

Detección de cáncer de seno usando imágenes de histopatología y modelos de aprendizaje profundo pre-entrenados

Breast Cancer Detection using digital histopathology images and pre-trained deep learning models

DOI: <http://dx.doi.org/10.17981/cesta.02.02.2021.04>

Artículo de investigación científica. Fecha de recepción: 13/11/2021. Fecha de aceptación: 14/12/2021.

Harold Agudelo Gaviria

Universidad Santiago de Cali. Cali (Colombia)
harold.agudelo00@usc.edu.co

Milton Sarria-Paja 

Universidad Santiago de Cali. Cali (Colombia)
milton.sarria00@usc.edu.co

How to cite this article:

H. Agudelo & M. Sarria-Paja, "Detección de cáncer de seno usando imágenes de histopatología y modelos de aprendizaje profundo pre-entrenados", *J. Comput. Electron. Sci.: Theory Appl.*, vol. 2, no. 2, pp. 27–36, 2021. <https://doi.org/10.17981/cesta.02.02.2021.04>

Resumen

El cáncer es una enfermedad que se puede originar en cualquier parte del cuerpo. Comienza cuando las células infectadas crecen de forma descontrolada sobrepasando a las células sanas. El cáncer de seno, en su mayoría carcinomas, es el tipo más común entre las mujeres de todo el mundo. Los procedimientos utilizados para la detección de la enfermedad son aproximaciones diagnósticas, algunos de estos son invasivos. Usando herramientas digitales, es posible desarrollar o implementar sistemas de diagnóstico asistido para agilizar el proceso y permitir mayor confiabilidad de los análisis. El presente estudio se realiza con imágenes digitales de histopatología a partir de la base de datos de acceso abierto. Se evalúan tres escenarios, partiendo desde un esquema de aprendizaje de máquinas clásico, regresión logística combinado con análisis de componentes principales. Luego se incluye el uso de modelos profundos pre-entrenados y finalmente se evalúa un modelo profundo con una red neuronal convolucional. El rendimiento de cada uno de los métodos sometidos a estudio se evaluó calculando las medidas diagnósticas de precisión, sensibilidad y especificidad, logrando así encontrar el modelo que mejor se adecua a la tarea abordada. Se observa que los modelos pre-entrenados aportan información altamente discriminante a pesar de haber sido entrenados para una tarea completamente diferente. En general los modelos profundos permiten mejorar significativamente la especificidad del sistema al comparar con el enfoque clásico.

Palabras clave— Aprendizaje profundo; cáncer de seno; imágenes digitales; diagnóstico asistido

Abstract

Cancer is a disease that can start anywhere in the body. It begins when infected cells grow out of control, outpacing normal cells. Breast cancer is the most common type in women around the world. Most of them are carcinomas, these originate in the cells that cover the organs and tissues of the body. The procedures used to detect the disease are diagnostic approaches, and some are invasive. Using digital tools, it is possible to develop or implement assisted diagnostic systems to streamline the process and allow greater reliability of the analyzes. The present study is carried out with digital histopathology images. In this study three scenarios were evaluated, starting from a classical machine learning scheme, logistic regression combined with principal component analysis. Then we include the use of pre-trained deep models and finally a deep model based on a convolutional neural network. The performance for each approach was evaluated by calculating three diagnostic measures such as precision, sensitivity and specificity. It is observed that the pre-trained models provide highly discriminative information despite having been trained for a completely different task. In general, deep models allow to significantly improve the specificity of the system when compared with the classical approach.

Keywords— Deep learning; breast cancer; digital images; assisted diagnosis



I. INTRODUCCIÓN

El cáncer es una enfermedad que se puede originar en cualquier parte del cuerpo. Comienza cuando las células infectadas crecen de forma descontrolada sobrepasando a las células sanas. Este fenómeno dificulta que los sistemas funcionen correctamente. Para muchas personas, la enfermedad puede tratarse muy eficazmente; de hecho, ahora hay un mayor número de personas que tienen una vida plena después de haber recibido su tratamiento, es por esto por lo que es de vital importancia su detección temprana [1].

El cáncer de seno es el tipo más común en las mujeres de todo el mundo. En su mayoría son carcinomas, estos se originan en las células que cubren los órganos y tejidos del cuerpo [2]. Los procedimientos utilizados para la detección de la enfermedad son aproximaciones diagnósticas, es decir que no confirman la presencia de la enfermedad. Entre las aproximaciones diagnósticas se tienen las muestras de sangre, orina u otras sustancias del cuerpo, Tomografías Computarizadas (TC), exploración nuclear, ecografía, Resonancia Magnética (RM), Tomografía por Emisión de Positrones (TEP) y rayos X. El único método por el cual se llega a la confirmación es la biopsia, la cual requiere de aguja, endoscopio o cirugía, y consiste en la remoción de tejido de cualquier parte del cuerpo para su posterior estudio, esto la hace 100% invasiva [1], [3].

El uso de imágenes y el desarrollo de algoritmos para su procesamiento ha tomado fuerza en el ámbito científico. En particular, en el campo de la inteligencia artificial y visión por computador el número de aplicaciones ha aumentado en los últimos años. Gracias a los avances tecnológicos implementados recientemente es posible procesar imágenes médicas a gran escala y velocidad. Con estas imágenes se logra identificar órganos y tejidos, además de recolectar datos que ayudan a caracterizar y cuantificar las patologías, siendo esta una herramienta de diagnóstico muy potente [4]. La inspección visual de tumores es una técnica usada por los profesionales de la salud para realizar valoraciones iniciales. Recientemente se demostró que el diagnóstico puede ser más eficaz empleando algoritmos de aprendizaje automático. Los resultados mostraron que un médico experimentado puede diagnosticar con un 79.97% de precisión, basado únicamente en el análisis visual de las imágenes, sin tener en cuenta otros exámenes, como la biopsia, que permiten confirmar el diagnóstico. Un sistema automático que emplea únicamente imágenes, logra un 91.1% de diagnósticos correctos, siendo una herramienta eficaz para tomar decisiones [5].

En el presente trabajo se hace uso de diferentes modelos pre-entrenados como VGG16, VGG19, RESNET50 y una arquitectura de red neuronal convolucional entrenada completamente a partir de los datos disponibles. Estas arquitecturas de aprendizaje moderno se comparan con técnicas clásicas de clasificación [6], empleando diferentes enfoques para encontrar la mejor estrategia.

El presente manuscrito se encuentra distribuido en las siguientes secciones. Inicialmente en la primera sección se introduce al lector en la temática a tratar. En la siguiente sección se hace un corto resumen sobre los trabajos relacionados. En la sección 3 se definen los materiales usados y las técnicas para llevar a cabo la metodología. En la cuarta sección se presentan los resultados obtenidos y finalmente en la sección final se presentan las conclusiones.

II. TRABAJOS RELACIONADOS

En la bibliografía se pueden encontrar trabajos importantes donde se presentan enfoques que utilizan redes neuronales profundas para el análisis de imágenes en el diagnóstico de cáncer. Por ejemplo, en el trabajo presentado por universidades estadounidenses y colombianas en 2014 [7] se aborda el problema de detección de cáncer de mama empleando imágenes de 162 pacientes. Los autores comparan redes neuronales convolucionales con un sistema de clasificación basado en bosque aleatorio (random forest). Se reportan resultados de exactitud entre 77% y 84%, empleando diferentes configuraciones de los sistemas. Por otro lado, en China [8] se realiza un estudio empleando una arquitectura híbrida de redes convolucionales sobre una base de datos de 7909 imágenes. Los resultados reportados se aproximan a 90% de exactitud en la clasificación. Otro trabajo importante es reportado en España y Colsanitas de Colombia [9], donde se emplean modelos pre-entrenados como VGG16 y VGG19 para la clasificación de cáncer de mama. El conjunto de datos consta de 845 imágenes, de las cuales 437 corresponden a carcinoma ductal invasivo. Los resultados reportados alcanzan el 95% de exactitud en la clasificación.

También es importante mencionar el trabajo presentado por la república de Korea en 2018 [10], donde se desarrolla un sistema de ayuda diagnóstica asistida por computador usando el algoritmo YOLO, este sistema capaz de detectar y clasificar objetos en imágenes utilizando una red neuronal convolucional. El sistema tiene la capacidad de detectar la ubicación de las masas mamarias y la clasificación en benignas o malignas en una mamografía [10]. En otro trabajo que presenta mejores resultados, se compara una arquitectura de Red Neuronal Convolutiva (CNN), contra arquitecturas como INCEPTION V3, VGG16 y RESNET50, sobre algunas bases

de datos de imágenes de radiografías de mama [11]. En este caso, los mejores resultados los obtuvo el modelo CNN logrando una precisión del 98.94%. Dicho modelo se utilizó como línea de base para construir el Marco de Detección del Cáncer de Mama. Para evaluar el sistema se probó en una base de datos independiente (MIAS-Medición Independiente de Aprendizajes), obteniendo un 98.23% de precisión [11].

III. MATERIALES Y MÉTODOS

Como se mencionó anteriormente, el cáncer de seno es el tipo de cáncer más común en las mujeres de todo el mundo con casi 2.1 millones de casos nuevos reportados en 2018, en su mayoría carcinomas, originados en las células que cubren los órganos y tejidos del cuerpo [12]. Usando herramientas digitales, es posible desarrollar o implementar sistemas de diagnóstico asistido para agilizar el proceso y permitir mayor confiabilidad de los análisis. Los sistemas automáticos cumplen con unas etapas que garantizan su eficacia.

A. Estructura general de un sistema de diagnóstico asistido

Un sistema de diagnóstico asistido se implementa con el objetivo de detectar patrones en los datos con la ayuda de herramientas computacionales. Este tipo de sistemas sirven como sistema de apoyo a los expertos, quienes tienen en cuenta otros factores para finalmente emitir un diagnóstico. Las motivaciones para la construcción de tales herramientas son: a) Reducción de errores observacionales humanos y b) Reducción del tiempo y el esfuerzo asociado con el diagnóstico. Cuando se trata de sistemas de diagnóstico usando imágenes médicas, se pueden plantear una serie de etapas que son necesarias para llevar a cabo la tarea en cuestión:

Adquisición de imágenes: El proceso de captura consiste en convertir un objeto o una escena, en una representación apta para ser procesada por una computadora. Tal adquisición o conversión puede ser realizada por medio de un escáner o cámara fotográfica. El resultado varía dependiendo del proceso de digitalización usado y del método de codificación [13].

Pre-proceso: Para esto se utilizan diferentes técnicas que aplican una operación matemática sobre la imagen. En este trabajo se emplean técnicas para reducir la redundancia e interdependencia en las variables de entrada. En este caso nuestras variables de entrada son los valores de los píxeles directamente. Para ello se usan técnicas de de-correlación (Whitening, Analisis de Componentes principales, entre otros). El objetivo de estos procesos es modificar las imágenes para mejorar su representación en un espacio de características y facilitar el proceso de aprendizaje [14].

Segmentación: Su objetivo es dividir la imagen en partes que tienen alta correlación con objetos o áreas de interés. Se distinguen dos tipos de segmentación: parcial y completa, donde la primera encuentra regiones en la imagen que corresponden a superficies, sin agruparlas en objetos, mientras que la segunda detecta los objetos reales [15].

Extracción de características: También conocido como representación y descripción. En esta etapa, a la imagen le son extraídos un conjunto de características que describen las propiedades físicas de los objetos como son: el color, textura, área, perímetro, ancho, alto, promedio de intensidad, centroide, orientación, color, entre otros [16].

Aprendizaje e Inferencia: Requiere de un entrenamiento, para eso se hace necesario usar bases de datos que con la ayuda de especialistas se han etiquetado y se convierten en ejemplos para ser presentados al algoritmo de aprendizaje [6]. En la Fig. 1 se muestra de forma resumida un diagrama de bloques para un sistema de reconocimiento de patrones para uso general. Este sistema se puede usar para el problema que se esta abordando en este manuscrito.

