le añadió otra convolución separable en profundidad dada por de tamaño de 7x7, con en el fin aumentar la cantidad de características más genéricas de las imágenes histológicas de cáncer de mama.

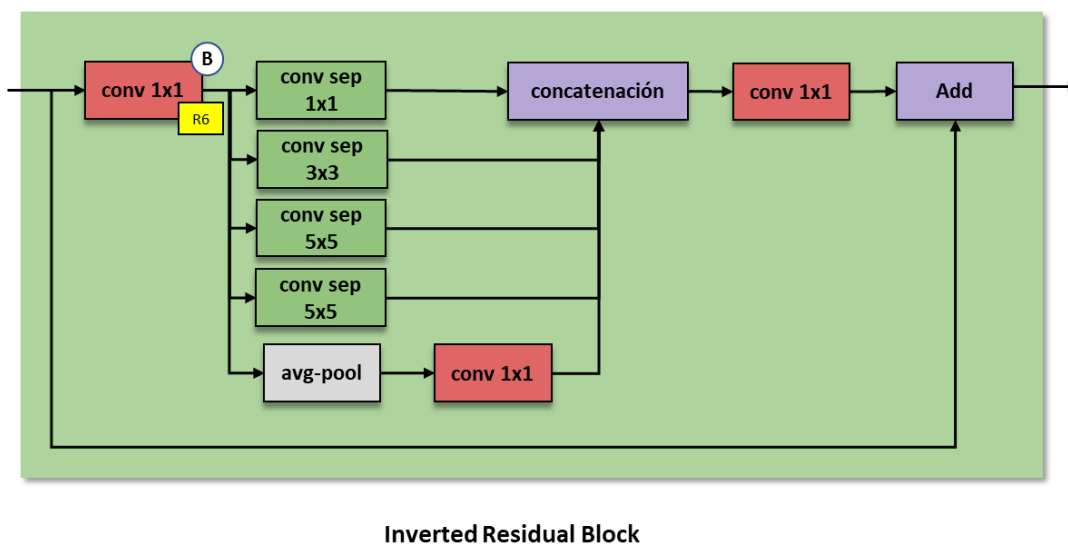
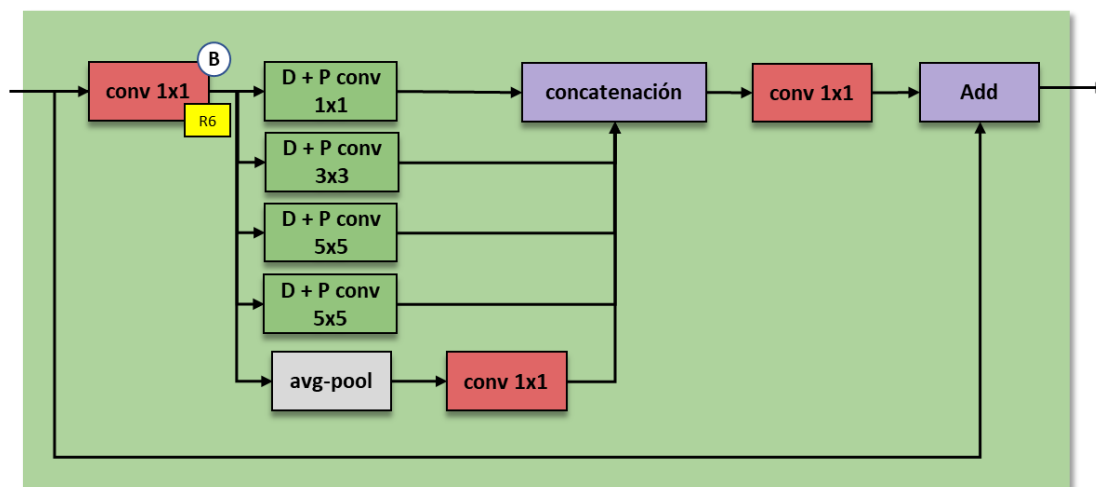


Figura 4.15 Arquitectura del bloque Inverted Residual Block de Light-hist-v2.

Adicional a estas propuestas, se realizó una tercera llamada Light-hist-v3, la cual se muestra en la Figura 4.16 que sustituye las convoluciones separables en profundidad por las llamadas convoluciones separables usadas en Xception. Esta se realizó como variable de las dos anteriores buscando incluir las ventajas que la red Xception tiene. Sin embargo, esta última propuesta fue descartada por tener un rendimiento inferior a las otras dos.



Inverted Residual Block

Figura 4.16 Arquitectura del bloque Inverted Residual Block de Light-hist-v3.

