**INFORME DE LA PRIMERA FASE DEL PROYECTO DE AULA**

**IDENTIFICACIÓN DE MASAS EN IMÁGENES MAMOGRÁFICAS, UTILIZANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL**

**Estudiantes**

**Kelly Johana Pimienta Clavijo**

**Darío Andrés Ramos Cañas**

**Docente**

**Tony Jiménez Márquez**

**Grupo:**

**01**

**Código**

**SS429-A**

**Universidad Popular del Cesar**

**Ingeniería de Sistemas**

**2021-1**

**INTRODUCCIÓN**

El cáncer de mama es la forma más común de cáncer entre las mujeres y es la principal causa de muerte entre los 35 a 55 años de edad y se estima que una de cada nueve mujeres desarrollará cáncer de mama alguna vez en su vida (Jemal et al, 2011). Actualmente no existe un método para prevenir el cáncer de mama, a menudo el tratamiento concluye con la extirpación de la mama. En Colombia, aunque la incidencia actual de cáncer de mama no es tan alta en comparación con Estados Unidos, es posible que crezca en los próximos años (Torres et al, 2009). No obstante, el cáncer de mama constituye una de las pocas enfermedades de su estilo que se pueden diagnosticar precozmente, antes de que aparezcan los primeros síntomas y que afectan, sobre todo, al tamaño y la forma de la mama.

Para el año 2007, el cáncer de mama se había constituido el tipo de cáncer con mayor número de muertes entre las mujeres a nivel mundial, además de ser el primero en número de nuevos casos al año (American Cancer Society, 2007). Esta enfermedad posee un alto potencial mortal, pero de ser detectada a tiempo puede tratarse evitándose el fallecimiento de la paciente en un alto porcentaje de casos. El proceso de detección consiste en identificar la clase de lesión que se encuentre en la mama, la cual puede ser una masa o una calcificación, mientras que el proceso de diagnóstico consiste en identificar en qué zona de la mama se encuentra la lesión e identificar si ésta es benigna o maligna.

Las microcalcificaciones son pequeños depósitos de calcio en el tejido mamario, indicadores de la presencia de cáncer de mama en el 60% de los casos; por ello, resulta muy importante reconocerlas cuando se desea detectar esta enfermedad tempranamente. Actualmente, la principal herramienta que existe para la detección de una microcalcificación, en etapas cuando ésta aún no es palpable, es la mamografía (Salvador, 2006). La lectura adecuada de la mamografía es determinante en este proceso de detección temprana, pero no resulta una tarea sencilla dada su alta dependencia de la experticia del especialista y la influencia de variables de índole humano que afectan al proceso de interpretación (cansancio, fatiga, entre otros).

A fin de incrementar la eficiencia y eficacia con la que se lleva a cabo el proceso de detección de la enfermedad, surge la propuesta de desarrollar un sistema de Identificación De Masas En Imágenes Mamográficas, Utilizando Técnicas De Inteligencia Artificial

**ESTADO DEL ARTE**

* **A Nivel Internacional**

A nivel internacional Aníbal Guerra, Joel Rivas en Venezuela en el (2011) se realizó Un software para la **Detección de microcalcificaciones en imágenes mamográficas usando redes neuronales,** el cual se constituyó como base de un mecanismo de segunda opinión en el proceso de detección de microcalcificaciones, a través del estudio de imágenes mamográficas. El software trabaja a partir de una mamografía digitalizada, la cual es procesada para ingresarla como dato de entrada a una Red Neuronal Artificial (RNA) del tipo perceptrón multicapas; ésta se encarga de detectar si la imagen presenta microcalcificaciones. La RNA se implementó en lenguaje C++, con una arquitectura de una capa oculta y el algoritmo de aprendizaje *Backpropagation*, en combinación con técnicas basadas en el análisis estadístico sobre la textura de imágenes. La *data*base para el desarrollo de la aplicación proviene de la base de datos de la Sociedad de Análisis de Imágenes Mamográficas (MIAS, en sus siglas en inglés). La efectividad alcanzada en la evaluación del software fue de 94.4% de aciertos en su predicción, mostrando así el potencial de la aplicación de ambas técnicas para el abordaje del problema planteado.