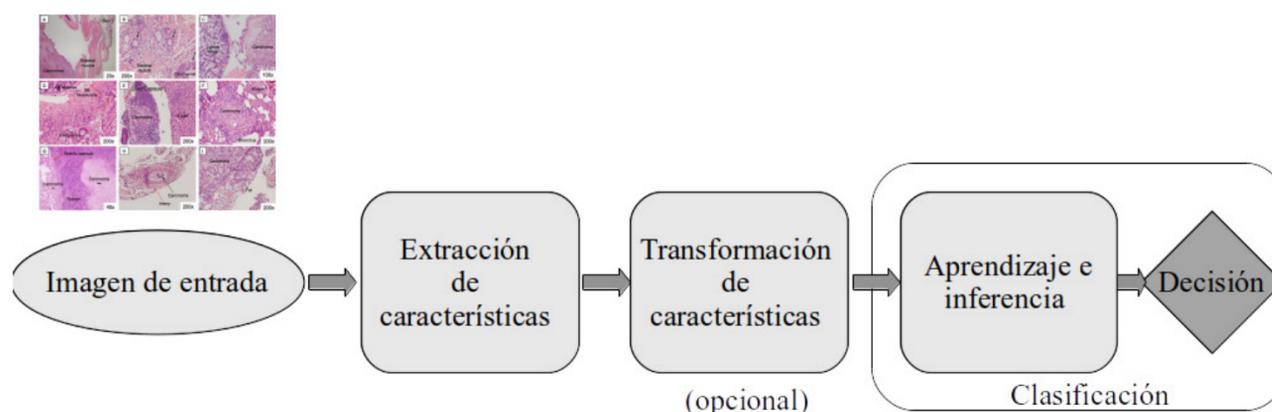


Fig. 1. Diagrama de bloques de un sistema de reconocimiento de patrones.

Fuente: Elaboración propia.

Las técnicas para extracción y/o transformación de características buscan representar la información relevante contenida en la imagen en un conjunto de variables reducido. Dos de las técnicas más utilizadas son Análisis de Componentes Principales (Principal Component Analysis-PCA) y Análisis Discriminante Lineal (Linear Discriminant Analysis- LDA) [6].

La herramienta PCA es una técnica para reducir la dimensión de entrada utilizando información estadística de segundo orden. La esencia de PCA es proyectar muestras de datos que están en un espacio de alta dimensión, a un espacio de baja dimensión. Esto se logra mediante una transformación lineal mientras se conservan las características de datos originales tanto como sea posible. En el presente trabajo y para todos los casos estudiados se selecciona el número de componentes principales conservando el 98% de varianza acumulada [6], [17]

También podemos mencionar técnicas como ZCA (Zero Component Analysis) y Whitening, estas aplican rotaciones sobre los datos originales y permiten eliminar interdependencia entre medidas diagnósticas, pero PCA además de hacer eso, selecciona solamente las componentes que contienen mayor varianza, esto la hace una herramienta más completa [6], [17].

Por otro lado, se debe tener en cuenta que es necesario estimar el desempeño de los sistemas implementados. Para ello se emplean medidas diagnósticas [18]. Las medidas diagnósticas que se calculan son: Sensibilidad, Especificidad y Precisión. Las mismas son calculadas en función a los falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos. Se complementan las medidas mencionadas con el cálculo de la Curva ROC (Receiver operating characteristic) Las curvas ROC (Característica Operativa del Receptor) son una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario [19].

B. Clasificación

El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como biología, sicología, medicina, visión por computador, inteligencia artificial, entre otros campos [6]. Un sistema de reconocimiento de patrones busca identificar el patrón como miembro de una clase ya definida (clasificación supervisada). También se tiene el caso donde el patrón se debe asociar a una clase todavía no definida (clasificación no supervisada, agrupamiento o clustering) [6]. El diseño de un sistema de reconocimiento de patrones se lleva a cabo normalmente en tres fases: Adquisición y preproceso de datos, Extracción de características y Toma de decisiones o agrupamiento [6].

Para el problema abordado se deben seleccionar cuidadosamente los algoritmos de clasificación. Se debe tener en cuenta el tamaño de la base de datos utilizada. Normalmente algoritmos como K-vecinos cercanos, o máquinas de soporte vectorial están concebidos para bases de datos pequeñas. Por esta razón, emplearemos una estrategia de clasificación sencilla y que permita realizar el proceso de aprendizaje sobre una base de datos de tamaño considerable.

Regresión logística (LR): Entre los numerosos algoritmos de aprendizaje automático, la LR es un modelo de clasificación que se utiliza ampliamente. Tiene una interpretación probabilística directa. Además de la información de la etiqueta de clase, puede obtener probabilidades de clasificación directas [6]. Gracias a ser un algoritmo simple, y los parámetros del modelo se pueden actualizar con técnicas de gradiente, se puede emplear como una línea base para comparar o combinar con técnicas más avanzadas [20]. En particular, para el presente trabajo se usa este método debido a la cantidad masiva de datos que se usarán.

Aprendizaje profundo (Deep learning): Es una sub-categoría del aprendizaje automático. En la literatura científica se le ha denominado *deep learning* y trata del uso de las redes neuronales artificiales en diferentes aplicaciones, incluyendo clasificación y regresión. Es uno de los campos más activos de investigación en ciencias computacionales. Uno de los desafíos se debe resolver al entrenar un sistema usando modelos profundos, es la cantidad de datos necesaria para estimar de forma adecuada los parámetros [21]. En los últimos años el aprendizaje profundo ha tenido gran impacto en múltiples disciplinas del saber entre las que tiene gran relevancia las aplicaciones médicas. Las aplicaciones de la tecnología de aprendizaje profundo a las imágenes de cáncer pueden ayudar a los patólogos en la detección y clasificación del cáncer en las primeras etapas de su desarrollo para permitir que los pacientes tengan tratamientos apropiados que puedan aumentar su supervivencia [22], [23].

Existen diferentes arquitecturas que se utilizan para aplicaciones de visión por computador. Sin embargo, las dos arquitecturas que tienen gran popularidad son las redes neuronales residuales [24], son redes neuronales que usan funciones residuales y de acuerdo con los autores, permiten reducir sustancialmente la complejidad en el proceso de optimización y en la medida que se incrementa la profundidad de la red pueden ganar mayor exactitud en la tarea para la cual se están entrenando. Por otro lado, la arquitectura desarrollada en la Universidad de Oxford en el Grupo de Geometría Visual (Visual Geometry Group-VGG) también conocida como VGG-

Net. La ventaja de VGG es que, al apilar múltiples capas compuestas por filtros de tamaño pequeño, aumenta el campo receptivo efectivo de la red, mientras se reduce el número de parámetros. En contraste con el uso de menos capas con filtros más grandes para el mismo campo receptivo [25]. Uno de los problemas que presentan estas arquitecturas es la gran cantidad de datos requeridos para estimar los parámetros del modelo.

La transferencia de aprendizaje (*transfer learning*) es una de las soluciones prácticas para reducir los datos necesarios para el entrenamiento de una red neuronal profunda. En este caso, se busca reutilizar la experiencia o conocimiento ganado en una tarea específica, para resolver problemas en tareas similares. Su innovación central es que se toma el conocimiento adquirido por un modelo profundo para ayudar en el diseño y construcción de un modelo nuevo [26].

C. Marco experimental

El objetivo del presente trabajo es evaluar diferentes enfoques de clasificación en el proceso de detección de cáncer de seno. La tarea se plantea sobre la evaluación automática de imágenes de histopatología que han sido previamente evaluadas por especialistas. Los esquemas de clasificación se evalúan usando medidas diagnosticas acorde a lo sugerido en la literatura [18]. A continuación, se describen los datos utilizados y el esquema general de evaluación para los sistemas de clasificación.

Base de datos: El conjunto de datos que se usa para el presente trabajo corresponden a imágenes del Carcinoma Ductal Invasivo (CDI), el más común de todos los cánceres de mama. El conjunto de datos fue depurado originalmente por Andrew Janowczyk en el año 2016 pero está disponible en dominio público en el sitio web de Kaggle. Las imágenes originales son de gran tamaño (en términos de dimensiones espaciales), por lo tanto para facilitar el trabajo se extraen un total de 277 524 parches de 50×50 píxeles, lo que resulta en 198 738 ejemplos negativos (es decir, sin cáncer de mama) y 78 786 ejemplos positivos (se encontró cáncer de mama). La base de datos no tiene el mismo número de ejemplos para las clases involucradas. En este caso se presenta más del doble de ejemplos de datos negativos que puntos de ejemplos positivos. Por esta razón se emplean diferentes métricas de desempeño para evaluar de forma objetiva los sistemas de clasificación. Para más detalles sobre la base de datos se recomienda consultar [23].

Estrategia de validación: El primer paso después de obtener la base de datos consiste en separar los ejemplos en tres conjuntos disyuntos: Entrenamiento, Validación y Prueba (Fig. 2).

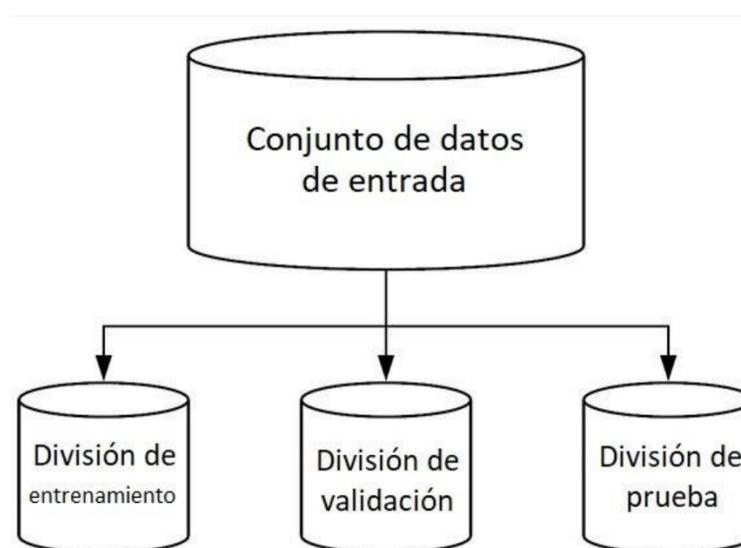


Fig. 2. División de la base de datos en tres sub-conjuntos: Entrenamiento, Validación y Prueba.
Fuente: Elaboración propia

Los tres subconjuntos en los que se divide la base de datos tienen un propósito específico. El conjunto de entrenamiento corresponde al conjunto de datos para estimar parámetros y equivale al 70% de los datos. El conjunto de validación corresponde al conjunto de datos para ajustar los modelos y validar el funcionamiento, corresponde al 10% del total de los datos. Finalmente, el conjunto de prueba se usa para estimar las medidas de rendimiento y presentar resultados finales, corresponde al 20% del total de los datos. Los tres conjuntos conservan las proporciones en el número de registros para cada clase [6].

Empleando estos tres conjuntos se evalúan los siguientes escenarios empleando los sistemas de clasificación descritos anteriormente.

- *Escenario 1*: Se emplea un enfoque clásico de aprendizaje de máquinas como es regresión logística. Para este escenario se usan las imágenes directamente como entrada al sistema. Posteriormente se aplica una técnica de procesamiento como ZCA, whitening o PCA. Este pre-proceso permite adecuar la información presente en la imagen de tal forma que se pueda presentar como entrada al modelo de regresión logística. Este caso corresponde al modelo de referencia contra el que se comparan los siguientes escenarios.
- *Escenario 2*: En este caso se emplean modelos neuronales profundos. El objetivo es evitar el entrenamiento de todo el modelo neuronal al emplear modelos pre-entrenados. En este caso se emplean los modelos VGG16, VGG19 y RESNET50. Para conocer los detalles de la arquitectura de estos modelos se recomienda consultar el trabajo de la Universidad de Oxford [25]. Los modelos pre-entrenados tienen parámetros que han sido ajustados para una tarea específica, sin embargo, si se omite la etapa de clasificación, estos modelos pueden ser usados como extractores de características. Dichas características pueden ser usadas posteriormente para ser usadas en un esquema de clasificación estándar como el usado en el escenario 1, es decir, regresión logística.
- *Escenario 3*: En el tercer escenario se entrena una Red Neuronal Convolutiva (CNN) con una arquitectura simple, que se puede entrenar sin que requiera altos recursos computacionales. Es un modelo que imita hasta cierto punto el modelo VGG, sin embargo, en este caso se entrenan todos los parámetros y se busca que sea una arquitectura de entrenamiento rápido para reducir el tiempo de cálculo y estimación de parámetros. La diferencia con el modelo VGG, es que se utiliza convolución separable en lugar de capas de convolución estándar.

Para el *escenario 3* se emplea una red neuronal con filtros de 3×3 , y se emplean una función de activación de rectificación lineal (*relu*) para todas las capas. Se tiene una capa de entrada con 32 filtros, seguida de max-pooling y dropout de 0.25. Posteriormente se agregan dos capas con 64 filtros cada una, seguidas de max-pooling. Finalmente 3-capas con 128 filtros cada una, finalizando con max-pooling. Se agrega una capa completamente conectada con 256 unidades para alimentar la capa de clasificación que usa softmax.

La convolución separable en profundidad se ha usado con anterioridad y el ejemplo más notable es la arquitectura Xception [27]. Una explicación detallada sobre las diferencias entre convolución separable y convolución estándar está fuera del alcance de este trabajo. Sin embargo, la convolución separable tiene las siguientes características: Es más eficiente. Requiere menos memoria. Requiere menos cálculo. Puede funcionar mejor que la convolución estándar en algunas situaciones.

IV. RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados obtenidos de los tres escenarios propuestos previamente. En primer lugar, para el *escenario 1*, es necesario determinar cuál es la técnica de pre-procesamiento adecuada que permita alcanzar los mejores resultados con el sistema de clasificación seleccionado.

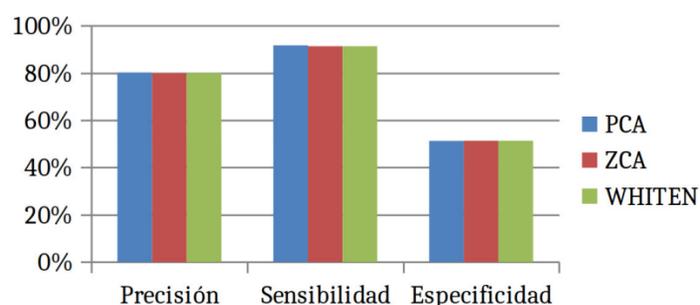


Fig. 3. Selección de la técnica de pre-procesamiento.
Fuente: Elaboración propia.

En primera instancia se realiza la comparación entre las técnicas PCA, ZCA y WHITENING, usando imágenes de 50×50 píxeles. En la Fig. 3 se grafican los valores obtenidos al evaluar las técnicas mencionadas. En general para los tres casos estudiados se puede apreciar que los valores de precisión, sensibilidad y especificidad son similares. Sin embargo, para efectos de reducir la carga computacional, se selecciona la técnica PCA ya que permite seleccionar un conjunto de componentes principales y así reducir la dimensión de entrada al clasificador. Para reducción de dimensión se aplica PCA reteniendo el 98% de varianza acumulada. Adicionalmente, es la técnica que obtuvo porcentajes ligeramente mayores en las medidas diagnósticas de precisión y sensibilidad. Este corresponde a nuestro modelo de referencia. El siguiente paso corresponde a evaluar el *escenario 2*.

En segunda instancia, se evaluó el *escenario 2*, que corresponde a realiza una comparación usando los modelos VGG16, VGG19 y RESNET50 para extraer características sobre las imágenes y posteriormente alimentar dichas características a un sistema como el descrito en el *escenario 1*, es decir PCA + Regresión logística. Para analizar el efecto de redimensionar las imágenes, se realiza también el experimento al extrapolar las imágenes originales a dimensiones de 64×64 y 100×100 . Al igual que en el *escenario 1*, se aplica PCA reteniendo el 98% de varianza acumulada a las características obtenidas del modelo pre-entrenado. Los resultados se muestran en la Fig. 4.

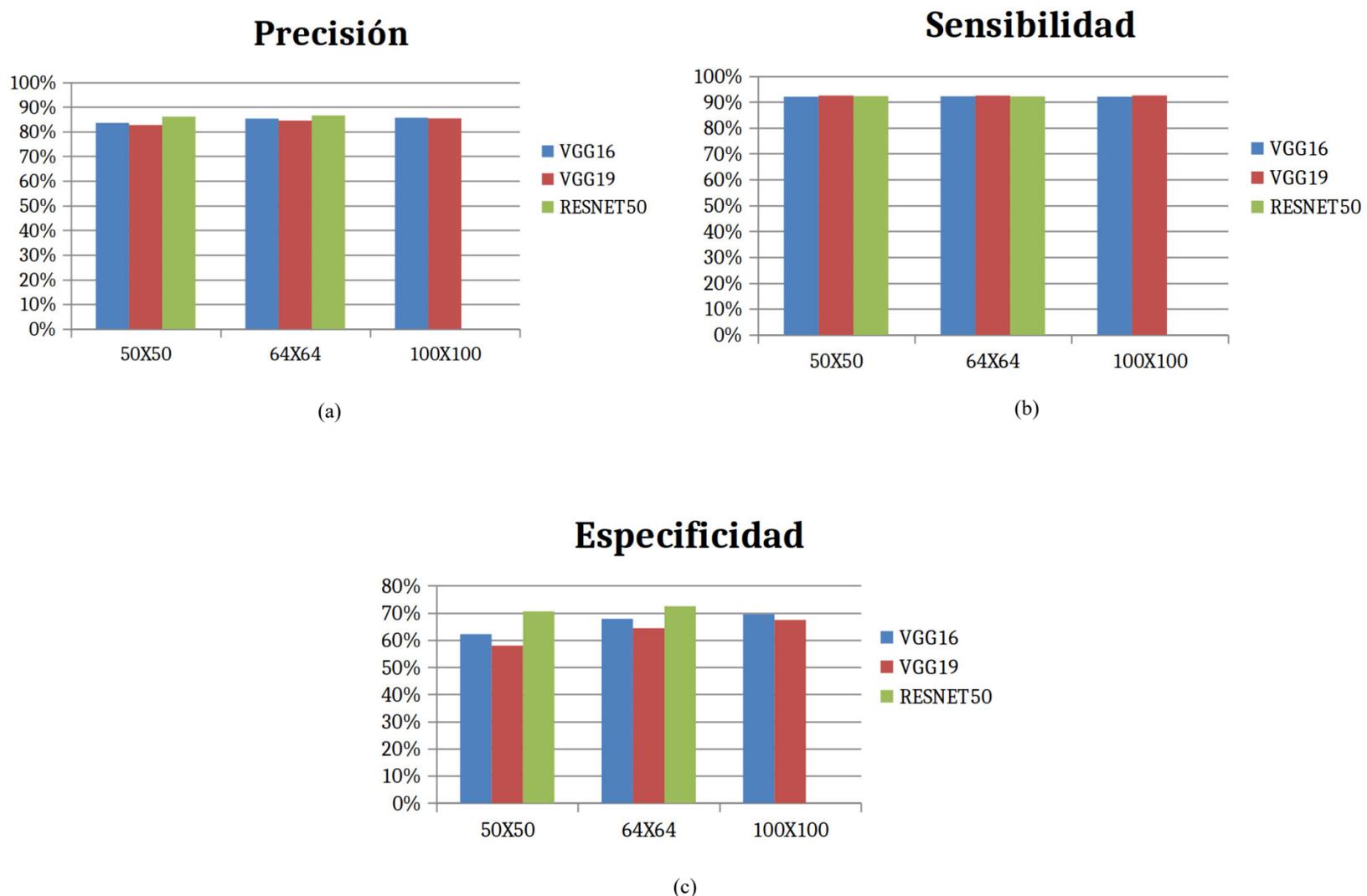


Fig. 4. Pruebas diagnósticas sobre un sistema de clasificación usando modelos pre entrenados para extracción de características.
Fuente: Elaboración propia.

Se observan pequeñas diferencias entre los porcentajes obtenidos, sin embargo, el modelo RESNET50 presenta algunas diferencias favorables con respecto a los modelos VGG. El mejor porcentaje de precisión se logra con el modelo RESNET e imágenes de 50×50 , donde se tiene un porcentaje de 86.05% (Fig. 4a). También se puede observar que cuando se extrapolan las imágenes a un tamaño de 100×100 el sistema de clasificación presentaba problemas de convergencia.

Al analizar los resultados de sensibilidad, se puede observar que las diferencias son mínimas. En este caso la VGG19 presenta el mejor valor de sensibilidad con un 92.47%, usando imágenes de 50×50 (Fig. 4b). Finalmente, se observa los valores de especificidad (Fig. 4c). En este caso se pueden observar diferencias notables, donde el modelo RESNET50 alcanza un porcentaje de 70.57% usando imágenes de 50×50 .

De acuerdo con estos resultados, se puede concluir que la técnica más efectiva para extraer características en el contexto que se aborda en este manuscrito es RESNET50. A pesar de no ser superior que VGG19 al medir sensibilidad, es la técnica que presenta porcentajes altos y balanceados en las tres medidas diagnósticas.

Finalmente, se evaluó el *escenario 3*, en este caso se usan las imágenes en su tamaño original (50×50 pixeles), y se comparan los resultados de la CNN que se ha entrenado completamente vs los mejores sistemas del *escenario 1* y *escenario 2*. Es decir, PCA + Regresión logística, y cuando se usa el modelo RESNET50 para extracción de características. Los resultados se muestran en la Fig. 5.

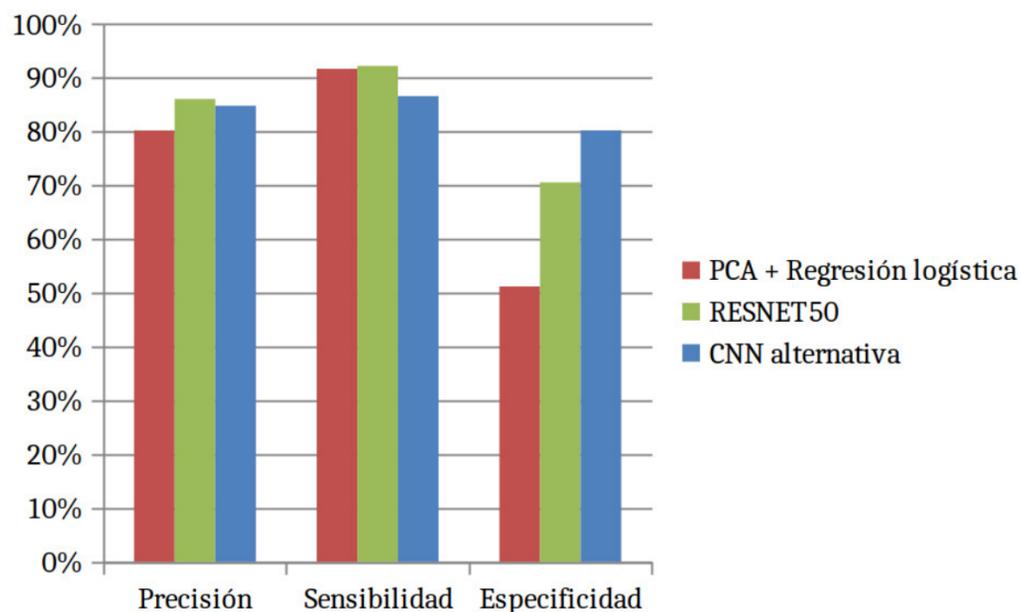


Fig. 5. Comparación entre los diferentes sistemas evaluados en los escenarios 1, 2 y 3.
Fuente: Elaboración propia,

Como se puede observar en la Fig. 5, el sistema de clasificación basado en características extraídas del modelo RESNET50 muestra ligera superioridad en cuanto a precisión y sensibilidad. Sin embargo al medir especificidad se puede observar que la CNN descrita en el escenario 3 muestra una clara superioridad. Se puede observar que en general el sistema basado en RESNET50 y la CNN muestran resultados similares y es difícil determinar cual es el mejor sistema. Para realizar un comparación global de los sistemas que se están evaluando, en la Fig. 6 se presenta la curva ROC para cada caso.

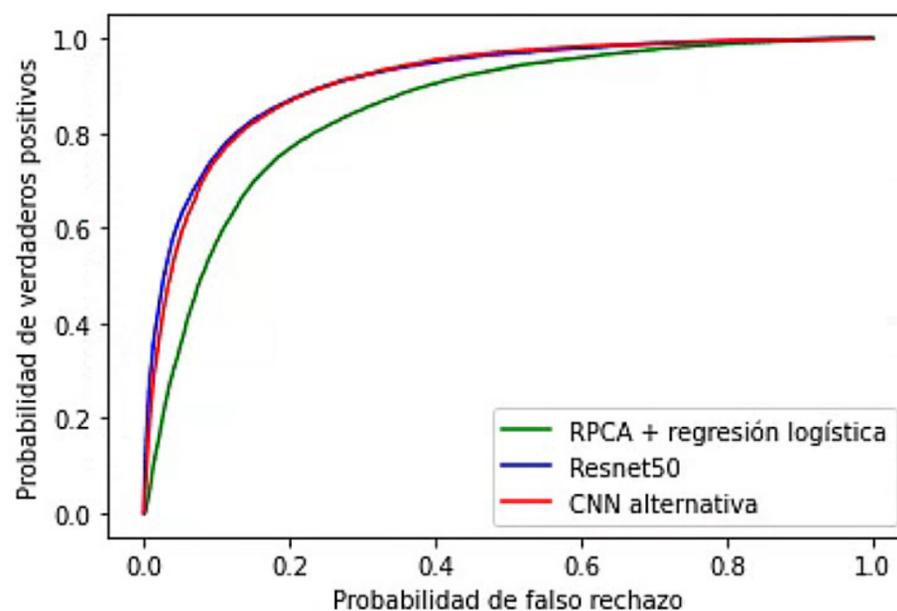


Fig. 6. Comparación entre los diferentes sistemas evaluados en los escenarios 1,2 y 3 usando la curva ROC.
Fuente: Elaboración propia,

La curva ROC muestra claramente que los modelos que usan en alguna de sus etapas técnicas de aprendizaje profundo son superiores al enfoque clásico. También se puede observar que no es necesario entrenar completamente un modelo profundo, es posible emplear modelos pre-entrenados y empleando el concepto de transferencia de aprendizaje se puede evitar el proceso de entrenar todo un modelo profundo.

CONCLUSIONES

En este trabajo se aborda el problema de detección de cáncer de seno por medio de imágenes digitales de histopatología. Se propone comparar diferentes enfoques de aprendizaje supervisado para clasificación automática de imágenes. Uno de los principales problemas que se debe considerar es el tamaño de la base de datos, es decir, el número de registros de entrada es considerable, y las técnicas clásicas de aprendizaje supervisado no están diseñadas para estimar parámetros a partir de bases de datos de tamaño considerable. Se requiere por lo tanto abordar técnicas que permitan uso eficiente de memoria y entrenamiento por lotes.

Los enfoques que se abordan desde el punto de vista clásico corresponden a un clasificador lineal basado en el paradigma de regresión logística combinado con PCA. Se utilizan las imágenes originales como entrada al sistema y se aplica PCA para reducir la dimensión del espacio de entrada. El segundo enfoque que se aborda hace uso de modelos pre-entrenados para extracción de características. Tales características son la entrada al sistema de regresión logística combinado con PCA. Finalmente se implementa un modelo que utiliza una estructura similar a los modelos pre-entrenados como lo es la arquitectura VGG, y en este caso se entrenan todos los parámetros, lo que representa un modelo completo no solo para extraer características sino también para clasificación.

Los resultados muestran que es posible alcanzar tasas de acierto superiores a 90%, sin embargo, es necesario analizar otros aspectos del rendimiento. La tasa de acierto puede dar una idea errónea sobre el desempeño global de los sistemas implementados. Para esto se usan medidas diagnósticas y de acuerdo con los resultados en las técnicas clásicas se puede observar que se tiene mejor rendimiento en la variable sensibilidad logrando niveles hasta 91.68% y en especificidad logrando niveles hasta el 51.33%. Cuando se usan modelos pre-entrenados para extracción de características, se puede obtener resultados de sensibilidad hasta 92.47% y de especificidad hasta el 70.63%. Lo que representa un incremento absoluto de alrededor de 20% en la especificidad, sin afectar otras medidas. Esto se puede notar al comparar sus resultados con los obtenidos a partir de las técnicas clásicas como se observó en las Fig. 5 y Fig. 6.

Otros artículos abordan un problema similar, pero con bases de datos diferentes. Se resalta que los resultados de este trabajo son comparables o están en el rango de los resultados reportados en esos otros artículos. Por lo tanto, los aportes que ofrecen los sistemas basados en aprendizaje profundo pueden utilizarse como una herramienta para ayudar al profesional a dar diagnósticos más precisos. Un trabajo futuro incluye el uso de arquitecturas más profundas con mayor cantidad de información, contando con un sistema de alta velocidad de procesamiento.

AGRADECIMIENTOS

Agradecemos a la Universidad Santiago de Cali, en particular al equipo de la facultad de Ingeniería por los recursos facilitados durante el desarrollo del presente trabajo.

REFERENCIAS

- [1] T. J. Key, P. K. Verkasalo & E. Banks, "Epidemiology of breast cancer," *Lancet Oncol*, vol. 2, no. 3, pp. 133–140, Mar. 2001. [https://doi.org/10.1016/S1470-2045\(00\)00254-0](https://doi.org/10.1016/S1470-2045(00)00254-0)
- [2] M. Akram, M. Iqbal, M. Daniyal & A. U. Khan, "Awareness and current knowledge of breast cancer," *Biol Res*, vol. 50, no. 1, pp. 1–23, Oct. 2017. <https://doi.org/10.1186/s40659-017-0140-9>
- [3] A. S. Y. Leong & A. Zhuang, "The changing role of pathology in breast cancer diagnosis and treatment," *Pathobiology*, vol. 78, no. 2, pp. 99–114, 2011. <https://doi.org/10.1159/000292644>
- [4] A. Bereciartua, "Desarrollo de algoritmos de procesamiento de imagen avanzado para interpretación de imágenes médicas. Aplicación a segmentación de hígado sobre imágenes de Resonancia Magnética multisequencia," *Doctoral dissertation*, UPV-EHU. Bilbao, ES, 2016. Disponible en <http://hdl.handle.net/10810/20703>
- [5] A. Mert, N. KJİJç, E. Bilgili & A. Akan, "Breast Cancer Detection with Reduced Feature Set," *Comput Math Methods Med*, pp. 1–12, 2014. <https://doi.org/10.1155/2015/265138>
- [6] C. M. Bishop, *Pattern recognition. Machine learning*. SG: Springer, 2006.
- [7] A. Cruz-Roa, A. Basavanahally, F. González, H. Gilmore, M. Feldman, S. Ganesan & A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," presented at *Medical Imaging 2014: Digital Pathology*, vol. 9041, SPIE, SD, CA, USA, 20 Mar. 2014, . <https://doi.org/10.1117/12.2043872>
- [8] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo & J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," *BMC Med Inform Decis Mak*, vol. 19, no. 1, pp. 1–17, Oct. 2019. <https://doi.org/10.1186/s12911-019-0913-x>
- [9] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre & A. M. Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," *Sensors*, vol. 20, no. 16, pp. 1–17, Aug. 2020. <https://doi.org/10.3390/s20164373>
- [10] M. A. Al-masni, M. A. Al-antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera & T. S. P. Kim, "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," *Comput Meth Prog Bio*, vol. 157, pp. 85–94, 2018. <https://doi.org/10.1016/j.cmpb.2018.01.017>

- [11] H. Chougrad, H. Zouaki & O. Alheyane, “Deep Convolutional Neural Networks for breast cancer screening,” *Comput Methods Programs Biomed*, pp. 19–30, 2018. <https://doi.org/10.1016/j.cmpb.2018.01.011>
- [12] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A Torre & A. Jemal, “Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 68, no. 6, pp. 394–424, 2018. <https://doi.org/10.3322/caac.21492>
- [13] R. Wainschenker, J. Massa & P. Tristan, Procesamiento digital de imagenes. (2011). *Clase Teórico Práctica N° 1*. Buenos Aires, Argentina: UNICEN. Obtenido de <https://users.exa.unicen.edu.ar/catedras/pdi/FILES/TE/CP1.pdf>
- [14] Á. Martínez, “Bases Teóricas, Digitalización y Análisis de Imágenes,” in *Manual de Telepatología*, L. Alfaro, M. García y A. Puras, ed., Pamplona, ES: Club de Informática Aplicada a la Sociedad Española de Anatomía Patológica, 2001. Disponible en <https://www.seap.es/manual-de-telepatologia>
- [15] J. C. Russ, *The image processing handbook*. BR, USA: CRC press, 2006. <https://doi.org/10.1201/9780203881095>
- [16] G. Mandloi, “A survey on feature extraction techniques for color images,” *IJCSIT*, vol. 5, no. 3, pp. 4615–4620, 2014. Available from <http://www.ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503424.pdf>
- [17] J. Ma & Y. Yuan, “Dimension reduction of image deep feature using PCA,” *J Vis Commun Image Represent*, vol. 63, no. 5, pp. 102578, 2019. <https://doi.org/10.1016/j.jvcir.2019.102578>
- [18] G. S. Fletcher, *Clinical epidemiology: the essentials*. FI, USA: LWW, 2019.
- [19] Z. H. Hoo, Z. H. Candlish & D. Teare, “What is an ROC curve?,” *Emerg Med J*, vol. 34, pp. 357–359, 2017. <https://doi.org/10.1136/emermed-2017-206735>
- [20] R. Xiao, X. Cui, H. Qiao, X. Zheng, Y. Zhang, C. Zhang & X. Liu, “Early diagnosis model of Alzheimer’s disease based on sparse logistic regression with the generalized elastic net,” *Biomed Signal Process Control*, vol. 66, no. 3, pp. 102362, 2020. <https://doi.org/10.1016/j.bspc.2020.102362>
- [21] I. Goodfellow, Y. Bengio & A. Courville, *Deep learning*. CA, MA, USA: MIT press, 2016.
- [22] M. Coccia, “Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence,” *Technol Soc*, no. 60, pp. 101198, Oct. 2019. <https://doi.org/10.1016/j.techsoc.2019.101198>
- [23] A. Janowczyk & A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *J Pathol Inform*, vol. 60, pp. 101198, 2016. <https://doi.org/10.4103/2153-3539.186902>
- [24] K. He, X. Zhang, S. Ren & J. Sun, “Deep residual learning for image recognition,” presented at Proceedings of the *IEEE conference on Computer Vision and Pattern Recognition*, CVPR, Las Vegas, NV, USA, 27-30 Jun. 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [25] K. Simonyan & A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv*, pp. 1–14, 2014. Available: <https://arxiv.org/abs/1409.1556>
- [26] Z. Zheng, J. Fu, C. Lu & Y. Zhu, “Research on rolling bearing fault diagnosis of small dataset based on a new optimal transfer learning network,” *Measurement*, vol. 177, pp. 109285, 2020. <https://doi.org/10.1016/J.MEASUREMENT.2021.109285>
- [27] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” presented at Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, Honolulu, HI, USA, 21-26 Jul. 2017, pp. 1251–1258. Available: <https://arxiv.org/abs/1610.02357>

Harold Agudelo Gaviria recibió su título de ingeniero electrónico de la Universidad Santiago de Cali (Colombia). Fue miembro del semillero en aplicaciones de electrónica de la facultad de ingeniería en la misma Universidad.

Milton Sarria-Paja recibió su título de Ingeniero Electrónico y Msc. En ingeniería de la Universidad Nacional de Colombia. En el año 2017 recibió su título de Doctor en Telecomunicaciones de la Universidad de Quebec (Canada). Sus áreas de interés se enmarcan en aplicaciones de técnicas de procesamiento digital de señales y aprendizaje de máquinas sobre señales de origen biomédico. <https://orcid.org/0000-0003-4288-1742>