CAPÍTULO 5

RESULTADOS

Una vez diseñadas las redes propuestas, estas fueron probados con la base de datos de BreakHis descrita en el capítulo dos. Se utilizaron 70% de las imágenes para entrenamiento, 15% para prueba y 15% para validación, tal y como se sugiere en [21], donde se presenta la base de datos. Se realizaron diversos experimentos para desarrollar las redes de Light-hist, Light-hist-v2 y Light-hist-v3. Además, se realizaron diversos experimentos con BreakHis para comparar estas redes con las más populares en la literatura para definir la contribución de estas redes en el conocimiento científico.

Por ello, este capítulo se divide en la sección 5.1 que muestra los resultados de las redes Light-hist, Light-hist-v2 y Light-hist-v3 en términos de las métricas de exactitud, precisión, exhaustividad y puntuación -F1. La sección 5.2 presenta un análisis comparativo de las redes propuestas con los publicados en el estado del arte.

5.1 Resultados de desempeño de las redes propuestas

En esta sección se presentan los resultados obtenidos al realizar pruebas de entrenamiento a las redes propuestas, estos se basan en una clasificación binaria de imágenes de histología de cáncer de mama, donde para las pruebas se contabilizan los TP, FP, TN y FN usando las métricas de las ecuaciones 2.1-2.4

La primera prueba consistió en entrenar las redes con 70 épocas para observar el comportamiento del aprendizaje de las redes. En la Tabla 5.1 se observan los resultados de esta prueba y en la Tabla 5.2 las métricas de exactitud, precisión, exhaustividad y puntuación -F1.

Tabla 5.1 Resultados de las redes propuestas con las imágenes de pruebas a 70 épocas.

Red	TP	FP	TN	FN
Light-hist	288	84	678	74
Light-hist-v2	219	153	715	37
Light-hist-v3	304	68	580	172

Como se observa en la Tabla 5.2, la red Light-hist-v2 obtiene desempeño más alto con una *exactitud* de 83.10%, que es la métrica más popular en el estado del arte para comparar los métodos dedicados a la clasificación de imágenes de histología. En segundo lugar, está la red Light-hist-v3, pero esta red presenta altos FN como se muestra en la Tabla 5.1, lo que significa que clasifica más imágenes de tumores malignos (cáncer de mamá) como tumores benignos. En esta situación representa alto riesgo para un paciente, por eso se decide descartar esta red. En tercer lugar, se tiene la red Light-hist que presenta más FN pero menos FP que significa que clasifica imágenes como tumores malignos siendo estos benignos.

Tabla 5.2 Métricas de las redes propuestas con las imágenes de pruebas a 70 épocas

Red	<i>Exactitud</i>	<i>Precisión</i>	<i>Exhaustividad</i>	<i>Puntuación-F1</i>
Light-hist	69.04	94.89	51.76	66.98
Light-hist-v2	83.10	58.87	85.55	69.75
Light-hist-v3	78.65	81.72	63.87	71.70

Dentro de los artículos reportados en el libro Machine Learning in Medical Imaging [82] por Q. Wang et. Al, se recomienda realizar un entrenamiento a 1000 épocas para mejorar el desempeño, pero sin llegar a un sobreajuste de la red. Por lo que se realizó una segunda prueba donde se utilizó un máximo de 1000 épocas, pero con una detención temprana a 200 épocas. Esto significa que el entrenamiento se detiene si el error ya no disminuye después de 200 épocas, donde se utilizan los pesos con el menor error.

En la Tabla 5.3 se presentan los resultados del entrenamiento con más épocas de las redes Light-hist y Light-hist-v2 y en la Tabla 5.4 se muestran las métricas obtenidas. Se puede observar que con este entrenamiento aumenta notablemente el desempeño de las redes llegando a no tener imágenes de tumores malignos mal clasificadas y solo 3 imágenes de tumores benignos mal clasificadas dentro del conjunto de prueba.

Tabla 5.3 Resultados de las redes propuestas con las imágenes de pruebas a 1000 épocas.

Red	TP	FP	TN	FN
Light-hist	330	42	716	36
Light-hist-v2	369	3	752	0

Tabla 5.4 Métricas de las redes propuestas con las imágenes de pruebas a 1000 épocas.

Red	<i>Exactitud</i>	<i>Precisión</i>	<i>Exhaustividad</i>	<i>Puntuación-F1</i>
Light-hist	93.06	88.70	90.16	89.43
Light-hist-v2	99.73	99.19	100.0	99.59

Para evaluar las redes con un análisis estadístico y garantizar que los resultados sean independientes de la división entre datos de entrenamiento y datos de prueba se utilizó la validación cruzada (presentada en el capítulo 2), dividiendo el conjunto de entrenamiento en 5 carpetas. Se utilizaron 5 carpetas ya que, entre más carpetas, aumenta la varianza y el tiempo de entrenamiento. En la Tabla 5.5 se presentan las métricas de LightHist-v2 al realizar la evaluación a 70 épocas, quien fue la red que logro un mejor desempeño al evaluarlo con los datos de prueba como se observa en la tabla Tabla 5.2. Se puede notar que obtiene una exactitud promedio similar a los obtenidos con las imágenes de prueba, por lo que se asume que se tiene una generalización del aprendizaje.