* **A Nivel Nacional**

A nivel nacional Celis Esteban, Sergio Augusto Sarmiento Ortiz, Jhoan Felipe en Bucaramanga Colombia en el (2020) se realizó un **Prototipo de sistema de información para la detección de carcinomas mamarios, basado en técnicas de inteligencia artificial,** el cual permitió optimizar la labor del cuerpo médico para detectar casos de cáncer y mitigar el riesgo de muerte y otros problemas que presentaban los pacientes. El desarrollo de la interfaz gráfica se realizó con Flutter Web y luego de ciertas pruebas, se definió que la opción conveniente para el sistema era la implementación de la arquitectura VGG16. La red neuronal, fue entrenada con un conjunto de imágenes diagnósticas mamográficas disponibles en la web, con el fin de adoptar un modelo, capaz de clasificar mamografías en cinco categorías: normal, microcalcificación benigna, nódulo benigno, microcalcificación y nódulo maligno.

* **A Nivel Local**

A nivel local [Gómez Betancur, Duván Alberto](https://repositorio.unal.edu.co/browse?type=author&value=G%C3%B3mez%20Betancur,%20Duv%C3%A1n%20Alberto) en Colombia en el (2012) se realizó un **Método de detección de distorsiones de la arquitectura de la glándula mamaria a partir de imágenes radiológicas**, el cual asistió a los especialistas en el proceso de decisión diagnóstica como segundo intérprete en el análisis de mamografías, mediante la integración de cuatro etapas principales: preprocesamiento, detección de regiones de interés que sean candidatas a la posible presencia de distorsión de la arquitectura de la glándula mamaria, extracción y selección de características de las regiones de interés detectadas y finalmente clasificación de esas regiones de interés con base en las características extraídas de las mismas. El método propuesto se valida mediante el análisis de imágenes mamográficas de la base de datos DDSM, logrando valores de precisión general hasta de un 90.7% lo cual lo convierte en una base importante en la búsqueda de la reducción del alto número de diagnósticos errados que conducen a las altas tasas de morbilidad por cáncer de mama que se presentan en el mundo.

**PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA**

**Morbilidad y Mortalidad del Cáncer de Mama.**

Según el Observatorio Global de Cáncer (Global Cancer Observatory, 2018) en su informe publicado el 12 de septiembre del 2018, el tipo de cáncer más frecuente en las mujeres de todo el mundo es el cáncer de mama con un total de 2’088.849 casos registrados. Aproximadamente uno de cada cuatro nuevos casos de cáncer diagnosticados en las mujeres es cáncer de mama, además, es la principal causa de muerte por cáncer en mujeres (626.679). (International for Research on Cancer, 2018). Según la CAC en Colombia, hasta el 01 de enero de 2018, se registraron 59.837 casos de cáncer de mama en el SGSSS. Específicamente, para el periodo comprendido entre el 02 de enero de 2017 y el 01 de enero de 2018, se reportaron 6.348 nuevos casos de cáncer de seno, de los cuales 303 (4.8%) corresponden a casos de carcinoma in situ, y 6.045 (95.2%) fueron casos de cáncer de mama invasivo, lo que representa un aumento del 37% en los registros, con relación al periodo anterior (2016 - 2017). Además, para este mismo periodo, el total de pacientes fallecidos fue 2.522. (Fondo Colombiano de Enfermedades de Alto Costo Cuenta de Alto Costo[CAC], 2018) Sin embargo, en el GLOBOCAN 2018, los últimos datos mundiales sobre el cáncer, las cifras para Colombia varían significativamente, donde los casos nuevos reportados fueron 13.380 y el número de muertes 3.702. (Global Cancer Observatory, 2018).

**Examen Mamográfico.**

Normalmente, la detección de carcinomas mamarios se ve obstaculizada por la complejidad que conlleva el análisis de mamografías realizado por los radiólogos en su labor. Estos análisis no cuentan con resultados muy precisos y la incertidumbre aumenta en ocasiones por diversas causas: en primer lugar, los procesos analíticos de las mamografías realizados por los expertos son hechos de manera meramente visual, esto genera que se obtengan diferentes opiniones en torno a una observación subjetiva (Radiological Society of North America, 2019) de parte de cada profesional.

