

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio Ciencias e Ingenierías

**Clasificación de lesiones en mamografías mediante una
red neuronal convolucional**

Proyecto de investigación

Samik Real Enríquez

Ingeniería en Sistemas

Trabajo de titulación presentado como requisito
para la obtención del título de
Ingeniería en Sistemas

Quito, 22 de abril de 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE TITULACIÓN

Clasificación de lesiones en mamografías mediante red neuronal convolucional

Samik Real Enríquez

Calificación:

Nombre del profesor, Título académico:

Noel Pérez, Ph.D.

Firma del profesor:

Quito, 22 de abril de 2019

Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombres y apellidos: Samik Real Enríquez

Código: 00117771

Cédula de Identidad: 1717479131

Lugar y fecha: Quito, 22 de abril de 2019

Resumen

El cáncer de mama es el segundo causante de muerte por cáncer a nivel mundial. Aproximadamente 411 mil mujeres mueren por este cáncer al año (Kamanger, 2006). Se ha demostrado que la detección temprana de este cáncer incrementa la probabilidad de supervivencia del paciente de manera drástica. Es por esto que es necesario mejorar las técnicas de lectura de mamografías para mejorar el diagnóstico de los pacientes. Los modelos tradicionales y modelos CAD son altamente dependientes de su implementación y de las características proporcionadas de manera manual. Por esta razón, el uso de redes convolucionales representa una ventaja enorme, ya que las características son extraídas de manera automática por el modelo. En este trabajo se implementó exitosamente una red neuronal convolucional para la clasificación de lesiones de mamas en mamografías, capaz de establecer un compromiso entre confiabilidad y efectividad. Este modelo fue entrenado con una base de datos de 3946 mamografías, conteniendo microcalcificaciones y masas, obteniendo un valor aproximado de 68% en la métrica de exactitud.

Palabras claves: cáncer de mama, clasificación, Deep Learning, CNN, mamografías,

Abstract:

Breast cancer is one of the 2 most deadly cancers for women worldwide. This type of cancer is attributed to killing 411 000 people each year. Research has shown that early detection can improve survival rate in patients with breast cancer (Kamanger, 2006), which is why it is important to improve mammography Reading techniques in order to accurately diagnose patients. Traditional machine learning models and CAD models are highly dependent on their implementation and the features that are manually extracted. This is why convolutional neural networks are important, because they can extract features automatically and scale much better due to the decrease in number of parameters used. A convolutional neural network was implemented in this work for the classification of mammography lesions. This model was trained with 3946 mammographies containing microcalcifications and masses. This model obtained an accuracy of approximately 68%.

Key words: breast cancer, classification, Deep Learning, CNN, mammographies

TABLA DE CONTENIDO

<i>CAPÍTULO 1: INTRODUCCIÓN</i>	9
1.1. Antecedentes	9
1.1.1. Cáncer de mama.....	9
1.1.3. Masas.....	10
1.1.4. Microcalcificaciones.....	11
1.1.5. Detección temprana del cáncer de mama.....	13
1.1.6. CAD para el cáncer de mama.....	13
1.1.7. Redes neuronales artificiales.....	15
1.1.8. Hiperparámetros.....	18
1.1.9. Deep Learning.....	18
1.2. Motivación y objetivos	19
1.3. Tareas de investigación	20
<i>CAPÍTULO 2: ESTADO DEL ARTE</i>	21
2.1. Estado del arte de modelos <i>Deep Learning</i>	21
2.2. Trabajos relacionados	22
<i>CAPÍTULO 3: MÉTODO PROPUESTO</i>	27
3.1 Modelos <i>Deep Learning</i> seleccionados	27
3.2 Metodología Experimental	28
3.2.1 Creación del conjunto de datos.....	28
3.2.2 Preprocesamiento de las imágenes.....	29
3.2.3 Validación de los resultados.....	30
<i>CAPÍTULO 4: CONCLUSIONES</i>	37
4.1 Conclusiones	37
4.2 Trabajo futuro	37
<i>CAPÍTULO 5: REFERENCIAS BIBLIOGRÁFICAS</i>	38

TABLA DE FIGURAS

Figura 1. Mamografías de la vista medio lateral oblicua de la mama derecha con masas benignas, bien definidas y con baja densidad. Imagen extraída de la base de datos BCDR. A) Paciente #134 con 23 años, B) Magnificación del paciente #134, C) Paciente #429 con 35 años, D) Magnificación del paciente #429 (Pérez, 2015)	11
Figura 2. Mamografías de la vista medio lateral oblicua de la mama derecha con masas malignas y de forma irregular. Imagen extraída de la base de datos BCDR. A) Paciente #105 con 51 años, B) Magnificación del paciente #105, C) Paciente #143 con 67 años, D) Magnificación del paciente #143 (Pérez, 2015).....	11
Figura 3. Mamografías de la vista medio lateral oblicua de la mama izquierda con agrupaciones de microcalcificaciones benignas y de tamaño homogéneo. Imagen extraída de la base de datos BCDR. A) Paciente #32 con 49 años, B) Magnificación del paciente #32, C) Paciente #293 con 48 años, D) Magnificación del paciente #293 (Pérez, 2015).....	12
Figura 4. Mamografías de la vista medio lateral oblicua de la mama izquierda con agrupaciones de microcalcificaciones malignas y de morfología poco simétrica. Imagen extraída de la base de datos BCDR. A) Paciente #457 con 58 años, B) Magnificación del paciente #457, C) Paciente #488 con 73 años, D) Magnificación del paciente #488 (Pérez, 2015)	13
Figura 5: Ejemplos de arte generado artificialmente (Elgammal, 2017).....	23
Figura 6. Ejemplo de mamografías extraída de la base de datos DDSM: las dos primeras (de izquierda a derecha) son vistas craneocaudales, las dos últimas son vistas medio lateral oblicuo.	29
Figura 7. Ejemplo de localización de lesiones en mamografías usando el programa en MATLAB (izquierda) y recorte del área de interés de la lesión (derecha).....	30
Figura 8. Gráfica de pérdida del modelo 1	31
Figura 9. Gráfica de exactitud del modelo 1	32
Figura 10. Gráfica de pérdida del modelo 2	33
Figura 11. Gráfica de exactitud del modelo 2.....	33
Figura 12. Gráfica de pérdida del modelo 3	34
Figura 13. Gráfica de la métrica de exactitud del modelo 3	35

GLOSARIO

ANN: Red neuronal artificial

Backpropagation: Retropropagación del error

BCDR: Repositorio digital de cáncer de mama

CAD: Detección/Diagnóstico asistido por computadora

CADe: Detección asistida por computadora

CADx: Diagnóstico asistido por computadora

CC: Craniocaudal

Convolutional Neural Network: (CNN)Red neuronal convolucional

DDSM: Digital Database for Screening Mammography

Decay: Tasa de decaimiento

Dropout: Tasa de desconexión

FC: Capa de neuronas completamente conectadas

Learning Rate: Tasa de aprendizaje

Machine Learning: aprendizaje automático

MC: Microcalcificación

MLO: Medio lateral oblicuo

Momentum: Momento lineal

ROI: Región de interés

Stochastic Gradient Descent: Gradiente estocástico de descenso

CAPÍTULO 1: INTRODUCCIÓN

El cáncer de mama es el segundo causante de muerte por cáncer de mujeres en el mundo. Las estadísticas muestran que 1,152,161 nuevos casos de cáncer de mama son encontrados a nivel mundial; y con 411,093 muertes (Kamanger, 2006). Se ha demostrado que el diagnóstico temprano del cáncer de mama incrementa la probabilidad de una recuperación completa y disminuye la mortalidad de los pacientes padeciendo este cáncer (Pérez, 2014).

1.1. Antecedentes

1.1.1. Cáncer de mama

El cáncer es la mutación de los genes encargados de la replicación celular y de la regulación del crecimiento de las células. Estos genes se encuentran en el núcleo de las células y actúan como un control para activar o desactivar diferentes células y que células viejas se mueran mientras que las nuevas tomen control. Cuando una mutación ocurre, estas células no se mueren y empiezan a dividirse sin control, creando tumores (Pérez, 2015).

Un tumor es benigno cuando las células son parecidas a las normales y crecen de manera lenta, sin invadir tejido cercano o esparcirse a otros lugares en el cuerpo. Los tumores malignos son cancerígenos y crecen con bastante velocidad, poniendo en riesgo la vida del paciente (Pérez, 2015).

1.1.2. Mamografías

Las mamografías son imágenes creadas mediante el uso de una baja dosis de rayos-x en el área de las mamas. La lectura de la mamografía es el método más efectivo de detección de cáncer de mama, ya que se pueden detectar las lesiones antes de que estas sean palpables clínicamente. La ventaja de las mamografías, en relación con otros tipos de imágenes de mamas, es que genera imágenes de alta calidad y requiere de una baja exposición a radiación. Por esta razón, las mamografías son el único método de imágenes para la revisión de rutina de cáncer de mama (NG KH, 2003).

1.1.3. Masas

Una masa está definida como una lesión que ocupe un espacio en las tres dimensiones espaciales. Se puede asegurar que ocupa un espacio en las tres dimensiones cuando se observa la lesión en la vista craneocaudal y medio lateral oblicuo. De lo contrario, esto se denomina una asimetría. Las masas benignas generalmente tienen una baja densidad con márgenes bien definidos y una cobertura de grasa sobre la lesión; mientras que las masas malignas tienen generalmente una forma poco irregular, sin simetría y no tienen la cobertura de grasa (Paredes, 2009).

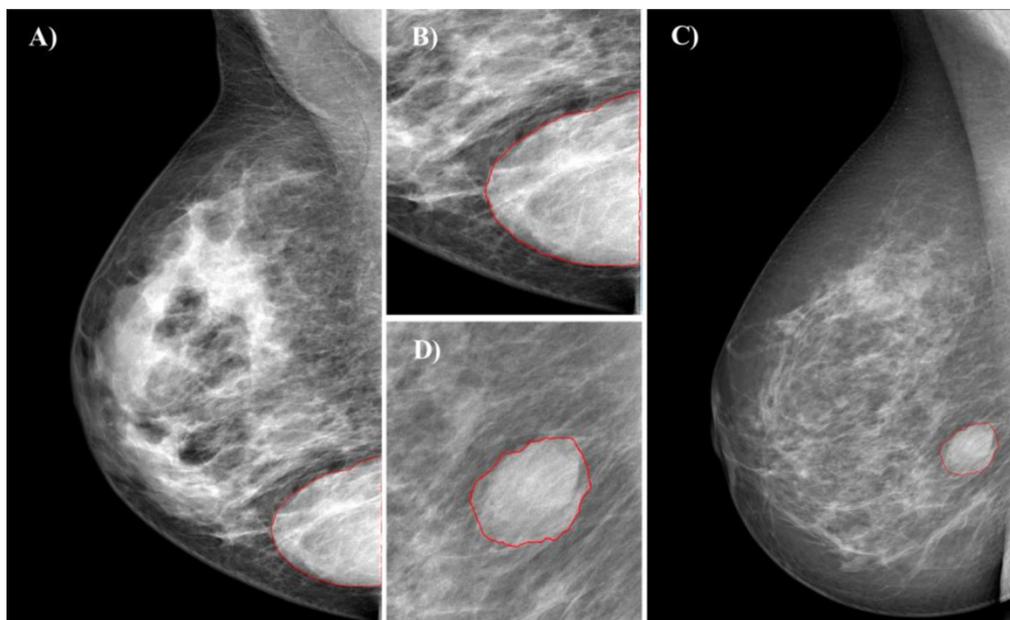


Figura 1. Mamografías de la vista medio lateral oblicua de la mama derecha con masas benignas, bien definidas y con baja densidad. Imagen extraída de la base de datos BCDR. A) Paciente #134 con 23 años, B) Magnificación del paciente #134, C) Paciente #429 con 35 años, D) Magnificación del paciente #429 (Pérez, 2015)

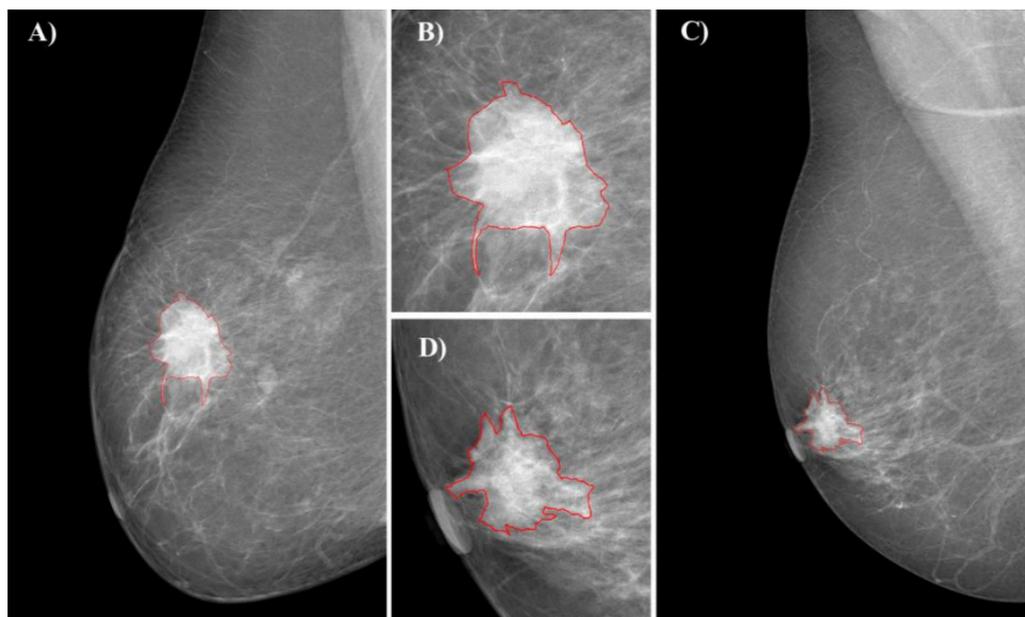


Figura 2. Mamografías de la vista medio lateral oblicua de la mama derecha con masas malignas y de forma irregular. Imagen extraída de la base de datos BCDR. A) Paciente #105 con 51 años, B) Magnificación del paciente #105, C) Paciente #143 con 67 años, D) Magnificación del paciente #143 (Pérez, 2015)

1.1.4. Microcalcificaciones

Las microcalcificaciones son acumulaciones granulares de calcio con densidad mucho mayor al tejido normal. Las microcalcificaciones malignas suelen ser pequeñas, agrupadas entre varias microcalcificaciones de diferentes tamaños y con bordes irregulares. Las microcalcificaciones benignas suelen ser grandes, homogéneas en tamaño y sus bordes son redondos y regulares. Otra característica importante de distinción entre microcalcificaciones benignas y malignas es la morfología de las agrupaciones. Una morfología simétrica suele ser de microcalcificaciones benignas, mientras que una morfología poco simétrica suele ser de microcalcificaciones malignas (Suri, 2016).

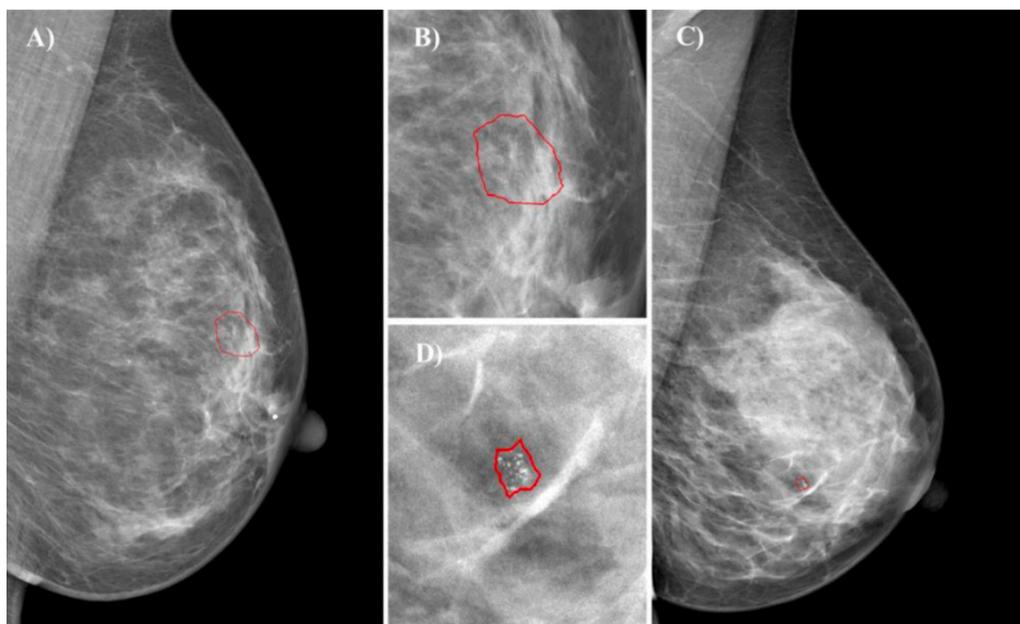


Figura 3. Mamografías de la vista medio lateral oblicua de la mama izquierda con agrupaciones de microcalcificaciones benignas y de tamaño homogéneo. Imagen extraída de la base de datos BCDR. A) Paciente #32 con 49 años, B) Magnificación del paciente #32, C) Paciente #293 con 48 años, D) Magnificación del paciente #293 (Pérez, 2015)

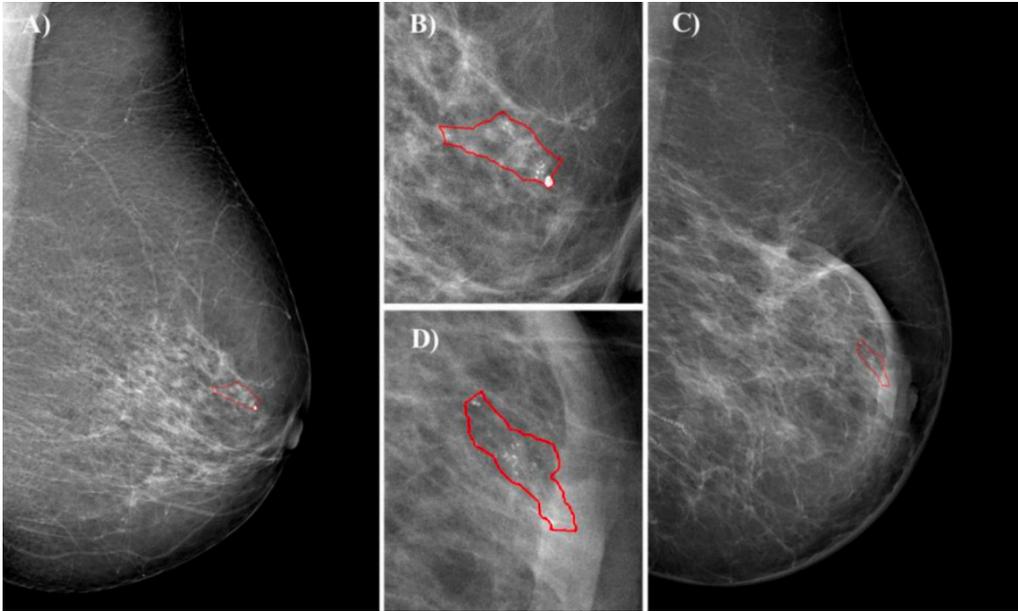


Figura 4. Mamografías de la vista medio lateral oblicua de la mama izquierda con agrupaciones de microcalcificaciones malignas y de morfología poco simétrica. Imagen extraída de la base de datos BCDR. A) Paciente #457 con 58 años, B) Magnificación del paciente #457, C) Paciente #488 con 73 años, D) Magnificación del paciente #488 (Pérez, 2015)

1.1.5. Detección temprana del cáncer de mama

La detección temprana del cáncer de mama es fundamental para incrementar la supervivencia del paciente. Existen dos tipos fundamentales:

- 1) Diagnóstico temprano, conocimiento de los síntomas e historial médico familiar en la población sintomática para facilitar el tratamiento adecuado. Este tipo de detección temprana es importante en especial para países en vías de desarrollo donde no se cuenta con muchos recursos económicos (Suri, 2016).
- 2) Controles de rutina de manera sistemática para detectar el cáncer de mama en pacientes asintomáticos (Suri, 2016).

1.1.6. CAD para el cáncer de mama

El término *CAD* para el cáncer de mama se puede asociar a dos métodos: detección asistida por computadora (*CADe*) o diagnóstico asistido por computadora (*CADx*). El primero se enfoca en la detección de regiones sospechosas en imágenes médicas, mientras que el segundo se encarga de diagnosticar la región sospechosa. Los sistemas *CAD* tienen generalmente las siguientes partes (Pérez, 2015):

1. Selección de la región de interés (*ROI*): el área de la lesión es seleccionada para la segmentación de manera manual, semiautomática o completamente automática.
2. Preprocesamiento de imagen: la imagen es mejorada, eliminando el ruido y aumentando el contraste.
3. Segmentación: la lesión es segmentada del área de interés, seleccionado en el primer proceso, mediante la identificación del contorno o por región de píxeles. Esta segmentación puede ser manual, semiautomática o completamente automática.
4. Extracción y selección de características: características medibles son extraídas de la región segmentada y agrupadas en un vector de características. Ejemplos de estas características son: perímetro de la lesión, tamaño, área, elongación, entre otros. El proceso de selección de características es realizado mediante técnicas de filtro, *wrapper*, o híbridos.
5. Clasificación automática: este último paso es esencial para sistemas *CADx*, ya que determina la clase de la lesión mediante las características proporcionadas. Las clases generalmente corresponden al tipo de lesión o condición maligna o benigna de la misma.

Se ha mostrado que ambos métodos *CAD* ayudan a los especialistas a mejorar su diagnóstico en pacientes (Eadie, 2012) y a disminuir los falsos negativos y falsos positivos en los diagnósticos (Araujo, 2017).

1.1.7. Redes neuronales artificiales

La primera conceptualización de *machine learning* fue en 1958, con el aporte de Frank Rosenblatt, quien implementó el Perceptrón electrónico (Rosenblatt, 1958). Este Perceptrón electrónico demostró poder inferir y aprender mediante el asociacionismo (Wang, 2017). El trabajo de Paul Werbos en 1974 dio origen a la técnica de *backpropagation*, el cual es usado en la mayoría de los modelos de aprendizaje hasta hoy en día (Wang, 2017). A partir de estos dos aportes, se realizaron varios avances en técnicas y algoritmos aplicadas para la estadística y para *machine learning*.

La importancia de esta área es que, mediante diferentes algoritmos, se puede aprender rasgos y características de forma generalizada y eficiente para determinar la clase de la instancia en cuestión, a diferencia de la programación tradicional que necesita instrucciones explícitas. Por esta razón, el área de *machine learning* ha ganado mucho interés en diferentes ramas, no solo de la ciencia en computación (Domingos, 2012).

Hay muchos tipos de modelos y técnicas dentro del área de *machine learning*, pero los más estudiado históricamente son los modelos de clasificación, a los cuales se les introducen datos (imágenes, texto, números, etc.) y el modelo devuelve la clase a la que pertenece (Domingos, 2012).

Existen tres componentes principales de un modelo de aprendizaje: el algoritmo de clasificación, la métrica de evaluación y el optimizador (Domingos, 2012). El clasificador define

la forma en la cual va a aprender el modelo. Ejemplos de clasificadores son: *k-nearest neighbour*, *Naive Bayes*, redes neuronales artificiales, entre otros. La métrica de evaluación define un parámetro que será evaluado para determinar la eficiencia del modelo. Algunas métricas de evaluación son: precisión, porcentaje de error, error cuadrado, etc. El optimizador se encarga de establecer la forma en la cual el modelo buscará candidatos óptimos. Ejemplos de optimizadores son: *Greedy search*, *stochastic gradient descent* y optimizadores adaptivos como Adam y Nadam, los cuales modifican el *learning rate* de manera automática.

Para entrenar un modelo, es necesario proporcionarlo con data, separándola en tres subconjuntos de datos disjuntos: entrenamiento, validación y prueba. El modelo se entrena con el subconjunto de entrenamiento, modificando los pesos y evaluando el desempeño mediante el optimizador según la métrica de evaluación seleccionada sobre este conjunto. El entrenamiento y la evaluación completa del modelo sobre el conjunto de datos del entrenamiento y de validación, respectivamente, se denominada una época. Los pesos del modelo son ajustados al culminar cada época. Los pesos son parámetros que se utilizan para clasificar las instancias de los conjuntos de datos en las clases correspondientes. De esta manera se obtiene el resultado del modelo, el cual está en función de la métrica seleccionada, por ejemplo, la métrica de exactitud (*accuracy*). El conjunto de prueba es usado para probar nuevamente la capacidad de generalización del modelo. Por esta razón es muy importante que los conjuntos de entrenamiento, validación y prueba sean disyuntos.

Si se obtienen una buena clasificación con el conjunto de datos del entrenamiento, pero una mala clasificación con el conjunto de prueba, lo más probable es que exista *overfitting*. Este fenómeno ocurre cuando el modelo predice las clases del conjunto de entrenamiento, pero no es

capaz de clasificar adecuadamente instancias nuevas del conjunto de prueba, mostrando poca capacidad de generalización (LeCun, 2015).

Existen tres categorías principales de entrenamiento dentro del área de *machine learning*: supervisado, no supervisado e híbrido. Entrenamiento supervisado recibe como entrada instancias de datos de las cuales se conoce la clase de salida, mientras que el entrenamiento no supervisado recibe instancias sin su clase de salida. En ambos casos el modelo aprende a clasificar las instancias adecuadamente. En el caso de entrenamiento híbrido, se usa una mezcla de supervisado y no supervisado para entrenar al modelo (LeCun, 2015).

Un modelo de entrenamiento supervisado implica que las instancias de datos tienen una etiqueta adjunta con su clase, la cual es empleada para calcular el error (distancia entre la clase real y la clase predicha por el modelo en el entrenamiento) mediante la función objetivo y optimizador. Estas redes están compuestas por neuronas, las cuales conforman diferentes capas. Las neuronas son entidades que transforman valores recibidos (*input*) a valores de salida (*output*) mediante una función de activación. Existen diferentes funciones de activación, pero la más usada tradicionalmente es la sigmoide, la cual tiene forma de “S” con asíntotas horizontales en 0 y 1, y sirve para transformar valores entre un rango de 0 y 1.

Estas neuronas, dependiendo de su ubicación en las capas, pueden ser neuronas de entrada, ocultas o de salida, como se ve en la figura 2. Las neuronas de entrada no tienen función de activación ya que su única función es pasar valores a las otras neuronas. Los datos se propagan desde las neuronas de entrada hasta las neuronas de salida, activándolas cuando sobrepase un determinado límite impuesto por el modelo mientras se va entrenando.

1.1.8. Hiperparámetros

Los hiperparámetros son valores determinados antes del proceso de entrenamiento. Estos valores van a determinar la velocidad y capacidad del modelo de aprender. Existen diferentes tipos de hiperparámetros de las redes neuronales artificiales, los más comunes son (LeCun, 2015), (Ioffe, 2015):

- *Learning rate*: magnitud con la cual se van a actualizar los pesos en cada época
- *Momentum*: factor de momento lineal o de inercia
- *Nesterov Momentum*: *momentum* con factores de corrección para mejorar rendimiento.
- *Decay*: factor con el cual va a disminuir el *learning rate* en cada época.
- *Dropout*: probabilidad de romper conexión entre las neuronas de dos capas. Sirve para impedir la memorización del conjunto de entrenamiento del modelo (*overfitting*).
- *Batch normalization*: Normalización de distribución de valores de salida de las neuronas de forma gaussiana. Esta técnica sirve para acelerar el aprendizaje del modelo e impedir *overfitting*.

1.1.9. Deep Learning

Deep Learning es el entrenamiento de una red neuronal artificial con un número mayor de capas intermedias comparado con modelos de redes neuronales artificiales tradicionales. El impedimento principal que impedía incrementar el número de capas intermedias en redes neuronales tradicionales era la función de activación Sigmoid, ya que esta función de activación restringe los valores de salida entre 0 y 1. Esto impide el aprendizaje de las capas más profundas.

La solución a este problema fue la creación de nuevas funciones de activación, como el *Rectifier Linear Unit (ReLU)*, la cual cumple la función matemática de $f(x) = \max(0, x)$ (Glorot, 2011).

Al incrementar el número de capas de una red neuronal artificial, la capacidad de abstracción también aumenta, convirtiéndoles en buenos candidatos para resolver tareas complejas y muchas veces imposibles con modelos anteriores.

Existen diferentes tipos de redes de *Deep Learning* por ejemplo, una red neuronal artificial con una profundidad de más de 3 capas ocultas ya es considerado un modelo de *Deep Learning*. Otro modelo de *Deep Learning* es el *Recurrent neural network*, la cual está compuesta por conexión entre neuronas de capas delanteras y traseras, facilitando el uso de evento e información pasada. Las dos redes neuronales más usadas en la clasificación de imágenes son *Convolutional Neural Network* y *Adversarial Neural Network*. La red convolucional funciona mediante filtros que itera por la imagen, representada por una matriz. Estos filtros son parámetros ajustables por la red en el aprendizaje. La red Adversarial constituye de dos redes Convolucionales compitiendo una contra la otra, la una para crear algo artificial irreconocible y la otra para detectar creaciones artificiales. Al competir estas dos redes, ambas van mejorando progresivamente (LeCun, 2015).

1.2. Motivación y objetivos

La detección temprana del cáncer de mama es fundamental para incrementar la supervivencia del paciente. Con este fin, se han desarrollado varias técnicas para incrementar la detección y clasificación adecuada de lesiones en imágenes médicas (como la doble lectura, donde dos especialistas leen la misma imagen o los métodos *CAD*). El principal inconveniente del método de doble lectura es que incrementa el costo por agotamiento físico y el trabajo de los especialistas encargados de leer las mamografías.

En cambio, los métodos *CAD* son altamente dependientes del buen preprocesamiento y la segmentación, además de la extracción de características y clasificación. Este método tiene su principal limitante relacionado con las características extraídas (conocidas de antemano) en la implementación.

Los métodos de doble lectura, al igual que los métodos *CAD*, tienen fuertes limitantes que impiden una mayor mejora en rendimiento y métrica de exactitud en el momento de clasificar imágenes médicas. Por esta razón se han hecho varios estudios sobre la extracción de características y clasificación de manera automática con *Deep Learning* (Kooi, 2017).

Este trabajo tiene como objetivo un modelo de clasificación automática *Deep Learning*, que establezca un compromiso entre confiabilidad y efectividad para la clasificación del cáncer de mamas en mamografías.

La hipótesis está encaminada a, si se implementa un modelo de aprendizaje automático *Deep Learning*, que establezca un compromiso entre fiabilidad y efectividad, entonces se mejora la clasificación del cáncer de mamas en mamografías.

1.3. Tareas de investigación

Las tareas de investigación realizadas son:

- El estudio del estado del arte del aprendizaje automático *Deep Learning*
- El estudio del estado del arte del aprendizaje automático *Deep Learning* en mamografías
- La determinación del modelo de aprendizaje automático *Deep Learning* para la clasificación del cáncer de mama en mamografías
- La aplicación y experimentación del modelo de aprendizaje automático *Deep Learning* seleccionado

CAPÍTULO 2: ESTADO DEL ARTE

2.1. Estado del arte de modelos *Deep Learning*

Cualquier tipo de red neuronal con más de tres capas es considerado un modelo de *Deep Learning*. De acuerdo con LeCun (2015), los principales modelos de *Deep Learning* son:

- *Recurrent Neural Network*: redes neuronales con capacidad de almacenar información de eventos pasado en las neuronas. Esta capacidad de almacenamiento de estados es permitida gracias a las conexiones de las neuronas a las capas delanteras y traseras. Estos tipos de redes son utilizadas para tareas repetitivas y secuenciales como traducción automática de textos y reconocimiento de audio.
- *Generative Adversarial Neural Network*: en este tipo de redes neuronales, dos redes compiten, el uno contra el otro. La función de una red es crear imágenes originales, mientras que la otra intenta clasificar las imágenes como imágenes producidas por humanos o de manera artificial. En estas redes, ambas partes se entrenan simultáneamente. Generalmente se usan redes convolucionales enfrentadas una contra la otra.
- *Deep Convolutional Neural Network*: redes constituidas por la parte de extracción de características y la parte de clasificación. Ambas partes son realizadas automáticamente. La parte fundamental de la extracción de características es el proceso de convolución, el cual es el producto punto entre el filtro y la imagen. Tanto el filtro como la imagen son representado por matrices, siendo la matriz del filtro necesariamente menor en dimensiones. Esta convolución se itera por toda la

imagen, obteniendo como resultado las características extraídas. La clasificación funciona igual que la clasificación de una red neuronal tradicional.

2.2. Trabajos relacionados

En el 2013, Graves desarrolló un modelo de reconocimiento de voz con una métrica de exactitud de 83%. Este modelo fue entrenado con memoria de corto y largo plazo (*Long short-term memory*). Con esta arquitectura de red, y en especial el uso de memoria de corto y largo plazo, se demostró que los *Recurrent Neural Networks* son ideales para tareas secuenciales, como reconocimiento de voz (Graves, 2013).

En el 2016, Chan implementó un modelo nombrado *LAS*, el cual es capaz de transcribir caracteres, palabras y oraciones a texto. La arquitectura empleada es de dos *Recurrent Neural Networks*. Una red se encarga de escuchar, mientras que la otra se encarga de deletrear las palabras. El modelo *LAS* no asume, mediante distribución de probabilidades de las secuencias de caracteres, el resultado final de la palabra, siendo esta una innovación con respecto a otros *Recurrent Neural Networks* aplicados a reconocimiento de voz y deletreo. Este modelo obtuvo un error de palabras de 14.1%, el cual es mayor al estado del arte de 8.0%, pero muestra otra forma muy prometedora de resolver problemas de reconocimiento de voz (Chan, 2016).

Mogren, en el 2016, diseñó un modelo entrenado con música clásica, capaz de generar música original de manera artificial. Este modelo tiene una arquitectura *Generative Adversarial Neural Network* compuesta dos *Recurrent Neural Networks*. La ventaja principal de usar esta arquitectura es el uso de *Long short-term memory* para incrementar el rango de tonos e intensidad en la música generada. Esta música generada gustó mucho a los participantes que se sometieron a escucharla (Mogren, 2016).

En el 2017, Elgammal utilizó una arquitectura de red neuronal *Adversarial Neural Network* para proponer un nuevo sistema de creación arte. En esta arquitectura se usa una red generadora de arte, mientras que la otra clasifica entre arte generado artificialmente o arte auténtico. Ambas redes son redes neuronales convolucionales, utilizando la función de activación *ReLU*. El arte generado por este modelo se distingue del original, ya que el modelo prioriza y recompensa el incremento de la desviación estándar entre el original y el generado, manteniendo patrones y estilos. Esto resulta en imágenes con el mismo estilo y patrones parecidos, pero con diferente contenido.



Figura 5: Ejemplos de arte generado artificialmente (Elgammal, 2017)

Sahiner, en 1996, fue pionero en el área de las redes convolucionales al utilizarlas para la clasificación de lesiones en mamografías. La arquitectura empleada fue una red convolucional de dos capas convolucionales intermedias con tres filtros en cada capa. El tamaño de los filtros fue variand entre 6x6 hasta 10x10. Se utilizaron 700 imágenes de tamaño 32x32 para el entrenamiento de este modelo, obteniendo una métrica de exactitud máxima de 82%. Aunque la base de datos fue pequeña, este estudio mostró el potencial de las reden convolucionales en el área médico.

Askelrod-Ballin, en el 2016, diseño un sistema integrado con varios módulos para la segmentación de mamas y de tejido fibroglandular en imágenes de mamografías, además de la clasificación de tumores. La arquitectura de este modelo consiste en cuatro módulos. El primer módulo está compuesto por una red *Deep Belief* con modelo gaussiano, el segundo módulo es conformado por varios filtros convolucionales, el tercer módulo cuanta con un modelo *Random Forrest* y el cuarto módulo consiste en la unión de regiones en las imágenes. Este modelo primero segmenta la mama de la imagen, luego extrae el tejido fibroglandular para detectar por último las regiones de lesión y clasificarlas entre lesiones malignas y benignas. El segundo módulo, el de los filtros convolucionales, es encargado de la clasificación de las lesiones, mientras que los demás módulos sirven de apoyo para la segmentación automática del modelo. Se utilizaron 850 imágenes para el entrenamiento y se obtuvo una exactitud de 77% (Askelrod-Ballin, 2016).

En el 2016, Arevalo implementó un sistema clasificar lesiones de mamas en mamografías. Este modelo cuenta con una arquitectura de seis capas totales. Cuatro capas convolucionales ocultas, con 64 filtros en cada capa están ubicadas al inicio de la red; y dos capas de de neuronas

completamente conectadas con 400 neuronas y 2 neuronas respectivamente, están ubicados al final. La base de datos de mamografías empleada en este estudio fue creada por los autores. El entrenamiento fue realizado con 736 recortes de las lesiones de imágenes de tamaño 150x150 en escala gris. La mejor métrica de exactitud obtenido en este modelo fue de 80% por este modelo. La adición de parámetros previamente calculados como la edad del paciente no variaron el porcentaje de la métrica de exactitud (Arevalo, 2016).

En el 2016, Kooi implementó un modelo convolucional entrenado con 45 mil imágenes. Este modelo clasifica las mamografías entre benigno y maligno, independientemente del tipo de lesión. La arquitectura empleada es de 5 capas convolucionales y 2 capas de neuronas completamente conectadas. En este estudio se obtuvo una métrica de exactitud del 89%. Al agregar características manuales (como la edad del paciente), se obtuvo una métrica de exactitud del 91% (Kooi, 2016). Este es el modelo con mejor exactitud hasta el momento de la escritura de esta tesis.

En el 2016, Qui creó un sistema novedoso de detección del riesgo que corren las mujeres de contraer cáncer a corto plazo basado en sus mamografías. La arquitectura de su modelo consiste en 3 capas convoluconales de 20 filtros, 10 filtros y 5 filtros respectivamente. La clasificación es realizada por 2 capas de redes neuronales completamente unidas. Se obtuvo un valor de 71.4% en la métrica de exactitud, mostrando la ventaja de usar modelos *Deep Learning* para la aplicación médica en mamografías.

En el 2017, Araujo diseñó un modelo de 5 capas convolucionales y 3 capas de neuronas completamente conectadas para clasificar lesiones de mamografías en cuatro clases distintas. Este

modelo fue entrenado con menos de 1000 imágenes diferentes para un total de 14 000 lesiones diferentes. Su métrica de exactitud inicial fue de 66.7%, el cual pudo ser mejorado al clasificar las lesiones en dos clases (Araujo, 2017).

CAPÍTULO 3: MÉTODO PROPUESTO

3.1 Modelos *Deep Learning* seleccionados

En este trabajo se proponen tres modelos de redes convolucionales para abordar el problema de clasificación de lesiones en mamografías.

El primer modelo cuenta con una arquitectura inicial igual a ResNet50, con dos capas de neuronas completamente conectadas de 512 neuronas y 4 neuronas respectivamente. Se utilizó un *dropout* de 0.5 entre las dos últimas capas. El *learning rate* utilizado fue de 5×10^{-5} , con *nesterov momentum* de 0.99 y un *decay* de 5×10^{-5} . La entrada de este modelo son imágenes de dimensión $250 \times 250 \times 3$, utilizando imágenes de color en formato RGB. Este modelo fue entrenado con 7892 por 20 épocas y con un *batch size* de 258. La inicialización de los pesos fue de manera aleatoria. Las clases de salida son 4: masas benignas, masas malignas, calcificaciones benignas y calcificaciones malignas.

El segundo modelo cuenta con una arquitectura de 4 capas convolucionales y 3 capas de neuronas completamente conectadas. La primera capa convolucional tiene 16 filtros, la segunda capa tiene 32 filtros, la tercera tiene 64 y la cuarta tiene 128 filtros. Todas estas capas usan filtros de tamaño 3×3 , función de activación *PreLu* y *batch normalization* para regularizar los valores de salida. Las tres últimas capas de neuronas completamente conectadas están conformadas por 64 neuronas, 16 neuronas y 2 neuronas respectivamente. Se utilizó un *dropout* de 0.5 entre las tres últimas capas. El *learning rate* empleado fue de 5×10^{-5} , con *nesterov momentum* de 0.99 y un *decay* de 5×10^{-5} . La entrada de este modelo son imágenes de dimensión $250 \times 250 \times 1$, utilizando imágenes en escala gris. Este modelo fue entrenado con 40 000 imágenes por 580 épocas y con un *batch size* de 258. La inicialización de los pesos fue realizada mediante la técnica de HeNormal

(Szegedy, 2015). Las clases de salida son 2: lesiones benignas y lesiones malignas, independientemente de qué tipo de lesión (masa o microcalcificación) sea.

REPETIR COMO EL PRIMER MODELO

El tercer modelo cuenta con una arquitectura inicial igual a ResNet50, con dos capas de neuronas completamente conectadas de 512 neuronas y 4 neuronas respectivamente. Se utilizó un *dropout* de 0.5 entre las dos últimas capas. El *learning rate* utilizado fue de 5×10^{-5} , con *nesterov momentum* de 0.99 y un *decay* de 5×10^{-5} . La entrada de este modelo son imágenes de dimensión $250 \times 250 \times 3$, utilizando imágenes de color en formato RGB. Este modelo fue entrenado con 7892 por 20 épocas y con un *batch size* de 258. La inicialización de los pesos fue realizada con los pesos de *ImageNet*. Las clases de salida en este modelo son 2: calcificaciones benignas y calcificaciones malignas.

3.2 Metodología Experimental

La metodología seleccionada para evaluar se describe mediante los siguientes puntos:

1. Creación del conjunto de datos
2. Preprocesamiento de las imágenes
3. Validación de los resultados basado en la métrica de exactitud

3.2.1 Creación del conjunto de datos

La base de datos fue creada a partir de *Digital Database for Screening Mamography* (DDSM) (Heath, 1998). Esta base de datos cuenta con 3946 imágenes de 2620 pacientes (Pérez, 2015). Cada imagen tiene un fichero *overlay* relacionado, en el cual se especifica el tipo de lesión y la posición de las lesiones. De estas imágenes, 1081 son de masas benignas, 1070 de masas malignas, 894 de calcificaciones benignas y 901 de calcificaciones malignas. Ejemplos de la base de datos se puede ver en la figura 6.

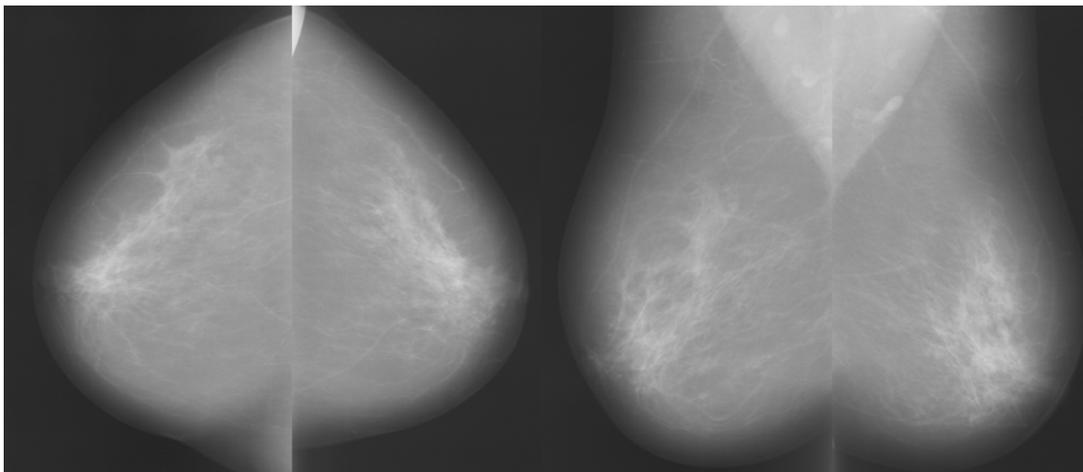


Figura 6. Ejemplo de mamografías extraída de la base de datos DDSM: las dos primeras (de izquierda a derecha) son vistas craneocaudales, las dos últimas son vistas medio lateral oblicuo.

3.2.2 Preprocesamiento de las imágenes

El preprocesamiento de las imágenes se realizó mediante el uso de un programa en MATLAB, producto de una investigación previa (Pérez, 2015), el cual lee las anotaciones de los *overlays* hechas por especialistas sobre las lesiones y recorta estas lesiones de las imágenes para tener imágenes únicamente del área de interés. El funcionamiento de este programa se puede ver en la figura 7. Este programa también mejora el contraste de las imágenes mediante una ecualización adaptiva (Pérez, 2015). Las imágenes también fueron centradas en cero, restando la media de los píxeles. Estas imágenes fueron divididas en subconjuntos disyuntos de entrenamiento y validación con 85% y 15% de las imágenes totales respectivamente.

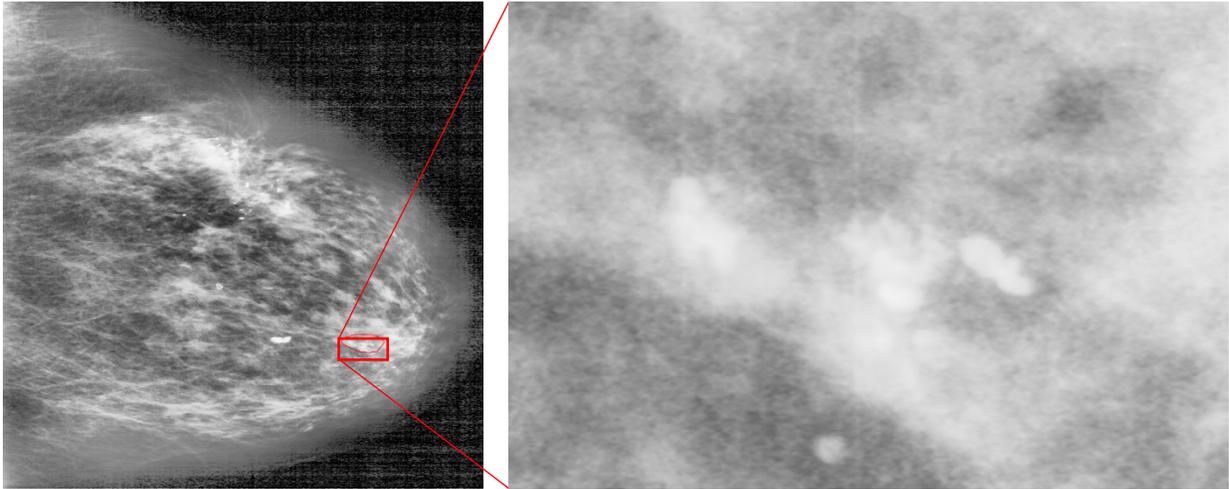


Figura 7. Ejemplo de localización de lesiones en mamografías usando el programa en MATLAB (izquierda) y recorte del área de interés de la lesión (derecha)

3.2.3 Validación de los resultados

La validación de los datos fue realizada basada en la métrica de exactitud del modelo para el conjunto de datos de validación. Para el modelo 1 en la figura 8, se ve que las curvas de los conjuntos de entrenamiento y validación mejoran en la gráfica de pérdida. A partir de la época 8 se ve que la curva validación empieza a subir de nuevo, a diferencia de la curva de entrenamiento de la pérdida. Esta diferencia entre las dos curvas de entrenamiento y validación muestran un claro *overfitting*, ya que el modelo mantiene la confiabilidad al poder clasificar adecuadamente las imágenes del entrenamiento, pero no es efectivo al fallar en la clasificación de las instancias del conjunto de validación.

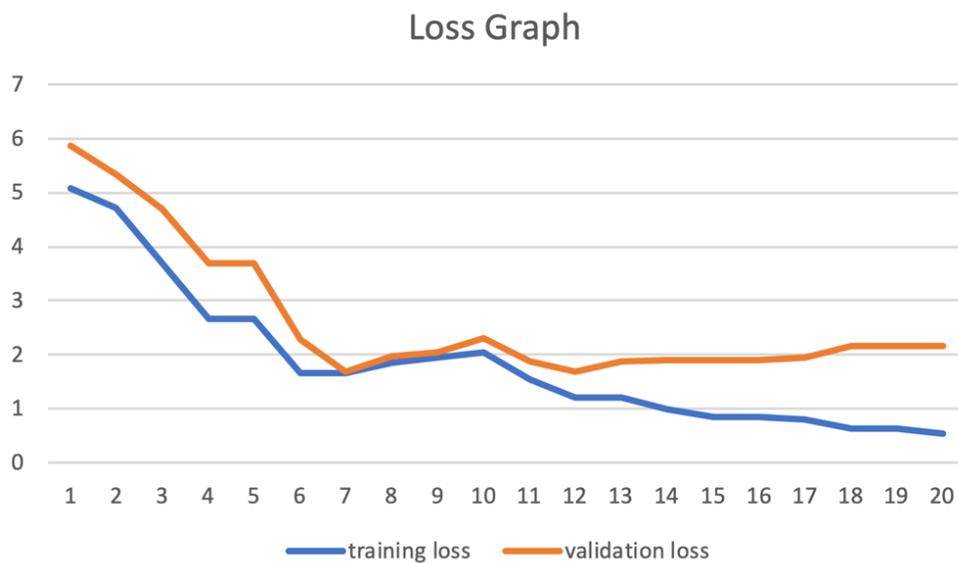


Figura 8. Gráfica de pérdida del modelo 1

En la métrica de exactitud en la figura 9 también se evidencia un claro *overfitting*, ya que la curva de entrenamiento sigue incrementando mientras que la curva de validación se mantiene constante o baja. Esto ocurre porque el modelo empieza a memorizar las imágenes del entrenamiento, lo cual impide que mejore la métrica de exactitud en el conjunto de datos de validación.

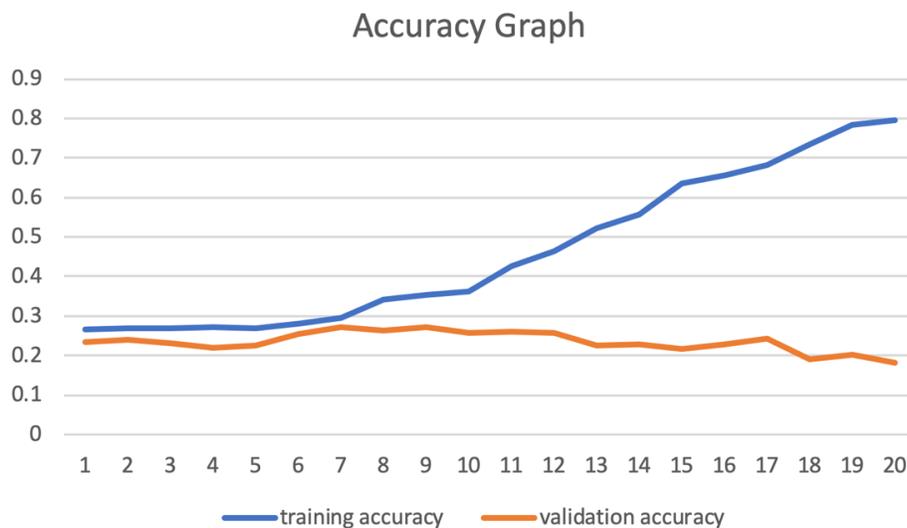


Figura 9. Gráfica de exactitud del modelo 1

En el modelo 2 en la figura 10 se puede apreciar que las curvas de los conjuntos de entrenamiento y validación decrecen en la pérdida e incrementan en la grafica de la métrica de exactitud de manera paralela. Esto muestra claramente la confianza del modelo asociado con la mejora de la métrica de exactitud del conjunto de entrenamiento. También se demuestra la efectividad del modelo, asociado con la capacidad del modelo de clasificar instancias previamente no vistas, evidenciando esto en la curva de la métrica de exactitud del conjunto de validación. Este modelo obtuvo un 62.83% en la métrica de exactitud para el conjunto de validación.

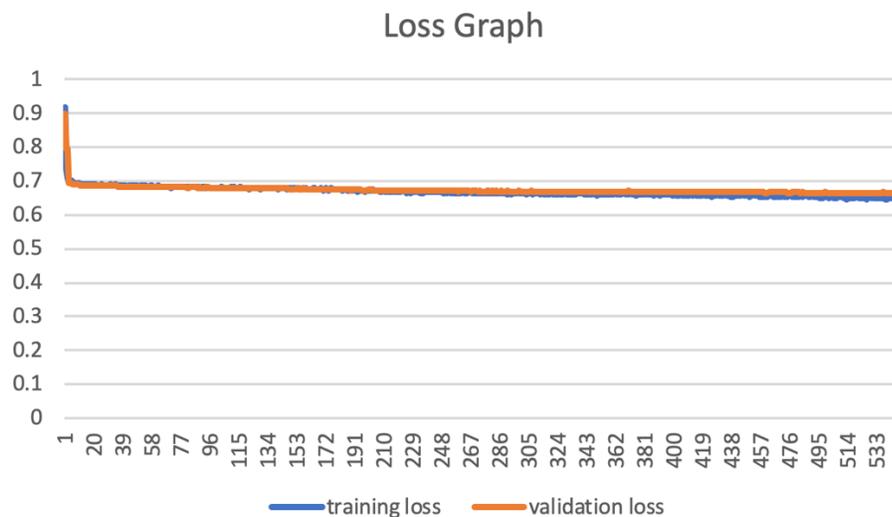


Figura 10. Gráfica de pérdida del modelo 2

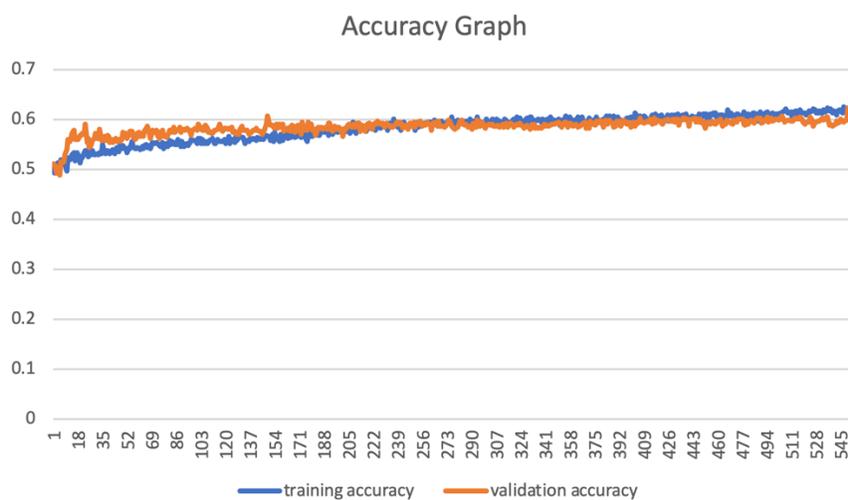


Figura 11. Gráfica de exactitud del modelo 2

En el modelo 3 en la figura 12 y 13 se puede ver que las curvas de los conjuntos de entrenamiento y validación tienden a decrecer en la pérdida y tienden a incrementar en la grafica de la métrica de exactitud de manera paralela. Esto muestra claramente la confianza del modelo

asociado con la mejora de la métrica de exactitud del conjunto de entrenamiento. También evidencia la efectividad del modelo, asociado con la capacidad del modelo de clasificar instancias previamente no vistas, evidenciando esto en la curva de la métrica de exactitud del conjunto de validación. Este modelo obtuvo un 67.63% en la métrica de exactitud para el conjunto de validación.

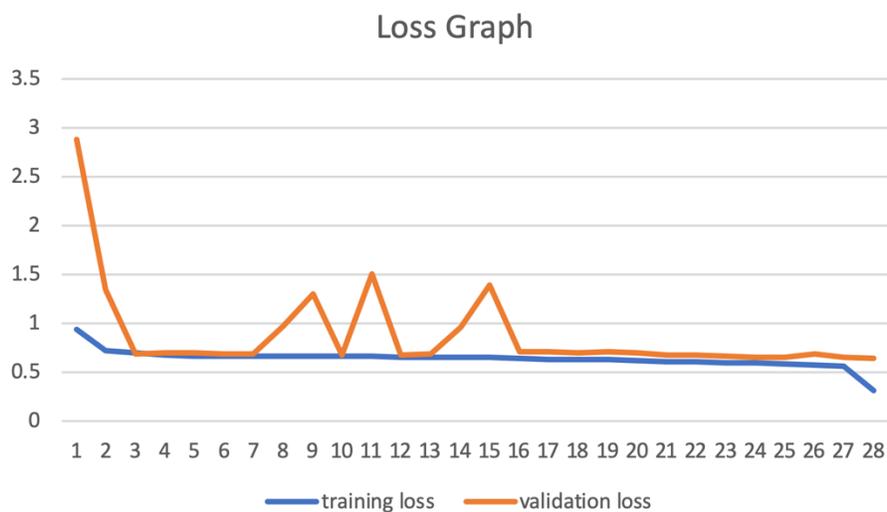


Figura 12. Gráfica de pérdida del modelo 3

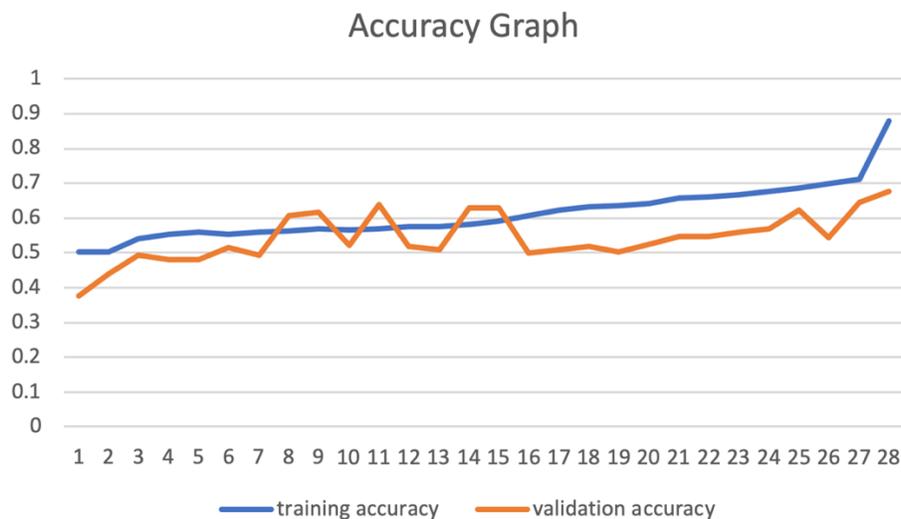


Figura 13. Gráfica de la métrica de exactitud del modelo 3

De los tres modelos propuestos, se obtuvieron valores de la métrica de exactitud de 25%, 62% y 67.63% respectivamente. Las gráficas de todos los modelos mostraron inicialmente un compromiso entre confiabilidad (con la métrica de exactitud en el conjunto de entrenamiento) y efectividad (con la métrica de exactitud en el conjunto de validación) del modelo. El primer modelo mostró rasgos de *overfitting* de manera temprana en el proceso de entrenamiento, siendo esto atribuible a la inicialización aleatoria de los pesos. El modelo 2 y el modelo 3 mostraron una buena capacidad de generalización en relación con el modelo 2, lo cual muestra la efectividad del uso de *batch normalization*, *dropout* e inicialización de parámetros diferente al aleatorio para evitar *overfitting* y mejorar la métrica de exactitud.

El modelo 3 logró mostrar el compromiso entre confianza y efectividad mucho antes que el modelo 2, lo que muestra que el uso de modelos previamente entrenados (*transfer learning*) es beneficioso. A pesar de que se usó una arquitectura compleja en el modelo 3, corriendo el riesgo

de *overfitting*, se demostró que los modelos CNN son capaces de utilizar características extraídas de otros experimentos para reajustarlos y emplearlos para la clasificación de lesiones en mamografías.

CAPÍTULO 4: CONCLUSIONES

4.1 Conclusiones

Debido a la importancia de un buen diagnóstico médico en el caso de cáncer de mamas para incrementar la esperanza de vida del paciente, se han desarrollado diferentes métodos de lectura de mamografías como la doble lectura y los métodos *CAD*. Estos métodos son ineficientes ya que incrementan el número necesario de médicos y están limitados a las características conocidas de antemano. Por esta razón es una ventaja desarrollar modelos de extracción automática de características y clasificación automática de lesiones de mama en mamografías. En este trabajo se implementó con éxito un modelo de clasificación de lesiones en mamografías con una métrica de exactitud de aproximadamente 68%.

El principal limitante de este trabajo fueron los número de imágenes a disposición, ya que se puede ver claramente cómo los modelos son entrenados, mejorando la métrica de exactitud del training y validación de manera paralela, hasta que se llega a un punto donde el modelo empieza a memorizar las imágenes del entrenamiento. Esto es por la falta de imágenes nuevas en el conjunto de datos con las que es entrenado el modelo. Otro limitante bastante grande fue la falta de capacidad de cómputo, ya que no se contó con equipos con GPUs CUDA para poder aprovechar el uso de los GPU para cálculos y entrenamiento del modelo.

4.2 Trabajo futuro

Recomendaciones para futuros trabajos en esta área es incrementar la base de datos de mamografías, ya que esto permitirá que los modelos puedan aprender y generalizar mejor sin que ocurra overfitting; y obtener tarjetas gráficas CUDA, capaces de recibir instrucciones para mandar cálculos de parte del CPU y aprovecharlos para el entrenamiento de los modelos.

CAPÍTULO 5: REFERENCIAS BIBLIOGRÁFICAS

- Araujo, C. et al., (2017). Classification of breast cancer histology images using Convolutional Neural Networks.
- Malich, A., Fischer, D., & Böttcher, J. (2006). CAD for mammography: the technique, results, current role and further developments. *European Radiology*. 16(1), p.1449–1460
- Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*. 65(6) p.386-88.
- Domingos, P., (2012). A Few Useful Things to Know about Machine Learning. [PDF]. Extraído el 19 de abril desde <https://homes.cs.washington.edu/~pedrod/papers/cacml2.pdf>
- Eadie, L., Taylor, P., and Gibson, A. (2012). A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European Journal of Radiology*. 81(1) pp. 70-76.
- Glorot, X., (2011). Deep Sparse Rectifier Neural Networks. Extraído el 5 de mayo del 2019 desde <http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Covariate Shift. Extraído desde <https://arxiv.org/abs/1502.03167>
- LeCun, Y. (2015). Deep Learning. *Nature*. 521, p. 436-44
- LeCun, Y., (1998). Efficient Backpropagation. [PDF]. Extraído el 5 de mayo del 2019 desde <http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>

LeCun, Y., et al., (1998). Gradient-Based Learning Applied to Document Recognition.

IEEE. [PDF]. Extraído el 5 de mayo del 2019 desde

<http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>

N. Pérez, M. A. Guevara, A. Silva, I. Ramos, and J. Loureiro. (2014). Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection. *IEEE*. 2, pp. 209-217.

Paredes, E. (2009). Atlas of Mammography. *Radiology*. 252(3) p. 663.

Pérez, N. (2015). *Improving Variable Selection and Mammography-based Machine Learning Classifiers for Breast Cancer CADx*. (Disertación doctoral). Extraído desde <https://repositorio-aberto.up.pt/bitstream/10216/79069/2/35122.pdf>

Kamangar, F., Dores, G., & Anderson, F. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of clinical oncology*. 24 (14), pp. 2137-2150.

Kooi, T., (2017). Discriminating Solitary Cysts from Soft Tissue Lesions in Mammography using a Pretrained Deep Convolutional Neural Network. *Medical Physics*. 44(3), p. 1017-1027.

Kooi, T., et al., (2016). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis*. 35 p. 303-312

Krizhevsky, A., et al. (2012). ImageNet Classification with Deep Convolutional Neural

Networks. [PDF]. Extraído desde <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

- Suri, J., y Rangayyan, R. (2006). Recent advances in breast imaging, mammography, and computer-aided diagnosis of breast cancer. Washington, US: SPIE press.
- Szegedy, C., et al. (2015). Going Deeper with Convolutions. [PDF]. Extraído desde <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf>
- Simonyan, K., et al. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. [PDF]. Extraído desde <https://arxiv.org/pdf/1409.1556.pdf>
- Ng KH., (2013). Advances in mammography have improved early detection of breast cancer. J Hong Kong Coll Radiol. 6(3) pp. 126-131.
- Wang, H., & Raj, B., (2017). On the Origin of Deep Learning. [PDF]. Extraído desde <https://arxiv.org/pdf/1702.07800.pdf>
- Heath, M. & Bowyer, K. (1998). Current status of the digital database for screening mammography. Digital Mammography. 13, pp. 457-460.
- Graves, A. et al., (2013). Speech Recognition with Deep Recurrent Neural Networks. IEEE. DOI: 10.1109/ICASSP.2013.6638947
- Ledig, C., et al., (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. [Artículo online]. Extraído desde http://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html

- Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., Barkan, E. (2016). A region based convolutional network for tumor detection and classification in breast mammography. *Lecture Notes Computer Science*. 100(8) pp. 197–205.
- Arevalo, J., González, F. A., Ramos-Polla R., Oliveira, J. L., Guevara Lopez, M. A., (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods Programs Biomedical*. 127, 248–257.
- Qiu, Y., Wang, Y., Yan, S., Tan, M., Cheng, S., Liu, H., Zheng, B., (2016). An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology. *Medical Imaging*. 9785 (3) p.978-91.