Tabla 5.5 Resultados LightHist-v2 usando validación cruzada a 70 épocas.

Carpeta	Exactitud	Precisión	Exhaustividad	Puntuación-F1
h1	83.4688	100.0000	65.4717	79.1334
h2	87.1725	69.6532	86.6906	77.2436
h3	90.5320	78.1609	90.3654	83.8213
h4	81.5149	90.5172	64.6817	75.4491
h5	90.6222	88.4726	82.7493	85.5153
Promedio	86.6621	85.3608	77.9918	80.2325

Un aspecto importante a mencionar es que los métodos supervisados propuestos en la literatura no presentan una validación cruzada o algún método que compruebe el aprendizaje del método.

5.2 Comparación de la red propuesta en comparación con el estado del arte

En esta sección se hace una comparación de las redes LightHist propuestos contra los publicados en el estado del arte que usan la base de datos BreakHis. Las pruebas se realizaron

para probar la capacidad de clasificación de las redes ya que en estos artículos no se reporta como realizaron los autores la partición de los datos, preprocesamiento de los datos, cantidad de parámetros, cantidad de operaciones ni validación cruzada.

Por ello, las comparaciones se realizaron en esta tesis mediante pruebas de las redes en la misma plataforma de cómputo en vez de utilizar directamente los desempeños presentados en los artículos. Es decir, las comparaciones se basaron en entrenar las redes populares en el estado del arte y las redes propias bajo las mismas condiciones, que son:

- La base de datos está dividida en 70% de los datos para entrenamiento, 15% para validación y 15% para pruebas
- Las redes fueron entrenadas a 1000 épocas con detención temprana.
- La computadora tiene un procesador Intel i7-8750H un procesador de video NVIDIA GeForce GTX 1060 con diseño Max-Q.
- La comparación se realizó con las métricas de Total de parámetros, exactitud, precisión, exhaustividad y puntuación – F1. Estas métricas fueron seleccionadas para comparar tanto el desempeño de las redes como su eficiencia computacional.

En la Tabla 5.6 se presenta una comparación de los resultados presentados en los artículos contra los datos obtenidos de las redes populares en la literatura y LightHist y LightHist-v2.

Tabla 5.6 Comparación de la red propuesto con los existentes en el estado del arte.

Red	Total de operaciones	Parámetros	Exactitud	Precisión	Exhaustividad	Puntuación-F1
LightHist	6993040	5532933	93.06%	88.70%	90.16%	89.43%
LightHist-v2	8174327	6655587	99.73%	99.19%	100	99.59%
InceptionResNetV2 [66]	2974763	55911649	87.81%	87.36%	78.31%	82.59%
Xception [72]	3956117	39737897	89.41%	94.35484	78.17%	85.50%
MobileNetV2 [71]	2530654	14056513	54.18%	98.65%	41.84%	58.76%
Método 2 basado en InceptionResNetV2 [52]	-	-	97.90%	98.46%	99.13%	98.47
InceptionResNetV2 (augmented data)	-	-	99.79%	99.66%	99.95%	99.81%

[52]						
Método 3 basado en InceptionResNetV2 [43]	-	-	86%	-	-	-
Método 4 basado en InceptionResNetV2 [77]	-	-	92.40%	-	-	-
Método 2 basado en Xception [77]	-	-	86.37%	-	-	-
MobileNetV2 [77]	-	-	92.4%	-	-	-

En muchos de los artículos no se reportan la cantidad de parámetros y de operaciones, además que no presentan suficiente información para calcular estas métricas. No obstante, estas redes tienen más capas que las redes base de InceptionResNetv2 y Xception, por lo que se asume que la cantidad de parámetros y operaciones son mayores a los que reportan en InceptionResNetv2 y Xception.

Como se puede ver en la tabla, los resultados muestran que las redes LightHist cuentan con un desempeño comparable con las redes publicadas. Sin embargo, una fuerte diferencia es que las redes propuestas tienen menos cantidad de parámetros respecto a las redes populares. De acuerdo con [85], [86], esta baja cantidad de parámetros permite que las redes LightHist ocupen menos memoria que el resto de las propuestas en la literatura. Por lo tanto, las redes propuestas pueden ser aplicadas en procesadores ASIC como los que se encuentran en teléfonos celulares o los que se encuentran en pequeñas computadoras como las Raspberry.

Los resultados de desempeño de la Tabla 5.6 indican que las redes LightHist tienen resultados similares a los presentados a la literatura. Sin embargo, como se mencionó anteriormente, en los artículos reportados en la literatura no se reporta algún análisis estadístico del aprendizaje. Por ello, en esta tesis se optó por entrenar las redes Xception y MobileNet bajo las mismas circunstancias que LightHist v2. Xception fue seleccionada obtuvo los resultados más altos en exactitud, mientras MobileNet-v2 los más bajos. Se realizó en entrenamiento con validación cruzada a 70 épocas usando validación cruzada con $h=5$, para evitar aumentar la

varianza y el tiempo de entrenamiento. Los resultados se muestran en Tabla 5.7 para Xception y en la Tabla 5.8 para MobileNet-v2.

Tabla 5.7 Resultados de Xception usando validación cruzada a 70 épocas.

<i>Carpeta</i>	<i>Exactitud</i>	<i>Precisión</i>	<i>Exahustividad</i>	<i>Puntuación-F1</i>
h1	91.4182	89.6254	84.0541	86.7503
h2	83.9205	59.8266	84.1463	69.9324
h3	79.7115	68.6782	67.3239	67.9943
h4	83.6790	86.2069	69.2841	76.8246
h5	85.0316	96.8300	68.4318	80.1909
Promedio	84.7522	80.2334	74.6480	76.3385

Tabla 5.8 Resultados de MobileNet-v2 usando validación cruzada a 70 épocas.

<i>Carpeta</i>	<i>Exactitud</i>	<i>Precisión</i>	<i>Exahustividad</i>	<i>Puntuación-F1</i>
h1	57.1816	96.5418	42.0326	58.5664
h2	69.9187	23.9884	54.2484	33.2665
h3	76.8260	95.6897	57.9130	72.1560
h4	67.0875	78.1609	48.4848	59.8460
h5	79.7115	85.5908	62.9237	72.5275
Promedio	63.5501	60.2651	48.1405	45.9165

Comparando los resultados del LightHist-v2 de la Tabla 5.5 con Xception de la Tabla 5.7 y la Tabla 5.8, se muestra que LightHist-v2 tiene resultados similares a Xception siendo aproximadamente 2% mayor en términos de exactitud, pero comparando otras métricas como precisión, exhaustividad y puntuación F1 va 3% al 5% mayor en estas métricas, esto nos muestra la calidad de la red ya que las clases no están balanceadas, y contra MobileNet-v2 de la Tabla 5.8, se muestra una gran diferencia que va del 21% mayor en exactitud hasta un 25% mayor en precisión lo que quiere decir que el incluir convoluciones en paralelo con diferentes tamaños de kernels resulto efectiva para la extracción de características que definan a las imágenes de histología de cáncer de mama, con ello, la generalización de la red. Aunque la exactitud puede decirnos de inmediato si la red se está entrenando correctamente y cómo

funciona en general. Sin embargo, no proporciona información detallada sobre la aplicación de este problema. En cuanto a la precisión ayuda cuando el costo de los falsos positivos es alto, si la precisión de la red que utilizamos es muy baja, se les dirá a muchos pacientes que tienen melanoma, lo que incluirá algunos diagnósticos erróneos. La exhaustividad ayuda cuando el costo de falsos negativos es alto, que en este caso dependería un diagnóstico médico. Una puntuación significa que tiene pocos falsos positivos y pocos falsos negativos, por lo que está identificando correctamente las amenazas reales y no le molestan las falsas alarmas. Esto nos indica que la red LightHist-v2 aprendió a clasificar de buena manera las imágenes de histología de cáncer de mama.

CAPÍTULO 6

CONCLUSIONES Y TRABAJO A FUTURO

La clasificación de imágenes de histología es una tarea difícil de realizar debido a los desafíos que se presentan en cada imagen como diferencia en la iluminación, la tinción, el enfoque del lente del microscopio, la cámara, etc. Estos desafíos hacen que cada imagen sea única y con características abstractas que se necesitan años de estudio para identificarlas correctamente, pero con ayuda de la tecnología, se ha logrado crear modelos de redes neuronales capaces de realizar esta tarea en pocos segundos. Por ello, en este trabajo se propusieron varias redes neuronales convolucionales que realicen una clasificación de imágenes histológicas de cáncer de mama que diferencia entre tumores malignos y benignos. Para desarrollar las redes propuestas, se realizó un análisis de redes ya existentes en la literatura, para conocer sus características, comportamiento y resultados de evaluación ante la problemática de clasificación binaria de imágenes de histología de cáncer de mama. Además, también se analizó la red MobileNet-v2, que es una red para aplicaciones móviles y fue la base de las redes a proponer. Con base en esto, se desarrollaron las redes LightHist, LightHist-v2 y LightHist-v3, que son modelos de CNN de bajo costo computacional con convoluciones separables en profundidad en paralelo. Para el entrenamiento y prueba, se utilizó la base de datos BreakHist que cuenta con 7909 imágenes de histología de tumores de mama separados en malignos y benignos y estos a su vez en subclases, a los que se le aplicó un tratamiento previo de normalización y tinción antes de ser entrenada la red. El desempeño de las redes LightHist fue similar a las presentadas en el estado del arte, pero LightHist-v2 fue la que presentó mejor desempeño logrando hasta un 99.73 de exactitud con los datos de pruebas. El estudio comparativo arrojó que las redes LightHist lograron un desempeño comparable con los propuestos en el estado del arte, pero con menor cantidad de parámetros. Esto genera una contribución en el estado del arte ya que mejora la red que puede ser utilizado en dispositivos móviles y procesadores ASCII. Este tipo de procesadores tienen la importancia de que la tecnología va encaminada a crear aparatos móviles capaces de realizar tareas que antes eran exclusivas para las computadoras y estos aparatos se conecten entre sí a través del Internet de