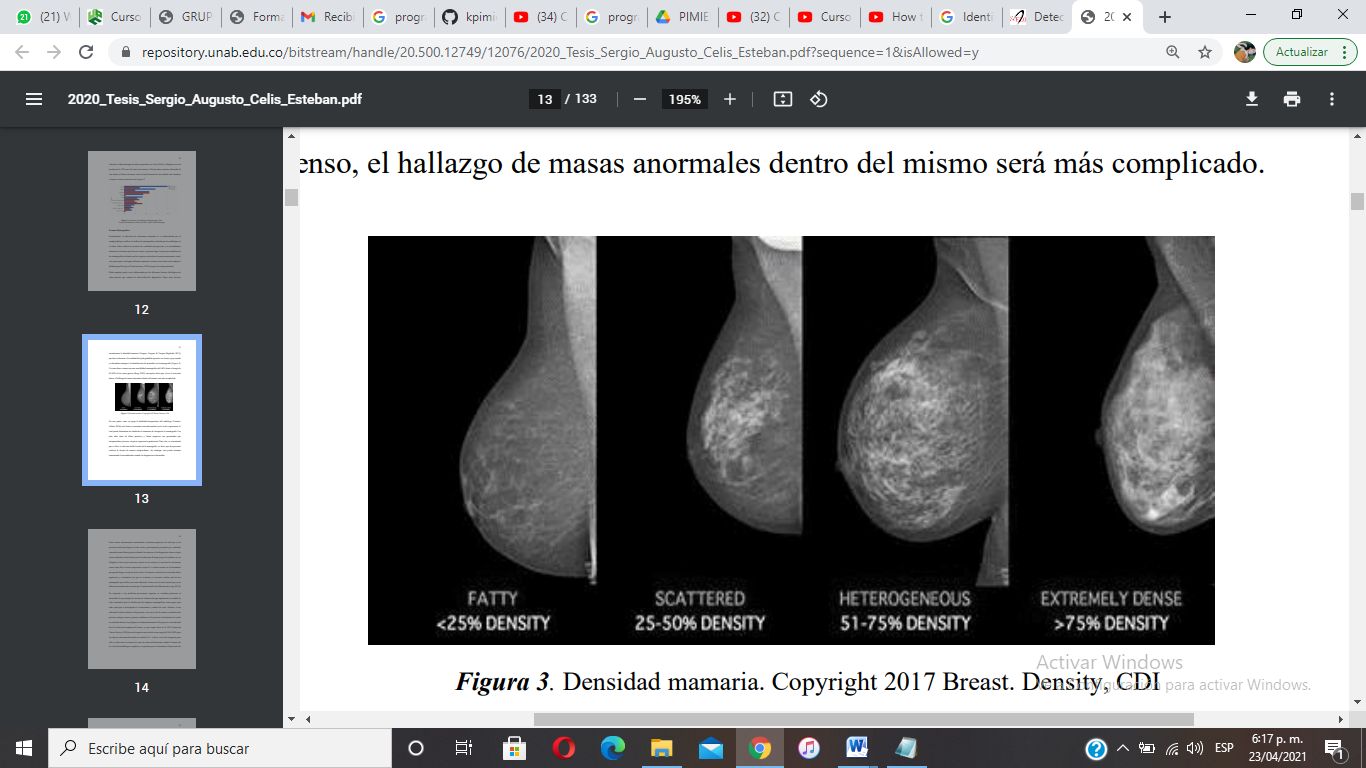
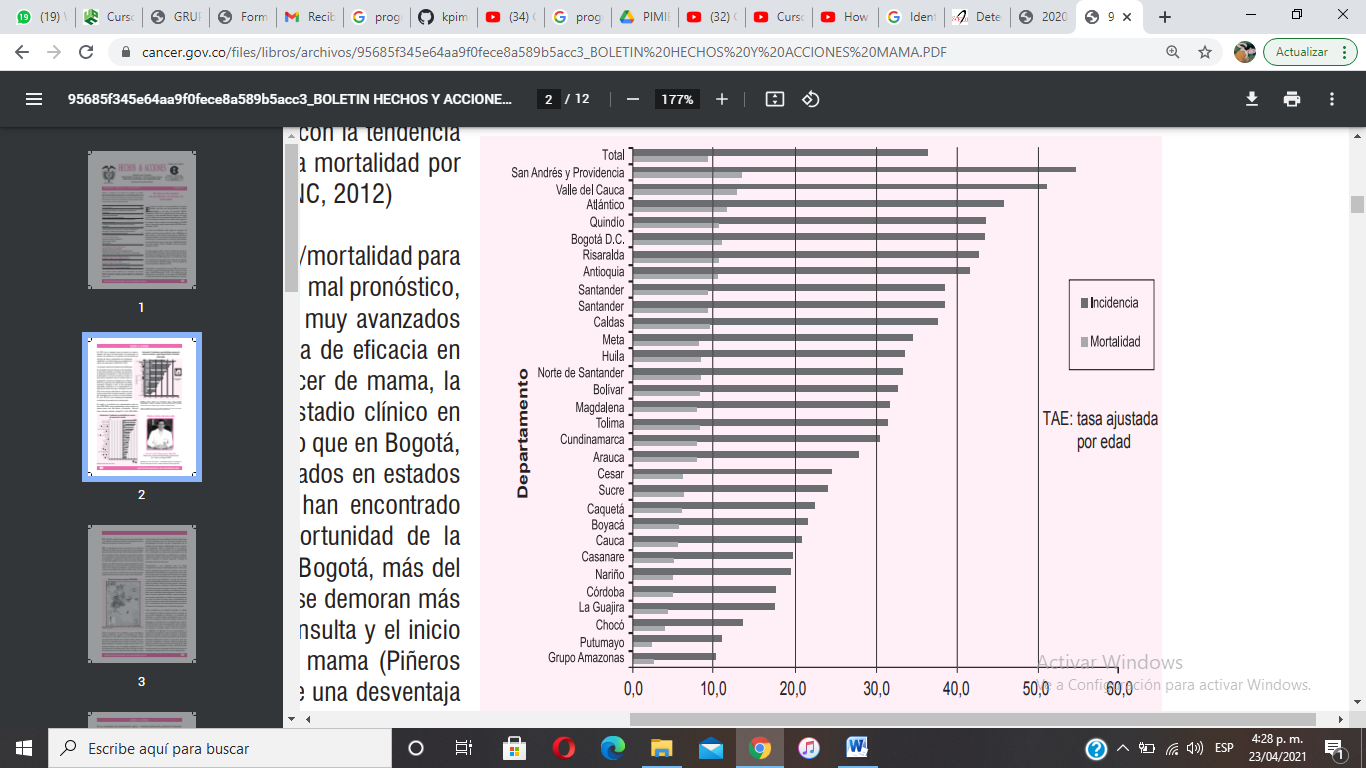
Dicha opinión, puede verse influenciada por los diferentes factores fisiológicos de cada paciente que reducen la efectividad del diagnóstico. Entre estos factores encontramos la densidad mamaria (Vergara, Vergara, & Vergara Dagobeth, 2013), que hace referencia a la cantidad de tejido glandular presente en el seno y que cuando es abundante entorpece la identificación de anomalías en la mamografía (Figura 1). Un seno denso cuenta con una sensibilidad mamográfica del 60% frente al rango de 85-95% de los senos grasos (Berg, 2019), esto quiere decir que, al ser el seno más denso, el hallazgo de masas anormales dentro del mismo será más complicado.