las cosas, así un patólogo pueda examinar múltiples muestras a la vez obteniendo un diagnóstico aproximado a través de su teléfono celular.

Por ello, se asume que la red LightHist-v2 se puede correr en GPU embebida o procesadores ASIC o bien incorporar las redes en aplicaciones móviles que pueda soportar un teléfono inteligente.

Como trabajo futuro, se propone la implementación de LightHist-v2 en una tarjeta embebida para ser aplicada como equipo médico o bien un teléfono inteligente. Además, se propone que la red de LightHist-v2 se entrene con diferentes tipos de imágenes histológicas que contengan información de tejidos que puedan utilizarse para predecir diferentes patologías. Por ello, se recomienda que LightHist-v2 se utilice como base para diferentes problemáticas planteadas en el proyecto de “Algoritmos De Visión Artificial Para Realizar Análisis De Imágenes Digitales Para Diagnóstico Clínico Parasitológico, Urológico, Hematológico y en Muestras Histológicas de Patología” que realiza el Instituto Tecnológico de Chihuahua en conjunto con la Facultad de Medicina y Ciencias Biomédicas de la Universidad Autónoma de Chihuahua, y del proyecto “Clasificación de Parásitos en Muestras de Copro Mediante Aprendizaje de Máquina, con clave 10071.21-P del Tecnológico Nacional de México. Para realizar el entrenamiento, se podría continuar la investigación creando un estándar de las imágenes en cuanto a su tamaño, normalización de tinción, color e intensidad con el fin de crear una red capaz de categorizar las imágenes que un laboratorio requeriría.

GLOSARIO

Acrónimo	Descripción
a_x	Aumento en el eje horizontal
a_y	Aumento en el eje vertical
ANN	Red Neuronal Artificial
<i>average-pooling</i>	Capa de agrupación de promediado
<i>brillo</i>	Valor de cambio de intensidad
CC_{std}	Costo computacional de la convolución estándar
CC_{sep}	Costo computacional de la convolución separable en profundidad
CIS	Carcinoma-in-situ
CNN	Red Neuronal Convolucional
$conv(ent,k)_{f,f}$	Capa de convolución
D_F	Dimensión de un mapa de características de entrada
D_K	Dimensión espacial del núcleo
DL	Aprendizaje profundo
<i>dim</i>	Dimensión de salida
DCNN	Red Neuronal Convolucional Profunda
$g(x)$	Salida de la neurona
<i>ent</i>	Imagen o matriz de entrada
ent_0	Intensidad de iluminación
f	Tamaño del filtro o kernel
FLOPS	Número de operaciones de punto flotante por segundo (del inglés floating point operations per second).
FN	Falsos Negativos
FP	Falsos Positivos
h	Número de carpetas
HE	Hematoxilina y eosina
HGG	Alto grado de glioma
k	Filtro o kernel
KNN	k vecinos más
L	Número de canales de salida
LSTM	Long-Short-Term-Memory
MANA	Análisis de núcleos adaptativos multiescala
<i>max-pooling</i>	Capa de agrupación de valor máximo
ML	Aprendizaje automático
MLP	Perceptrón multicapa
n_H	Tamaño en altura
n_{Hnuevo}	Nuevo tamaño de altura

n_{wi}	Tamaño en ancho
$n_{winuevo}$	Nuevo tamaño de ancho
n_C	Número de canales
OD	Matriz de valores de densidad óptica
p	Relleno añadido
PNG	Gráficos de Red Portátiles (siglas en inglés de <i>Portable Network Graphics</i>)
$pooling$	Capa de agrupación
q_{int}	Valor de intensidad de un pixel
Q_{norm}	Valor de intensidad normalizado de una imagen
q_{max}	Valor de intensidad máximo en una imagen
q_{min}	Valor de intensidad mínimo en una imagen
$ReLU$	Unidad Lineal Rectificada
RGB	Composición de color de una imagen en términos de intensidad en colores rojo, verde y azul (siglas del inglés <i>red, green, blue</i>)
$ResNet$	Red Neuronal Residual
RNN	Redes Neuronales Recurrentes
S	Distribución de probabilidad de cada clase resultante de la función Softmax
s	Tamaño del salto
Sh	Rango para aplicar el <i>shear_range</i> en una imagen
SVM	Maquina de vectores de soporte
SVD	Descomposición de Valores Singulares
t_j	Capa totalmente conectada
TN	Verdaderos Negativos
TP	Verdaderos Positivos
v	Matriz de entrada
w^T	Matriz de pesos
x	Vector de coordenadas horizontales
x_{max}	Valor máximo en el vector horizontal
x_{disp}	Valor de desplazamiento horizontal
y	Vector de coordenadas verticales
y_{min}	Valor máximo de las coordenadas verticales
y_{disp}	Valor de desplazamiento vertical
θ	Angulo de rotación

REFERENCIAS

- [1] M. H. Ross and P. Wojciech, *Histología Texto y Atlas*, 7th ed. Barcelona, España: Wolters Kluwer, 2015.
- [2] S. Chin Neoh *et al.*, “An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images,” *Sci. Rep.*, vol. 5, pp. 1–14, 2015.
- [3] H. Mohan, *Patología*, 6th ed. Madrid, España: Panamericana, 2010.
- [4] A. Stevens, J. S. Lowe, and B. Young, *Wheater. Histopatología Básica*. 2003.
- [5] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. New York, NY: Cambridge University Press, 2014.
- [6] M. Z. Alom *et al.*, “A state-of-the-art survey on deep learning theory and architectures,” *Electron.*, vol. 8, no. 3, pp. 1–67, 2019.
- [7] S. Theodoridis, K. Nos, and K. M. Bas, *Pattern Recognition*, 2nd ed. San Diego, CA: Elsevier, 2003.
- [8] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, 1969.
- [9] C. C. Aggarwal, *Neural Networks and Deep Learning*. Yorktown Heights, NY: Springer International Publishing AG, 2018.
- [10] J. Hurwitz and D. Kirsch, *Machine Learning for dummies*. Hoboken, NJ: John Wiley & Sons, Inc., 2018.
- [11] L. Deng, *Deep Learning Methods and Applications*. Redmond, WA: Foundations and Trends in Signal Processing, 2014.
- [12] A. Al Nahid and Y. Kong, “Involvement of Machine Learning for Breast Cancer Image Classification: A Survey,” *Comput. Math. Methods Med.*, vol. 2017, no. i, 2017.
- [13] K. Fukushima, “Neocognitron: A self-organizing neural network model for a

- mechanism of pattern recognition unaffected by shift in position,” *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.
- [14] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
 - [15] C. C. J. Kuo, “Understanding convolutional neural networks with a mathematical model,” *J. Vis. Commun. Image Represent.*, vol. 41, no. November, pp. 406–413, 2016.
 - [16] T. Wen and A. Edelman, “Support Vector Machines: Algorithms and Applications,” pp. 249–257, 2003.
 - [17] A. P. Qu *et al.*, “Segmentation of Hematoxylin-Eosin stained breast cancer histopathological images based on pixel-wise SVM classifier,” *Sci. China Inf. Sci.*, vol. 58, no. 9, pp. 1–13, 2015.
 - [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.
 - [19] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, “MuDeRN: Multi-category classification of breast histopathological image using deep residual networks,” *Artif. Intell. Med.*, vol. 88, pp. 14–24, 2018.
 - [20] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, “Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module,” *PLoS One*, vol. 14, no. 3, pp. 1–21, 2019.
 - [21] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A Dataset for Breast Cancer Histopathological Image Classification,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2016.
 - [22] C. Szegedy *et al.*, “Going deeper with convolutions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015.
 - [23] J. Ker, Y. Bai, H. Y. Lee, J. Rao, and L. Wang, “Automated brain histology

- classification using machine learning,” *J. Clin. Neurosci.*, no. xxxx, 2019.
- [24] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, “Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network,” *J. Digit. Imaging*, 2019.
 - [25] R. Yan *et al.*, “Breast cancer histopathological image classification using a hybrid deep neural network,” *Methods*, no. February, pp. 1–9, 2019.
 - [26] Z. Moshavash, H. Danyali, and M. S. Helfroush, “An Automatic and Robust Decision Support System for Accurate Acute Leukemia Diagnosis from Blood Microscopic Images,” *J. Digit. Imaging*, vol. 31, no. 5, pp. 702–717, 2018.
 - [27] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, “Automatic cell nuclei segmentation and classification of breast cancer histopathology images,” *Signal Processing*, vol. 122, pp. 1–13, 2016.
 - [28] E. Abdulhay, M. A. Mohammed, D. A. Ibrahim, N. Arunkumar, and V. Venkatraman, “Computer Aided Solution for Automatic Segmenting and Measurements of Blood Leucocytes Using Static Microscope Images,” *J. Med. Syst.*, vol. 42, no. 4, 2018.
 - [29] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, “Transfer learning based histopathologic image classification for breast cancer detection,” *Heal. Inf. Sci. Syst.*, vol. 6, no. 1, 2018.
 - [30] S. Reis *et al.*, “Automated Classification of Breast Cancer Stroma Maturity from Histological Images,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2344–2352, 2017.
 - [31] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, and J. Huang, “Weakly supervised mitosis detection in breast histopathology images using concentric loss,” *Med. Image Anal.*, vol. 53, pp. 165–178, 2019.
 - [32] G. Liang, H. Hong, W. Xie, and L. Zheng, “Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification,” *IEEE Access*,

- vol. 6, no. c, pp. 36188–36197, 2018.
- [33] A. Al Nahid, M. A. Mehrabi, and Y. Kong, “Histopathological breast cancer image classification by deep neural network techniques guided by local clustering,” *Biomed Res. Int.*, vol. 2018, 2018.
 - [34] D. M. Vo, N. Q. Nguyen, and S. W. Lee, “Classification of breast cancer histology images using incremental boosting convolution networks,” *Inf. Sci. (Ny)*, vol. 482, pp. 123–138, 2019.
 - [35] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, “EMS-Net: Ensemble of Multiscale Convolutional Neural Networks for Classification of Breast Cancer Histology Images,” *Neurocomputing*, vol. 366, pp. 46–53, 2019.
 - [36] M. Khoshdeli and B. Parvin, “Feature-Based Representation Improves Color Decomposition and Nuclear Detection Using a Convolutional Neural Network,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 625–634, 2018.
 - [37] B. S. Tchinda, M. Noubom, D. Tchiotsop, V. Louis-Dorr, and D. Wolf, “Towards an automated medical diagnosis system for intestinal parasitosis,” *Informatics Med. Unlocked*, vol. 13, no. August, pp. 101–111, 2018.
 - [38] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, “RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images,” *IEEE Access*, vol. 7, pp. 21420–21428, 2019.
 - [39] D. Bibin, M. S. Nair, and P. Punitha, “Malaria Parasite Detection from Peripheral Blood Smear Images Using Deep Belief Networks,” *IEEE Access*, vol. 5, no. c, pp. 9099–9108, 2017.
 - [40] M. Salvi and F. Molinari, “Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images,” *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–13, 2018.
 - [41] Y. Li, J. Wu, and Q. Wu, “Classification of Breast Cancer Histology Images Using

- Multi-Size and Discriminative Patches Based on Deep Learning,” *IEEE Access*, vol. 7, pp. 21400–21408, 2019.
- [42] T. H. Song, V. Sanchez, H. Eidaly, and N. M. Rajpoot, “Simultaneous Cell Detection and Classification in Bone Marrow Histology Images,” *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 4, pp. 1469–1476, 2019.
- [43] M. A. Anupama, V. Sowmya, and K. P. Soman, “Breast cancer classification using capsule network with preprocessed histology images,” *Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019*, pp. 143–147, 2019.
- [44] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [45] Ö. İnik, A. Ceyhan, E. Balcıoğlu, and E. Ülker, “A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network,” *Comput. Biol. Med.*, vol. 112, no. July, p. 103350, 2019.
- [46] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, “Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.
- [47] D. Bardou, K. Zhang, and S. M. Ahmad, “Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks,” *IEEE Access*, vol. 6, pp. 24680–24693, 2018.
- [48] T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-on, “Convolutional Neural Networks for Recognition of Lymphoblast Cell Images,” *Comput. Intell. Neurosci.*, vol. 2019, p. 12, 2019.
- [49] T. Qaiser *et al.*, “Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features,” *Med. Image Anal.*, vol. 55, pp. 1–14, 2019.

- [50] S. Rajaraman *et al.*, “Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images,” *PeerJ*, vol. 6, p. e4568, 2018.
- [51] Y. Gu and J. Yang, “Multi-level magnification correlation hashing for scalable histopathological image retrieval,” *Neurocomputing*, vol. 351, pp. 134–145, 2019.
- [52] J. Xie, R. Liu, J. Luttrell, and C. Zhang, “Deep learning based analysis of histopathological images of breast cancer,” *Front. Genet.*, vol. 10, no. FEB, pp. 1–19, 2019.
- [53] A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, “Recognition of peripheral blood cell images using convolutional neural networks,” *Comput. Methods Programs Biomed.*, vol. 180, p. 105020, 2019.
- [54] S. U. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, “A novel deep learning based framework for the detection and classification of breast cancer using transfer learning,” *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, 2019.
- [55] A. Anghel *et al.*, “A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology,” *Front. Med.*, vol. 6, no. September, pp. 1–13, 2019.
- [56] T. Noergaard, *Embedded Systems Architecture: A Comprehensive Guide for Engineers and Programmers*. 2013.
- [57] TIOBE, “TIOBE Index for November 2020,” 2020. [Online]. Available: <https://www.tiobe.com/tiobe-index/>.
- [58] S. Hosseinzadeh Kassani and P. Hosseinzadeh Kassani, “A comparative study of deep learning architectures on melanoma detection,” *Tissue Cell*, vol. 58, no. March, pp. 76–83, 2019.
- [59] L. Wall *et al.*, “Machine Learning with Python,” *Tutor. points Pvt Ltd*, p. 68, 2015.
- [60] D. Berrar, “Cross-validation,” *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*,

- vol. 1–3, pp. 542–545, 2018.
- [61] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64270–64277, 2018.
 - [62] R. Karim, “Illustrated: 10 CNN Architectures,” *Towards data science*, 2019. [Online]. Available: <https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#e4b1>. [Accessed: 11-Jun-2020].
 - [63] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, 2015.
 - [64] A. Garcia-Perez, F. Gheriss, D. Bedford, A. Garcia-Perez, F. Gheriss, and D. Bedford, “Measurement, Reliability, and Validity,” *Des. Track. Knowl. Manag. Metrics*, pp. 163–182, 2019.
 - [65] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016.
 - [66] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 4278–4284, 2017.
 - [67] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017.
 - [68] K. Bai, “A Comprehensive Introduction to Different Types of Convolutions in Deep Learning,” *[en línea]*, 2019. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215>. [Accessed: 18-Dec-2020].
 - [69] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann

- Machines,” in *ICML ’10: Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, p. Pages 807–814.
- [70] A. Krizhevsky and G. Hinton, “Convolutional deep belief networks on cifar-10,” *Unpubl. Manuscr.*, pp. 1–9, 2010.
 - [71] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018.
 - [72] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800–1807, 2017.
 - [73] L. Fei-Fei, J. Deng, and K. Li, “ImageNet: Constructing a large-scale image database,” *J. Vis.*, vol. 9, no. 8, pp. 1037–1037, 2010.
 - [74] M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” *Proc. - 2009 IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2009*, pp. 1107–1110, 2009.
 - [75] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001.
 - [76] A. Vahadane *et al.*, “Structure-preserved color normalization for histological images,” *Proc. - Int. Symp. Biomed. Imaging*, vol. 2015-July, pp. 1012–1015, 2015.
 - [77] S. H. Kassani, M. J. Wesolowski, and K. A. Schneider, “Classification of Histopathological Biopsy Images Using Ensemble of Deep Classification of Histopathological Biopsy Images Using Ensemble of Deep Learning Networks,” no. September, 2019.
 - [78] G. Aresta *et al.*, “BACH: Grand challenge on breast cancer histology images,” *Med. Image Anal.*, vol. 56, pp. 122–139, 2019.
 - [79] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal, “Automated analysis and

- classification of melanocytic tumor on skin whole slide images,” *Comput. Med. Imaging Graph.*, vol. 66, no. June 2017, pp. 124–134, 2018.
- [80] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, “Patch-based system for Classification of Breast Histology images using deep learning,” *Comput. Med. Imaging Graph.*, vol. 71, pp. 90–103, 2019.
- [81] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, “AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.
- [82] Q. Wang, Y. Shi, and Di. Shen, *Machine Learning in Medical Imaging*, 8th ed., vol. 1. Quebec, Canada: Springer International Publishing, 2017.
- [83] H. Li, S. Zhuang, D. ao Li, J. Zhao, and Y. Ma, “Benign and malignant classification of mammogram images based on deep learning,” *Biomed. Signal Process. Control*, vol. 51, pp. 347–354, 2019.
- [84] TensorFlow, “tf.keras.preprocessing.image.ImageDataGenerator,” [en línea], 2021. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator. [Accessed: 03-Jan-2021].
- [85] D. Moolchandani, A. Kumar, and S. R. Sarangi, “Accelerating CNN Inference on ASICs: A Survey,” *J. Syst. Archit.*, vol. 113, no. January, p. 101887, 2021.
- [86] C.-C. Wang, C.-T. Chiu, and J.-Y. Chang, “EfficientNet-eLite: Extremely Lightweight and Efficient CNN Models for Edge Devices by Network Candidate Search,” 2020.