Figura 1. Densidad mamaria. Copyright 2017 Breast. Density, CDI

En este punto, entra en juego la habilidad interpretativa del radiólogo, este factor se encuentra asociado muchas veces con la experiencia, lo cual puede determinar un obstáculo al momento de interpretar la mamografía. Las más altas tasas de falsos positivos y falsos negativos son presentadas por interpretadores jóvenes con poca experiencia profesional. Para esto, se recomienda que se lleve a cabo una doble lectura de la mamografía, es decir, que dos personas realicen la lectura de manera independiente, sin embargo, esto puede terminar aumentando la incertidumbre cuando los diagnósticos discuerdan.

Estas causas anteriormente mencionadas, ocasionan perjuicios de todo tipo en la paciente; desde psicológicos (como estrés y preocupación generados por resultados conocidos como falsos positivos donde a las mujeres se les diagnostica cáncer sin que exista realmente), hasta físicos (con la realización de biopsias que los médicos se ven obligados a hacer para tener una certeza en su examen, la creación de carcinomas como causa de la excesiva exposición a rayos X, o incluso retrasos en el tratamiento que pueden llegar, en el peor de los casos, a la muerte, a causa de los conocidos falsos negativos), y económicos (ya que en ocasiones es necesario realizar más de una mamografía que conlleva un costo adicional, o hacer uso de otras técnicas que en su defecto son mucho más costosas que el examen tradicional (Breastcancer.org, 2013)).

**Ilustración: Incidencia y mortalidad por cáncer de mama en mujeres, según departamento,Colombia, 2002-2006.**



Con el fin de apoyar a los especialistas en el proceso de decisión diagnóstica como segundo intérprete en el análisis de mamografías y optimizar la labor del cuerpo médico para detectar casos de cáncer y mitigar el riesgo de muerte y otros problemas que se presentan en los pacientes, este proyecto busca la implementación de un sistema de Identificación De Masas En Imágenes Mamográficas, Utilizando Técnicas De Inteligencia Artificial.

**OBJETIVOS**

**OBJETIVO GENERAL**

* Desarrollar un prototipo para la Identificación de Masas en Imágenes Mamográficas, Utilizando Técnicas de Inteligencia Artificial.

**OBJETIVOS ESPECIFICOS**

* Identificar los parámetros representativos para la Identificación de Masas en Imágenes Mamográficas, Utilizando Técnicas de Inteligencia Artificial con el fin de ser usados como entradas para la red neuronal
* Definir la arquitectura y el tipo de red neuronal artificial
* Realizar el entrenamiento de la red con un conjunto de datos significativos designados para esta tarea.
* Analizar los resultados de la aplicación de diferentes métodos entrenamiento de redes neuronales artificiales y verificar la confiabilidad del modelo.

**METODOLOGIA**

Los componentes a utilizar para el desarrollo del Frontend y del Backend se describirán a continuación y se hará referencia, a la información que se obtuvo de la revisión de literatura en proyectos aplicados al área de medicina. Se tomó como partida esta revisión y se procedió a hacer la respectiva selección.

**Tecnologías**

Es importante tener claro en que entorno se desarrollará el modelo. Para esto, se hizo un respectivo análisis de los lenguajes de programación para proyectos con Machine Learning, en el cual se revisaron las ventajas y desventajas de cada uno de ellos.

**Dataset**

Para el desarrollo del proyecto se creó un banco de datos propio con imágenes de mamografías digitales tomadas de internet, donde hubiese diferentes anomalías detectadas en estas imágenes, también un apoyo de la base de datos de la universidad de florida DDSM, en el cual se tomará una cantidad de imágenes para el proceso de entrenamiento y se dejara cierto porcentaje bajo para el proceso de simulación de la red. Luego de tomar estas imágenes se debe hacer un preprocesamiento o adaptación de estas imágenes, las imágenes deben de ser convertidas mediante un software de diseño gráfico en este caso se utilizará *Corel Draw* para la preparación de las imágenes, convertirlas a mapa de bits, escala de grises; este proceso debe ser llevado a cabo con cada una de las imágenes antes de ser presentadas a la red.

Culminada la etapa de preprocesamiento se debe adaptar los datos para que el modelo los reconozca y pueda usarlos.

**REFERENCIAS BIBLIOGRAFICAS**

**http://ve.scielo.org/scielo.php?script=sci\_arttext&pid=S0798-40652011000300002**

**https://repository.unab.edu.co/bitstream/handle/20.500.12749/12076/2020\_Tesis\_Sergio\_Augusto\_Celis\_Esteban.pdf?sequence=1&isAllowed=y**

**https://repositorio.unal.edu.co/handle/unal/10236**

**https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM**