

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**SISTEMA INTELIGENTE PARA APOYAR AL ANÁLISIS
MAMOGRÁFICO EN LA DETECCIÓN DE TUMORES DE MAMA
FEMENINO ENTRE LAS EDADES DE 40 A 60 AÑOS EN EL
HOSPITAL “LAS MERCEDES” DE CHICLAYO**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

JOSE CRISTHIAN RAMON CUEVA

ASESOR

MARIANA CHAVARRY CHANKAY

<https://orcid.org/0000-0001-5136-7177>

Chiclayo, 2021

**SISTEMA INTELIGENTE PARA APOYAR AL ANÁLISIS
MAMOGRAFICO EN LA DETECCIÓN DE TUMORES DE
MAMA FEMENINO ENTRE LAS EDADES DE 40 A 60 AÑOS
EN EL HOSPITAL “LAS MERCEDES” DE CHICLAYO**

PRESENTADA POR
JOSE CRISTHIAN RAMON CUEVA

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Roger Ernesto Alarcón García
PRESIDENTE

Gregorio Manuel León Tenorio
SECRETARIO

Mariana Chavarry Chankay
VOCAL

Dedicatoria

A Dios, por permitirme vivir grandiosos momentos y seguir haciéndolo hasta ahora

A mis padres Carmen y Jose, por el apoyo incondicional y por estar siempre en cada logro obtenido

A mi futura esposa Nathaly, quien ha estado en toda mi formación académica profesional apoyándome siempre en los buenos y malos momentos

Agradecimientos

A mi asesora de tesis Mariana Chavarry, por el apoyo constante en el desarrollo de esta investigación

A mis compañeros, por brindarme todo su apoyo incondicional

Al doctor Calderón, por apoyarme y brindarme toda la información necesaria para el desarrollo de este trabajo de investigación

Índice

Resumen	5
Abstract	6
Introducción.....	7
Revisión de literatura.....	9
Materiales y métodos	13
Resultados y discusión	15
Conclusiones	32
Recomendaciones	33
Referencias.....	34
Anexos	36

Resumen

La detección de tumores de mama es crucial para evitar las muertes a causa de esta enfermedad, los hospitales públicos padecen de este problema y cuyos centros de atención evidencian la carencia de un diagnóstico rápido y eficiente. Esta situación es la que ha permitido canalizar la investigación aplicada con el objetivo de apoyar al análisis mamográfico en la detección de tumores de mama mediante el uso de mamografías digitales. Para la obtención de las imágenes se vio la necesidad de buscar bases de datos públicas de las cuales se trabajaron con mini-MIAS y DDSM. El sistema inteligente ha sido desarrollado por medio de una fusión entre las Etapas del modelo de Machine Learning y RUP. La primera metodología se utilizó para la creación de un modelo haciendo uso de las técnicas de Deep learning, y dentro de ella se trabajaron con las redes neuronales convolucionales (CNN). la segunda metodología se utilizó para el desarrollo del sistema web local que sirve como una interfaz de interacción del experto con el modelo creado. Al finalizar el desarrollo del modelo se propuso evaluar su precisión, por lo cual se ha considerado la validación del experto con respecto a los resultados que retorna el sistema, concluyendo en una exactitud y precisión del modelo en un 88.5% y 93.5% respectivamente. Obteniendo estos resultados finales se puede decir que, haciendo uso de la inteligencia artificial, esta investigación es un aporte crucial en la ciencia de la salud.

Palabras clave: cáncer de mama, Deep Learning, mamografías digitales, CNN, Machine Learning

Abstract

The detection of breast tumors is crucial to prevent deaths due to this disease, public hospitals suffer from this problem and whose care centers show the lack of a fast and efficient diagnosis. This situation is the one that has allowed channeling applied research with the aim of supporting mammographic analysis in the detection of breast tumors through the use of digital mammography. In order to obtain the images, it was necessary to search public databases from which mini-MIAS and DDSM were used. The intelligent system has been developed by means of a fusion between the Machine Learning and RUP model stages. The first methodology was used for the creation of a model making use of Deep learning techniques, and within it convolutional neural networks (CNN) were worked with. The second methodology was used for the development of the local web system that serves as an interface of interaction of the expert with the created model. At the end of the development of the model it was proposed to evaluate its accuracy, for which the validation of the expert with respect to the results returned by the system has been considered, concluding in an accuracy and precision of the model in 88.5% and 93.5% respectively. Obtaining these final results, it can be said that, making use of artificial intelligence, this research is a crucial contribution to health science.

Keywords: breast cancer, Deep Learning, digital mammograms, CNN, Machine Learning

Introducción

El cáncer de mama es una enfermedad perniciosa que no excluye género, incluso puede afectar a la población masculina, en la actualidad se ha convertido en el afección más común entre las mujeres; llegando a convertirse en la enfermedad más frecuente a nivel mundial. Comparando los indicios de años anteriores podemos comprobar que el cáncer de mama no solo se ha incrementado, sino que, las detecciones comunes de esta enfermedad son en etapa avanzada. En el 2008 se detectó más de 1.4 millones de mujeres con Neoplasia mamaria en todo el mundo, cuya enfermedad comienza en el tejido mamario que se convierte en tumor maligno o benigno; esta situación preocupó a los médicos que informaron sobre el incremento de los casos al 100 % a comparación de los años anteriores como 1975 [1].

Según la Organización Mundial de la Salud (OMS) y la Organización Panamericana de la Salud (OPS) [2], señala que cada año en América se reporta 462 000 casos de cáncer de mama en mujeres, llegando a una cifra de 100 000 muertes originadas por esta enfermedad. Prospectivamente se prevé que el cáncer de mama aumentará un 34% en el 2030, debido a que los países de bajos recursos detectan el cáncer en su etapa avanzada, mientras que los países desarrollados logran diagnosticarlo en una etapa precoz. Así mismo, se indica que las muertes a causa de esta enfermedad en América latina y el Caribe, es detectada en la población de mujeres menores de 65 años.

Por consiguiente, según la Sociedad Americana de Oncología Clínica (ASCO) [3], señala que en el 2021 Estado Unidos prescribirá con cáncer de mama invasivo a 281 550 mujeres y cáncer de mama no invasivo a 49 290. Así mismo, se estima 44 130 muertes donde 43 600 mujeres y 530 hombres a raíz de esta enfermedad.

En el Perú el Ministerio de Salud (MINSA) afirmó que el número de médicos radiólogos que participan en la lectura de mamografía de tamizaje es de 113, de los cuales solo 32 estaban capacitados en mamografía; esto quiere decir que el número existente de radiólogos capacitados son pocos, generando que los análisis mamográficos realizados no sean muy exactos. La ausencia del control sobre las lecturas mamográficas dificulta el poder identificar los niveles de categorías en las que se encuentra el cáncer de mama [4]. Así mismo, el informe realizado por recursos humanos de MINSA informó que, en marzo del 2016, 62 cirujanos oncólogos en 8 regiones son elegibles para participar en cirugía de cáncer de mama, de los cuales 40 de ellos se encuentran en Lima, 6 en Callao, 5 en la Libertad, 4 en Arequipa, 2 en Cusco y Piura, 1 en las regiones de Ica y Junín; sin embargo, en las regiones de Lambayeque y Ucayali no existen especialista alguno.

El cálculo realizado por el Centros para el Control y Prevención de Enfermedades (CDC) indica que Lambayeque con un 11.9% ocupa el cuarto lugar con respecto a la tasa de mortalidad por cáncer de mama, lo cual se ve reflejado en el número de pacientes que reciben los hospitales públicos de la región, tales son: Regional y Las Mercedes, siendo éstos los únicos que cuentan con equipos y especialistas en cáncer en Lambayeque. Cabe destacar que el hospital con más número de recurrencia de pacientes de diversas provincias es “Las Mercedes”, pues es aquí donde se ofrece las evaluaciones y el descarte de cáncer de mama que son remitidas por los especialistas de Lima, debido a que cuentan con tecnología avanzada para realizar dichas evaluaciones, teniendo una espera mínima de 21 días o 3 meses como máximo para obtener los resultados, luego de acuerdo al diagnóstico son trasladados al Hospital Regional.

El área de oncología cuenta con un solo radiólogo que se encarga de analizar la mamografía superficialmente y dos oncólogos que asisten a los pacientes en el único turno de la mañana. La capacidad de atención es de 20 pacientes por día, lo cual sería diferente si el hospital contara con una aplicación que ayudara al radiólogo en las evoluciones de análisis mamográfico. Teniendo en cuenta de cómo las enfermedades arremeten rápidamente hacia el ser humano, la ciencia se encuentra a la vanguardia de las formas y estrategias, de controlar y contrarrestar dichas enfermedades; actualmente se ha encontrado un formato para diagnosticar el cáncer a través de las redes neurales, las cuales proporcionan resultados inmediatos y confiables. El área de la medicina considera que las redes artificiales poseen confiabilidad en cuanto a aplicación, monitoreo, verificación y velocidad para ofrecer una pronta respuesta que puede ayudar a contrarrestar esta enfermedad [5].

Así mismo, podemos darnos cuenta del gran problema que engloba el cáncer en cada nivel de localización, es por ello la importancia de formularse la siguiente pregunta ¿De qué manera se puede apoyar al análisis mamográfico automatizado en la detección de tumores de mama femenino entre las edades de 40 a 60 años en el hospital “Las Mercedes” de Chiclayo? Ante esta interrogante y la necesidad de ahondar en el problema, se llevará a cabo la investigación de tipo aplicada, dirigida a los pacientes pertenecientes al hospital Las Mercedes de Chiclayo, de las cuales se tomarán sus mamografías con la cuestión de apoyar al análisis mamográfico y así brindar una mejor detección temprana de tumores de mama.

Por esta necesidad, surgió el interés de proponer la implementación de un sistema inteligente con la finalidad de apoyar al radiólogo en el análisis mamográfico para la detección oportuna de tumores de mama y además de ayudar a mujeres con este mal que les aqueja. Para este plan se propone una serie de objetivos específicos que ayudarán a cumplir con el objetivo general, primero se determina las características asociadas a las imágenes mamográficas para detectar tumores de mama, luego crear un modelo computacional en base a las mamografías digitales; a través de ello, validar la precisión del modelo en la detección de tumores de mama y por último diseñar una aplicación para la detección de tumores de mama en base a las mamografías digitales que será entregado al personal especializado.

Así mismo, esta investigación se justifica desde el ámbito científico ya que mediante el desarrollo de un sistema inteligente apoyará al análisis mamográfico en la detección oportuna de tumores de mama generando así una idea innovadora que contribuye a la ciencia de la salud. Desde el ámbito financiero se justifica que este sistema no es muy costoso comparando con otros sistemas que están en el mercado tecnológico, además reducirá el tiempo de espera de un diagnóstico oportuno y temprano del cáncer de mama, obteniendo así más pacientes atendidos que ocasionará un alto prestigio a la institución. Continuando con el ámbito social se justifica ya que pretende apoyar al análisis mamográfico, donde se tendrá un margen mínimo de error y además de ayudar al área de oncología al tener un diagnóstico oportuno. También este sistema beneficiará indirectamente a los estudiantes de la carrera de medicina, al igual como los especialistas encargados de esta área. Por último, esta investigación se justifica tecnológicamente ya que se hará uso de tecnologías presentes tanto como software y hardware que permitirá realizar este sistema inteligente con la finalidad de apoyar al radiólogo con el análisis mamográfico.

Revisión de literatura

Para el desarrollo de esta investigación se recopiló información de diferentes fuentes que sirvieron para el sustento de la solución propuesta. En primer lugar, se detallan algunas investigaciones que sirven de ayuda para ampliar el conocimiento de lo que se pretende realizar. Pérez [6], propone resolver la problemática para detectar las lesiones cutáneas hacia la detección del melanoma, organizado por la colaboración internacional de imágenes de la piel. Por lo que, se aplicó la metodología de transferencia de aprendizaje para entrenar el modelo con las redes pre entrenadas AlexNet, VGG y ResNet, debido que, al emplear un entrenamiento desde cero tomaría mucho tiempo y para ese entonces no se contó con el recurso de hardware disponible. Asimismo, se implementó un sistema utilizando técnicas de Deep Learning basándose en redes neuronales convolucionales, logrando solucionar la problemática sobre la detección de melanoma en la piel. Para el entrenamiento se usó un dataset que consta de 10000 imágenes de manchas correspondientes al cáncer de piel, donde cada formato es en JPG con un tamaño de 600x450px. Para el desarrollo del entrenamiento, se utilizó la plataforma de código abierto Pytorch, ya que se puede utilizar todas las librerías de Tensorflow o keras. Esta investigación consideró como base para la división de las imágenes, tanto para el entrenamiento y testeo, al igual que el tamaño de las imágenes para esta investigación. Chanampe et al [7], relata la problemática que se vive en Argentina sobre el cáncer de seno, que es la causa más importante de decesos en mujeres mayores. Los autores al tener este problema se apuntan a utilizar técnicas Deep Learning que sirven para la categorización de distintos tipos de lesiones en ecografías mamarias. Para el entrenamiento de las ecografías se usaron las redes neuronales convolucionales profundas, permitiendo establecer un modelo de clasificación. Así mismo el modelo generado utilizó una red pre entrenada llamada ResNet ya que por medio una tabla de comparación es la que tiene menos índice de error del 3.57%. Respecto a la base de datos se trabajó con un mini-MIAS que cuenta con 322 imágenes que tienen una resolución de 1024x1024. Para la fragmentación de la base de datos se dividió en datos de entrenamiento, datos de validación y datos de prueba con sus respectivos porcentajes 80%, 10% y 10%. Finalmente, el autor concluyó con el desarrollo del modelo obteniendo una precisión del 62%, debido a la poca base de datos que posee el autor. Esta investigación tiene relación, debido a que se trabaja con imágenes mamográficas y se utilizó una red-pre entrenada para la elaboración del modelo. Y como recomendaciones, el autor menciona la adquisición de una gran cantidad de base de datos para alcanzar una alta precisión. Teniendo esto en claro se trató de buscar una gran cantidad de imágenes mamográficas para esta investigación. Por otro lado, García [8], relata la problemática que presenta la región de Loreto, donde más de una década el servicio de salud brindado es el más precario. La región presenta un incremento de enfermedades crónicas relacionadas con la alimentación, tal como: la hipertensión, diabetes, hígado graso y otras asociadas al riñón. Por lo que el aumento de los casos de insuficiencia renal en varios grupos es cada vez más frecuente, así como la escasa accesibilidad a hospitales y centros de salud. Ante esto, el autor propuso desarrollar un Sistema Inteligente para el diagnóstico previo de enfermedades renales a través de patrones de imágenes de muestras de orina de adultos, que posteriormente se convirtieron en datos de entrada para el entrenamiento de una red neuronal convolucional. Con esta propuesta se pudo dar apoyo a los diversos procesos de análisis médico en adultos de la ciudad de Iquitos y por medio del uso de dispositivos móviles los procesos también fueron realizados por los propios pacientes ya sea en casa o en establecimientos de salud. Finalmente, el autor concluyó que el desarrollo del sistema inteligente móvil permitió identificar dos tipos de resultados del diagnóstico previo: “Enfermo” y “Sano”; a partir del reconocimiento de imágenes de pruebas de orina capturadas desde un celular inteligente. Adema de haber logrado una eficacia del 97% y una sensibilidad mayor del 98% sobre el diagnóstico. La relación que se tiene con esta investigación, es que se realizó un sistema móvil

tomando en consideración este sistema ya que en esta investigación se desarrolló un sistema para web utilizando algunos recursos tecnológicos de esta investigación. Otra investigación por parte de Yacolca [9], abordó la problemática sobre el cáncer mama que según la OMS estimó que 627,000 mujeres murieron de este mal en el 2018. Por tal motivo se desarrolló un algoritmo que consiga identificar el tejido mamario basado en el criterio de atipia nuclear, para luego ser integrado como un módulo dentro de un sistema por computador según la escala de Nottingham. Para la construcción del algoritmo se utilizó una primera base de datos de imágenes BreCaHad, que contiene 162 imágenes histopatológicas del cáncer de mama, las cuales tienen las siguientes anotaciones: mitosis, núcleos no tumorales, apoptosis, núcleos tumorales, túbulos y no túbulos. Para esta investigación solo se utilizaron las anotaciones de núcleos tumorales. La segunda base de datos que se empleó es MITOS-ATYPIA 2014, que contiene imágenes correspondientes a HPF (campos de alta potencia), con anotaciones de mitosis y atipia nuclear. Además, para el entrenamiento se empleó las redes neuronales convolucionales (CNN) encargadas de identificar la región donde hay presencia de atipia nuclear. Por último, el autor recapitulo que con el desarrollo de este algoritmo se logró automatizar el trabajo realizado por el patólogo y facilitar el diagnóstico emitido. Además, hace énfasis que es necesario aplicar una segmentación semántica con el fin de que la red aprenda de forma más eficiente, ya que solo se centrar en partes más específicas del núcleo. Esta investigación se escogió por la razón de que utiliza tejidos mamarios para la identificación de atipia nuclear, que en otros aspectos es un entrenamiento más profundo respecto a mi investigación que ayudará en la lógica del algoritmo de mi trabajo de investigación. Y por último Quesquen [10], realizó un estudio en el hospital EsSalud de la ciudad Chiclayo, que cuenta con radiografía computarizada y una tecnología vigente llamada tomosíntesis del seno, ésta última genera mamografías tridimensionales o varias imágenes de cada seno. Es por ello que, el autor propone apoyar en la obtención de imágenes digitales de mayor resolución con formatos diferentes, ya que las microcalcificaciones son de tamaño submilimétricas. Se aplicó técnicas algorítmicas para la identificación automática de microcalcificaciones de mamografías digitales que mediante la segmentación que consiste en identificar cada uno de los objetos presentes de microcalcificaciones. Y como valor agregado en esta investigación es la construcción de una red neuronal que determina si las microcalcificaciones son benignas o malignas. En conclusión, el autor logró desarrollar un data set de entrenamiento con 9 imágenes benignas junto con 9 sospechosas y un data set para el testeo con 6 imágenes sospechosas, así como 6 benignas, así mismo se utilizó el software orientado a matrices de Matlab para el procesamiento de las imágenes digitales. También se logró desarrollar un sistema con una sensibilidad del 83%, una especificidad del 100%, una exactitud del 91%, una precisión del 100% y una tasa de error del 8.3% utilizando un framework de java Neuroph Studio.

Asimismo, para una mejor comprensión del cáncer de mama [11], se define como una enfermedad provocada por un tumor maligno que ocurre en la mayoría de población femenina, que es originada por células cancerígenas alojadas en el seno tendiendo a crecer y reproducirse de forma rápida. Existen 3 tipos de cáncer de mama, tales son [11]: según el grado de invasión del tumor, según el origen del tumor y las células que lo forman, según sus características biológicas y genéticas de las células. Para profundizar más sobre el tema se tuvieron en cuenta las categorías de la densidad que presenta los senos, por lo cual [12], lo define como la estructura del tejido mamario donde el epitelio y estroma mamario se atenúan más en los rayos X en comparación con la grasa, debido a esto aparece blancos en la mamografía, mientras que la grasa se ve oscura. Existen 4 categorías de densidad de los senos, tales son [13]: Predominante grasa, densidades fibroglandulares dispersas, heterogéneamente densa y extremadamente denso, estas categorías se observa en la Figura.1. Esta última categoría se presenta en un 10% en la población de mujeres y también es una de las categorías que más

complicaciones tiene el radiólogo al momento de detectar los tumores de mama. Respecto al sistema de evaluación BI-RADS [14], se define como un método, el cual permite categorizar los hallazgos mamográficos. Este es considerado un idioma común en el diagnóstico de enfermedades mamarias y tiene como objetivo estandarizar la terminología, sistematizar el informe mamográfico y clasificar las lesiones según el grado de sospecha. Por el lado de la mamografía [15], se determina como una radiografía centrada en todo el seno el cual abarca toda la extensión yuxta-axilar, axilar y planos profundos pre torácicos. Actualmente, es la única prueba para detectar el cáncer de mama temprano en mujeres asintomáticas y encargada de distinguir las áreas patológicas entre los tejidos normales. Existen 2 métodos para la obtención de la imagen mamográfica: mamografía convencional y la mamografía digital.

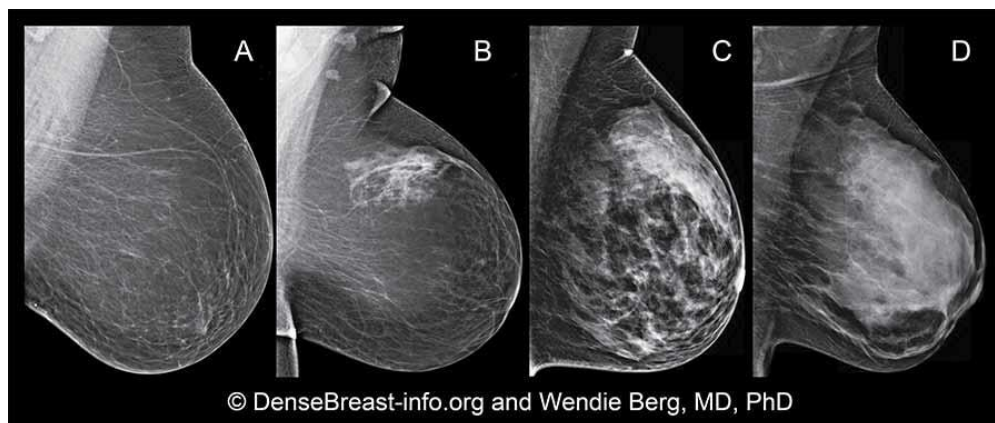


Figura 1. Las cuatro categorías de densidad del seno [13]

Por otro lado, ingresando al tema de inteligencia artificial y sus respectivos campos se visualiza en la figura 2, Francois Chollet [16], lo explica como el esfuerzo por automatizar las tareas intelectuales normalmente realizadas por los humanos. Es por ello que este tema tiene relación con la creación de un sistema inteligente que apoya en el diagnóstico de enfermedades humanas. Dentro de este campo se encuentra el Machine Learning o aprendizaje automático que es definido por [17], donde hace mención que este campo se encuentra en la intersección de la informática, la ingeniería y estadística que en conclusión es un sistema que esta entrenado en lugar de programado explícitamente. Existen 3 categorías en el aprendizaje automático, que son [18]: el aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por reforzamiento. En este sentido el modelo desarrollado se encuentra en el primer aprendizaje ya que se establecieron grupos ya clasificados para el aprendizaje. Para ir aún más profundo dentro del campo de la inteligencia artificial, nos encontramos con Deep learning o aprendizaje profundo, [16] aporta, como un subcampo específico del aprendizaje automático que por medio de una visión de representaciones por capas va aprendiendo a través de modelos llamados redes neuronales. Estos modelos se clasifican en tres [19]: redes neuronales recurrentes, redes neuronales generativas y redes neuronales convolucionales.

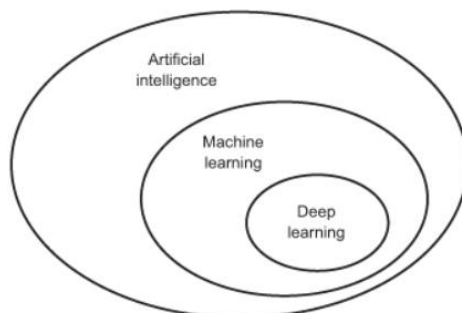


Figura 2. Inteligencia artificial, aprendizaje automático y aprendizaje profundo [16]

Tomando esta base de teorías y para entrar al proceso de desarrollo del modelo, se define las redes neuronales convolucionales por [20], que menciona que estas redes son muy similares a las redes neuronales ordinarias; que se componen de neuronas que tiene pesos y sesgos que pueden aprender. Cada neurona recibe como entrada una imagen y que por medio de una transformación genera una serie de capas ocultas. Cada capa oculta está formada por un conjunto de neuronas, donde cada neurona está completamente conectada a todas las neuronas de la capa anterior. La última capa totalmente conectada llamada “capa de salida”, obtiene el los resultados de las características extraídas de la imagen. En cuanto a su arquitectura de esta red se compone por las siguientes capas [21]: capa convolucional, se basa como la parte del cerebro donde se procesa la información visual usando una matriz de pesos llamados filtros que detectan atributos específicos como bordes diagonales, bordes verticales, etc. Función de activación ReLU [22], permite que la red neuronal tenga en cuenta las relaciones no lineales, es decir, que tenga en cuenta situaciones en las que la relación entre las entradas de valor de pixel y la salida de una red convolucional no sea lineal. Y la última capa llama MaxPooling [23], se utiliza para reducir agresivamente la dimensionalidad de los mapas de características y afinar las características ubicadas, a veces se debe insertar una capa de agrupación máxima después de una capa convolucional. Con la idea esencial de dividir cada capa en igual tamaño.

Materiales y métodos

Para este trabajo de investigación se aplicó un tipo de investigación aplicada, puesto que se realizó un aplicativo web local haciendo uso de un modelo basado en redes neuronales convolucionales que apoya al análisis mamográfico en la detección de tumores de mama y un nivel de investigación pre experimental, porque se ejerce un bajo control de las variables y tampoco se asignan grupos aleatorios para el experimento. De acuerdo a los métodos de investigación se utilizaron: el analítico que se hace un estudio del problema que se presenta el área de oncología sobre el cáncer de mama en el hospital las Mercedes, el deductivo que se establece una propuesta de solución al problema que se tiene sobre la demora de obtener un diagnóstico rápido de los análisis mamográficos, análisis de la literatura que se encargó de la búsqueda de antecedentes y bases teóricas que argumenten a la investigación, experimento que se encargó de entrenar del modelo de machine learning con la data de mamografías digitales y la implementación que se sirve para la ejecución de un sistema inteligente. Entre las técnicas aplicadas para determinar las características asociadas a las imágenes mamográficas se utilizó el análisis documental y entrevista al experto con el fin extraer información importante para la base de conocimiento a través de una guía de análisis documental y una guía de entrevista. También se usaron las técnicas de redes neuronales para crear un modelo computacional con el fin de detectar los tumores de mama en base a la adquisición de mamografías digitales que fueron utilizadas para el entrenamiento del modelo. Como parte importante también se aplicó un juicio de expertos al radiólogo a través de una lista de cotejo para verificar la validez de los resultados del modelo con respecto al análisis que hace el especialista.

Como parte de los procedimientos se utilizaron dos metodologías: Las etapas del modelo de Machine Learning que sirvió para la creación de un modelo computacional de procesamiento de imágenes y la metodología RUP para el desarrollo del sistema inteligente. La primera metodología se abarcan todas las etapas y dentro de cada una de ellas se aplicó la segunda metodología con el fin de desarrollar a la par el modelo junto con el sistema. Dentro de cada una de ellas se implementaron las siguientes etapas: etapa 1, que es definir el objetivo, es aquí donde se define la situación problemática que presenta la entidad y se realizó un análisis de los requerimientos (RUP) que se abarco el modelado del negocio y los requerimientos funcionales y no funcionales del sistema; etapa 2, llamada recolección de la data, es aquí donde se encargó de recolectar la mayor cantidad de mamografías digitales de diferentes bases de datos públicas ya que es la fuente principal para el entrenamiento del modelo; etapa 3, se utilizó para el preprocesamiento de la data recolectada que permitió estandarizar las imágenes digitales a un mismo tamaño, formato, color y dividir los datos en data de entrenamiento, testeo y validación, también aquí se desarrolló el análisis del sistema (RUP) y el diseño del sistema (RUP); etapa 4 se encargó de elegir el mejor algoritmo y red para el entrenamiento del modelo; etapa 5, aquí se desarrolló todo el proceso de la ejecución del modelo en base a la data de entrenamiento y de testeo; etapa 6, pertenece al proceso de validación que comprendió validación del modelo con la data de testeo y la data de prueba que como resultado se obtuvo los porcentajes de exactitud y precisión; por último la etapa 7 llamada modelo integrado en un sistema, se desarrolló el proceso de integración del modelo con el sistema y se aplicó también la fase de transición (RUP) donde se realizó las pruebas de caja negra y blanca.

Los aspectos éticos que se consideraron para el tema que involucra la salud de los pacientes se trabajó teniendo en cuenta el código ético y deontológico propuesto por el Colegio Médico del Perú [24], que habla sobre el respecto de la normativa internacional y nacional en cuanto a las investigaciones desarrolladas con seres humanos. Y para la protección y bienestar de los participantes en esta investigación fueron los siguientes: para la aplicación de técnicas de

recolección de datos, fueron utilizadas para la identificación de la problemática y aplicadas al radiólogo y oncólogo encargados del área de oncología. Esta información no será difundida ya que deberá respetar la privacidad de los participantes, tal como se indica en el código de ética y conducta profesional de Association for Computing Machinery (ACM) [25]. Asimismo, para la seguridad de los datos y secreto de la información fueron aplicados en cuanto al reporte de los resultados de los pacientes respecto a sus mamografías dado por el oncólogo. También para la construcción del sistema inteligente se utilizaron herramientas de software libre, con el fin de respetar los derechos de autor de acuerdo al Decreto Legislativo N° 822 [26] . Finalmente, para la recopilación de datos, se obtuvo información de varias fuentes bibliográficas trabajadas en cada una de las etapas, siendo citadas de forma honesta y realista según los datos disponibles, acorde al código ético del IEEE [27].

Resultados y discusión

Ahora, enfocándose en los resultados, las metodologías utilizadas en este estudio fueron las Etapas del Modelo de Machine Learning y la metodología RUP, la primera está compuesta por etapas que permiten la realización del modelo en cuanto a la parte inteligente y se apoya de la metodología RUP para realizar el aplicativo web que contiene la parte del negocio, análisis y diseño del sistema brindado, las Etapas del Modelo de Machine Learning se describen y se implementan a continuación:

1. Etapa 1: Definir el objetivo

En esta primera etapa se realiza la identificación de la problemática que presenta el área oncológica del hospital las Mercedes de Chiclayo, el lugar que permitió recolectar información. Con la ayuda de la metodología se muestra un listado de los problemas que se identificaron para determinar la posible solución. Todo lo descrito, se detalla a continuación en la Tabla I:

TABLA I
SITUACION PROBLEMÁTICA Y SOLUCIONES

Situación problemática
<ol style="list-style-type: none"> 1. Para obtener los resultados mamográficos en el hospital “Las Mercedes” se quiere un tiempo de 21 a 90 días como máximo, debido a que las evaluaciones mamográficas deben ser trasladadas a especialistas que se encuentran en la capital del país. De esta manera, el proceso requiere de un tiempo extra a lo planificado, reduciendo la tolerancia de los pacientes y deficiencia en la atención 2. El área de oncología cuenta con: un radiólogo que se encarga de analizar la mamografía de forma superficial y dos oncólogos que asisten a los pacientes en el único turno de la mañana. 3. La capacidad de atención es de 20 pacientes por día 4. Poca cantidad de mamografías que se podría obtener, ya que la base de datos del Hospital no se puede acceder debido a la coyuntura de la pandemia 5. Pocos recursos tecnológicos que se cuenta
Solución
<ul style="list-style-type: none"> • Para el problema 1, 2 y 3 se plantea crear un modelo computacional y un aplicativo para apoyar en el análisis mamográfico en la detección de tumores de mama en base a las mamografías digitales, el cual será usado para predecir si una mamografía tomada es de un Seno sano (Normal) o un Seno con tumores (Anormal) • Para el problema 4 se plantea utilizar bases de datos públicas ya certificadas • Para el ultimo problema se plantea trabajar con recursos tecnológicos de Open Source

Dentro de esta etapa también se realiza un análisis preliminar: diagrama de caso de uso, diagrama de actividades y requerimientos funcionales y no funcionales, esto pertenece al modelado del negocio de RUP que se detalla a continuación:

Modelado del negocio

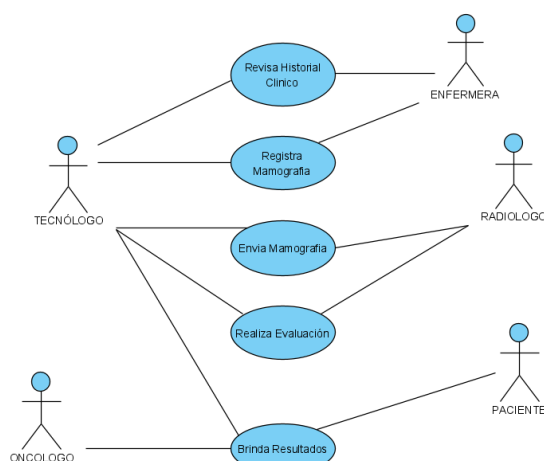


Figura 3. Diagrama de caso de uso del negocio

Dentro del modelado se realizó los diagramas de actividades de negocio de cada caso de uso que son:

- Revisar historial clínico donde interviene el tecnólogo que solicita el historial clínico del paciente a la enfermera, esta se le entrega para que luego sea revisado y mandar a tomar la mamografía al paciente.
- Registra mamografía: esta actividad lo inicia la enfermera que realiza la toma de la mamografía, luego es enviada y es recibida por el tecnólogo que es el encargado de registrar la mamografía.
- Enviar mamografía: esta actividad el tecnólogo verifica las mamografías digitales para que después se realice una copia de seguridad, luego es enviada y es recibida por los radiólogos que se encuentran en la capital del Perú.
- Realiza Evaluación: esta actividad comienza con la solicitud de los exámenes mamográficos que hace el tecnólogo, si estos exámenes aun no se han realizado por parte del radiólogo envía una respuesta de postergación de los exámenes y si lo están envía los resultados mamográficos solicitados.
- Brinda resultados: esta actividad comienza con el oncólogo que solicita los exámenes mamográficos, luego el tecnólogo verifica si los exámenes mamográficos estén listos, si lo están recibe los resultados el oncólogo y este se lo entrega al paciente.

Requerimientos

TABLA II
REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES

Funcionales	No funcionales
El sistema permitirá ingresar con usuarios autorizados.	Los datos modificados se actualizan para que todos los usuarios accedan a ellos en menos de 10 segundos.
El sistema le permitirá al tecnólogo que será el administrador del sistema subir las mamografías digitales de cada paciente.	El sistema es capaz de operar adecuadamente las 24 horas.
El sistema tendrá privilegios de usuarios en cuanto a la disposición de cada módulo	Los permisos de acceso al sistema solo son cambiados por el administrador de acceso a datos.
El sistema permitirá al radiólogo mostrar el resultado de cada mamografía de cada paciente	El sistema posee un diseño “responsive” que garantiza una vista completa en múltiples computadoras
El sistema generará un reporte de los exámenes de cada paciente y será enviado al oncólogo	Si el sistema sufre de algún fallo tendrá un backup de respaldo de la BD
El sistema permitirá al radiólogo evaluar sus mamografías de cada paciente.	
El sistema permitirá al administrador realizar copias y restauración de la BD	
El sistema permitirá al administrador asignar tipos de usuarios a los ya registrados	
El sistema permitirá registrar los datos del paciente	
El sistema permitirá listar los pacientes listos para ser evaluados como también los pacientes que aún no estén listos	
El sistema permitirá registrar todas las evaluaciones de cada mamografía digital	
El sistema permitirá almacenar como máximo 4 mamografías por cada paciente	
El sistema debe aceptar dos tipos de formato de archivo de la mamografía JPG y PNG.	
El sistema debe aceptar dos tipos de formato de archivo de la mamografía JPG y PNG.	
El sistema permitirá registrar el análisis propio del radiólogo de las mamografías	

2. Etapa 2: Recolección de la data

En esta etapa se explica las diferentes mamografías obtenidas, tanto de senos normales como de senos con tumores, las cuales componen la base de datos completa de este trabajo. Se utilizaron dos bases de datos públicas que tienen reconocimiento de diversas investigaciones de categorización y diagnósticos del cáncer de seno. Estas aún se encuentran a disposición de investigadores desde el año 2000, por lo cual la extensa revisión que han tenido, se difiere.

- **MINI-MIAS**

La Mammographic Image Analysis Society (MIAS) es una organización de investigación situada en Reino Unido, dedicada a la compresión de las mamografías y a su vez con su trabajo ha generado una base de datos de mamografías digitales del año 1994 [28]. Respecto al contenido de esta base de datos se dispone de 322 imágenes mamográficas, de las cuales 121 presentan tumores o anomalías, mientras tanto el resto son mamografías que no presentan anomalías (senos sanos). En cuanto a las características de las imágenes, encontramos un archivo de texto que especifica el carácter del tejido de fondo de la mama, la clase de anormalidad presente, la gravedad de anormalidad (benigno o maligno) y también nos especifica el tamaño de píxeles de 1024 x 1024 con un formato pmg en canal de blanco y negro.

- **DDSM**

Según [29], la DDSM, la cual proviene de sus siglas en inglés Digital Database for Screening Mammography, es una base de datos libre, creada a partir de la ayuda y empeño del hospital general de Massachusetts y la Universidad del Sur de Florida, departamento de ingenierías y ciencias computacionales, en el año 2000. En esta base de datos podemos encontrar un aproximado de 2500 estudios, que incluye dos imágenes de cada mama e información importante del paciente (edad que tenía cuando se realizó el estudio, densidad de la anomalía, etc.), pero sin mostrar información o algún carácter personal que pueda llevar a su identificación. Respecto al contenido de esta base de datos se tomaron las imágenes de senos normales y senos con tumor maligno segmentadas, las cuales se vio conveniente agregarlas a los datos de entrenamiento y de desarrollo. Con respecto a las características de las imágenes nos encontramos con un formato gif con canales RGB y su tamaño varía entre 160 x 220 píxeles.

3. Etapa 3: Preparar la data

En esta etapa se hará un procesamiento de la data o normalmente llamada limpieza de datos, esto se realiza para convertir la data en formas que produzcan mejores resultados.

En primero lugar al tener las imágenes mamográficas ya descargadas se formalizo el tipo de formato a JPG ya que las bases de datos de MINI-MIAS estaban en formato PMG y en cuanto a DDSM en formato GIF, por lo que se vio necesario convertirlas a un mismo formato y ser utilizadas en nuestro modelo de entrenamiento.

En segundo lugar, al ya tener las imágenes mamográficas normalizadas se ve la necesidad de utilizar algunos métodos de manipulación de imágenes para aumentar la base de datos, respecto a las imágenes de senos con tumores. Se utilizó una librería de Python llamada PIL que cuenta con varias herramientas para mejorar la calidad de la imagen y otras opciones de rotación, recorte, zoom y redimensiones de la imagen. Todo este proceso se obtuvo 1400 imágenes con las que se trabajará en el entrenamiento.

En cuanto a los parámetros fijos se maneja una división sobre el total de los datos, de los cuales se dividieron en tres conjuntos: entrenamiento, testeo y de validación. Sobre el porcentaje de distribución de los lotes es de 80% debido a que será tomado como su conjunto de entrenamiento, tal como lo sugiere el autor de manera empírica, es decir, por la experiencia obtenida de años de trabajo [16] ; 15% para el testeo y 5% para la validación (éstos últimos son distribuidos por decisión propia porque servirá de apoyo

para una mejor distribución para cada proceso y de esta forma obtener una evaluación de forma detallada).

La última transformación es crear una función que recorre todos los directorios y los archivos correspondientes de cada carpeta de entrenamiento y de testeo, así mismo, se establece un tamaño fijo de 150 X 150 y con una dimensión 1, es decir a escala de grises. Estas imágenes se cargaron en arrays que serán guardados en formato npy localmente para posteriormente ser llamados para el entrenamiento del modelo.

Análisis del sistema (RUP)

Dentro de esta etapa también se realizó el diagrama de caso de uso del análisis que se visualiza en la figura 4 y por cada caso uso se realizó los diagramas de realización que son detallados a continuación:

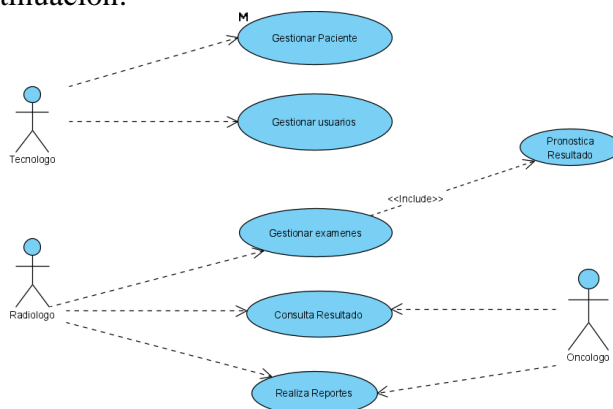


Figura 4. Diagrama de casos de uso de análisis

- En el paquete del módulo administrador se tiene el RCUA: inicio de sesión y RCUA: gestionar usuario
- En el paquete del módulo paciente se tiene el RCUA: gestionar paciente y RCUA: listar pacientes
- En el paquete del módulo exámenes se tiene el RCUA: gestionar examen, RCUA: consultar examen y RCUA: generar reporte

Diseño del sistema (RUP)

Por consiguiente, se realizó dentro del diseño: el diagrama de contexto de diseño, los diagramas de realización de cada caso de uso de diseño, por cada realización se realizó también los diagramas de secuencia y finalmente concluyendo con el diseño físico de la base de datos que se visualiza en la figura 5.

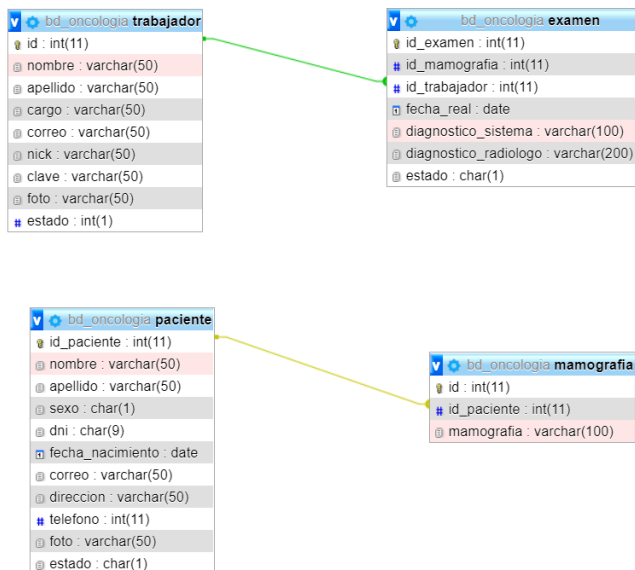


Figura 5. Diseño físico de la base de datos

Dentro en esta fase también se realizó el diseño de las interfaces del sistema que se visualiza en la figura 6 que corresponde a las interfaces más importantes que tiene el sistema.

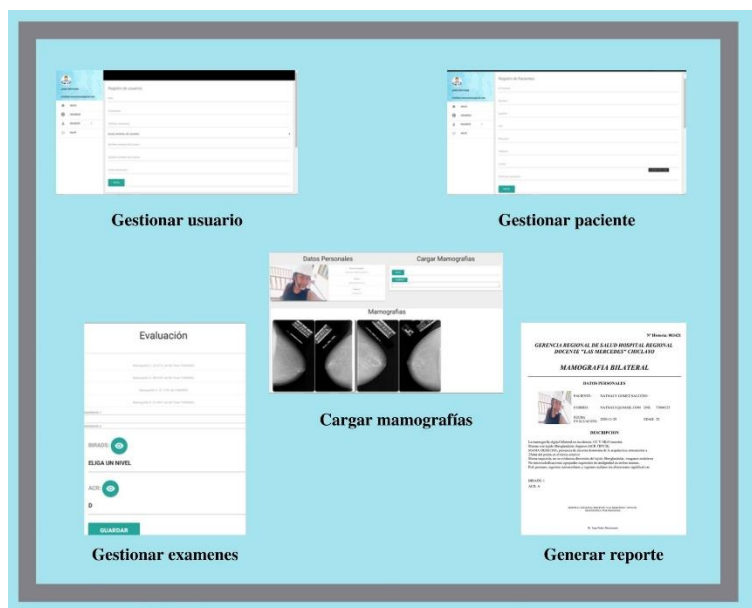


Figura 6. Interfaces del sistema

Para finalizar se diagramó el diseño de la arquitectura donde se detalla la integración del modelo con el sistema, ya que es un punto importante para la interacción del experto con el modelo. Esto se visualiza en la figura 7 y es detallada a continuación:

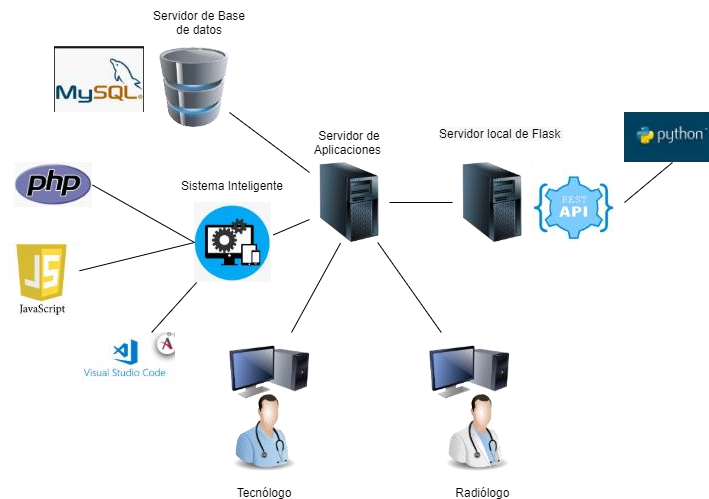


Figura 7. Diseño de la arquitectura

Para la construcción del modelo se desarrolló en el IDE de Google Colab que es una herramienta basada el IDE de Jupyter que permite ejecutar y programar en lenguaje Python ya sea para entrenamientos de imágenes o para ciencia de datos. Otro beneficio que nos ofrece es tener una máquina virtual con suficiente capacidad de procesamiento si en caso no se cuenta con una computadora de grandes características.

Para el desarrollo del sistema web local se utilizó el lenguaje de programación PHP, y para el front-end para el diseño se usó Material Design.

Como se describe anteriormente tenemos dos lenguajes diferentes PHP y Python. Por lo cual, se ve la necesidad de desarrollar una API REST en Python que nos permite la comunicación fluida entre los diferentes lenguajes.

Esta API REST es almacenada en un servidor local llamado Flask ya que por medio de una petición que hace el sistema envía la imagen y el servidor ejecuta el API REST retornando el diagnóstico.

4. Etapa 4: Elección del algoritmo

En esta etapa corresponde a elegir el algoritmo que más se adecue a nuestro problema a resolver. Para optar con que algoritmo vamos a trabajar, debemos de conocer en qué grupo se encuentra el problema a resolver de los cuales son:

- Aprendizaje supervisado
- Aprendizaje sin supervisión
- Aprendizaje por refuerzo

Para este trabajo de investigación se encontró dentro del aprendizaje supervisado ya que se establecieron grupos ya clasificados para el aprendizaje, es decir, se clasificaron la data en mamografías que tienen tumores y las que son sanas.

Asimismo, se seleccionó también los algoritmos de Aprendizaje profundo donde se ejecutan varias capas de algoritmos de redes neuronales, que cuenta con una cantidad de entradas que pasaran progresivamente a la siguiente capa donde cada una se encargará de filtrar contornos, bordes, etc. A más capas que pase la imagen más característica se aprenderá. Por lo cual se hará algunas comparativas de algunos modelos para poder elegir el adecuado en nuestro proyecto.

En la Tabla III se hace una comparativa de algunos algoritmos que se diferencian de los datos de procesamiento y del campo donde son utilizadas.

TABLA III
COMPARACIÓN DE LOS DIFERENTES MODELOS

<u>ALGORITMO</u>	<u>DATA DE PROCESAMIENTO</u>	<u>EJEMPLO DE APLICACIÓN</u>
Regresión Logística	Valores numéricos	Predicción de precios
Redes neuronales convolucionales	imágenes	Reconocimiento de objetos
Redes neuronales recurrentes	reconocimiento de escritura “a mano alzada”, composición musical, y por supuesto, reconocimiento del habla	Reconocimiento de voz Predicciones del clima
K-means	texto	Segmentación de clientes
Arboles de decisión	texto	Motores de recomendaciones
Clasificación Bayesiana	texto	Clasificación de email (spam)

De acuerdo con esta Tabla, las redes convolucionales se utilizan principalmente para el entrenamiento de imágenes. La eficiencia mostrada en los resultados de su investigación es una de las principales razones por las que el aprendizaje profundo y el aprendizaje automático han provocado un nuevo auge en el campo de la investigación. Por tal motivo se eligió trabajar con este algoritmo porque se van a trabajar con imágenes digitales.

5. Etapa 5: Entrenar el modelo

Al tener ya el algoritmo seleccionado, en esta etapa se seleccionó un porcentaje de la data que es del 80% del total y la utilizaremos como data de entrenamiento que sirvió para alcanzar el objetivo planteado inicialmente.

Antes de empezar con el entrenamiento se detalla las librerías que se utilizaron para el entrenamiento usando el lenguaje de programación Python[30]:

- **Pandas:** librería que es utilizado para la manipulación de datos de 1 dimensión y DataFrame para datos de 2 dimensiones.
- **Tensorflow:** Una biblioteca de Python desarrollada por Google para realizar cálculos numéricos utilizando diagramas de flujo de datos y para el aprendizaje profundo y otras aplicaciones informáticas científicas.
- **Numpy:** La biblioteca proporciona funciones matemáticas avanzadas para operar en estas estructuras de datos.

- **Os:** Es un módulo que nos proporciona un submódulo de ruta (`os.path`), que nos permite acceder a ciertas funciones relacionadas con nombres de archivos y rutas de directorio.
- **Matplotlib:** La biblioteca de gráficos estándar de Python y la más famosa. Puedes usar `matplotlib` para generar gráficos con la calidad necesaria y publicarlos gráficamente en papel y forma digital.
- **Sys:** El módulo `sys` es responsable de proporcionar variables y funciones directamente relacionadas con el intérprete.
- **Cv2:** Es una biblioteca de enlaces de Python diseñada para resolver problemas de visión por computadora. `cv2`. El método `imread()` carga una imagen del archivo especificado. Si la imagen no se puede leer (por falta de archivo, permisos incorrectos, formato no compatible o no válido), este método devuelve una matriz vacía.
- **Sklearn:** es una biblioteca de aprendizaje automático gratuita para Python. Cuenta con varios algoritmos como máquina de vectores de soporte, bosques aleatorios y vecinos `k`, y también admite bibliotecas numéricas y científicas de Python como `NumPy` y `SciPy`.
- **Skimage:** Es un paquete Python de código abierto diseñado para el preprocesamiento de imágenes
- **Tqdm:** es una herramienta que ayuda a visualizar la carga de archivos de manera dinámica

Ya comenzando con el entrenamiento del modelo, se realizó siguiendo la arquitectura de una red neuronal convolucional:

Capa convolucional:

En primera instancia se desarrolló el modelo como se visualiza en la figura 8 conteniendo 4 capas convolucionales, donde cada una de ellas contiene 64 neuronas, un filtro de 3 X 3, se utiliza la función de activación ReLU y al comenzar la primera capa se especifica el tamaño de la imagen de 150 X 150 y de dimensión 1, es decir a grises. Respecto a la salida de la red, contiene dos salidas con una activación de SOFTMAX donde determinaremos si es un seno SANO o un seno con TUMOR.

```
#crear la red neuronal convolucional
# Inicializamos el modelo
model_v1 = models.Sequential()
#Primera Capa con 64 neuronas con un filtro 3 x 3, una activación 'relu' y un tamaño de imagen 150 x 150 x 1
model_v1.add(layers.Conv2D(64, (3, 3), activation='relu', input_shape=(150, 150, 1)))
model_v1.add(layers.MaxPooling2D((2, 2)))

#Segunda Capa con 64 neuronas con un filtro 3 x 3, una activación 'relu' ya no se pasa el tamaño porque ya se
model_v1.add(layers.Conv2D(64, (3, 3), activation='relu'))
model_v1.add(layers.MaxPooling2D((2, 2)))

#Tercera Capa con 64 neuronas con un filtro 3 x 3, una activación 'relu' con una capa de maxPooling de 2 x 2
model_v1.add(layers.Conv2D(64, (3, 3), activation='relu'))
model_v1.add(layers.MaxPooling2D((2, 2)))
#Cuarta Capa con 64 neuronas con un filtro 3 x 3, una activación 'relu' sin una capa de maxpooling
model_v1.add(layers.Conv2D(64, (3, 3), activation='relu'))

# Planamos las imagenes
model_v1.add(layers.Flatten())
# La ultima capa de 64 neuronas
model_v1.add(layers.Dense(64, activation='relu'))
# Obtendremos dos salidas por que se evaluara si es Seno Normal o con Tumore y con una activación de 'Softmax'
model_v1.add(layers.Dense(2, activation='softmax'))
```

Figura 8. Capas del modelo entrenado

Capa MaxPooling:

Respecto a esta capa se implementó en la primera, segunda y tercera capa convolucional donde se le agregó una capa de MaxPooling con un filtro de 2X2.

Compilación del modelo:

Respecto a la compilación del modelo se visualiza en la figura 9, donde se utilizó como parámetro de Optimizador “Adam” que es un método de optimización estocásticas eficiente que solo requiere gradientes de primer orden con pocos requisitos de memoria y que el ritmo de aprendizaje se mantiene fijo durante todas las actualizaciones de pesos. El método calcula las tasas individuales de aprendizaje adaptativo para diferentes parámetros de estimaciones del primer y segundo momento de los gradientes. En base a su nombre se deriva de la estimación del momento adaptativo [31].

```
model_v1.compile(optimizer='adam',
                 loss='sparse_categorical_crossentropy',
                 metrics=['accuracy'])
```

Figura 9. Compilación del modelo

Ejecución del modelo:

Ya compilado el modelo entra a ejecutarse, ingresando los parámetros de los arrays de entrenamiento de senos sanos y los arrays de senos con tumores, también se le indicó las épocas que se va a entrenar el modelo, siendo para este caso 7 y en validación de la data se le envió los arrays de testeo ya sea de senos sanos y con tumores esto se visualiza en la figura 10.

```
H1 = model_v1.fit(X_trainReshaped,
                 y_train,
                 epochs=7, # Aqui indicamos las epocas que se van a entrenar
                 validation_data = (X_testReshaped,y_test),
                 callbacks=[tensorboard_callback])
```

Figura 10. Ejecución del modelo

6. Etapa 6: Validación del modelo

En esta etapa corresponde a la validación del modelo que fue realizada con la data de testeo y la data de validación, que fue en este caso el 15% y 5% respectivamente. Con esta data se ejecutó el modelo y después se evaluaron los resultados obtenidos. Cabe recalcar que existe la posibilidad de que el modelo no funcione bien con esta data de validación y tendremos que volver a la etapa 6 hasta que el modelo se ajuste bien a las dos particiones (data de entrenamiento y data de validación). Con esto se llega a obtener un modelo más correcto y seguro.

Respecto a la data de testeo se obtuvo:

Con la data de testeo se evaluó el porcentaje de exactitud que tiene las imágenes de testeo sobre las imágenes de entrenamiento y nos arrojó una exactitud del 85.7%, es decir que el modelo puede predecir exactamente si una mamografía es de un seno sano o con tumor. Esto se visualiza en la figura 11.

```
test_loss, test_acc = model_v1.evaluate(X_testReshaped, y_test)
print(test_acc)

7/7 [=====] - 0s 7ms/step - loss: 0.4177 - accuracy: 0.8571
0.8571428656578064
```

Figura 11. Porcentaje de exactitud de la data de testeo

Respecto a la data de validación se obtuvo:

Ahora con la data de validación que contiene 70 imágenes de las cuales están divididas, la mitad de mamografías sanas y el resto de mamografías con tumores. Se trabajo utilizando la matriz de confusión que se visualiza en la Tabla IV donde se obtuvieron: porcentaje de exactitud, porcentaje sensibilidad, porcentaje de especificidad y porcentaje de precisión.

TABLA IV
MATRIZ DE CONFUSIÓN [32]

		<u>PREDICCIÓN DEL MODELO</u>	
		Mamografías con tumores (enfermo)	Mamografías sin tumores (sano)
<u>PREDICCIÓN REAL</u>	Mamografías con tumores (enfermo)	VP = 25	FN = 10
	Mamografías sin tumores (sano)	FP = 4	VN= 31

Para determinar las métricas de rendimiento del modelo se trabajó con las siguientes formulas [32]:

- **Porcentaje de exactitud:**

$$Exactitud = \frac{VP + VN}{VP + FN + FP + VN}$$

$$Exactitud = \frac{25 + 31}{25 + 10 + 6 + 29}$$

$$Exactitud = \frac{56}{70} = \mathbf{80\%}$$

- **Porcentaje de Sensibilidad:**

$$Sensibilidad = \frac{VP}{VP + FN}$$

$$Sensibilidad = \frac{25}{25 + 10}$$

$$\text{Sensibilidad} = \frac{25}{35} = 71.4\%$$

- **Porcentaje de Especificidad:**

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

$$\text{Especificidad} = \frac{31}{31 + 4}$$

$$\text{Especificidad} = \frac{31}{35} = 88.5\%$$

- **Porcentaje de Precisión:**

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Precisión} = \frac{25}{25 + 4}$$

$$\text{Precisión} = \frac{25}{29} = 86.2\%$$

De acuerdo con los resultados obtenidos se determina que el modelo puede predecir con exactitud del 80% si una mamografía tiene tumores o no tiene. Respecto al porcentaje de sensibilidad se obtiene un 71.4% y el porcentaje de especificidad en un 88.5%, esto es debido a que el modelo tiene más facilidad en predecir mamografías sanas que no tienen tumores a diferencia de las que tiene. Y por último respecto al porcentaje de precisión nos arroja un 86.2%, debido a que se tiene predicciones correctamente clasificada entre el total de las predicciones predichas.

7. Etapa 7: Modelo integrado en un sistema

En esta etapa el modelo desarrollado será integrado en un sistema, donde retorna el porcentaje de no tener o de tener tumores de mama de cada mamografía. Si el porcentaje diera menos del 60% el sistema recomendaría que se realice una evaluación con RNM al paciente. También se tendrá por cada resultado de la mamografía su respectiva caja de texto que le sirve al radiólogo registrar su evaluación respecto a los resultados obtenidos en el sistema como se visualiza en la figura 12.

Evaluación

FECHA DE REGISTRO:
2021-06-02

RESULTADOS:

Mamografía 1: Hay una probabilidad del 99.20% de NO presentar TUMORES de mama

Mamografía 2: Hay una probabilidad del 93.06% de NO presentar TUMORES de mama

Mamografía 3: Hay una probabilidad del 81.28% de tener TUMORES de mama

Mamografía 4: Hay una probabilidad del 83.82% de tener TUMORES de mama

Figura 12. Resultados dentro del sistema

Dentro de la última etapa también se realizó la fase de transición: que corresponde a la realización de las pruebas de caja negra y caja blanca del sistema inteligente.

Respecto a la caja negra se aplicó en el registro de los usuarios, registro de los pacientes y el registro de las mamografías de cada paciente.

Y respecto a la caja blanca se aplicó en el inicio de sesión y en la validación del formato de la mamografía registrada.

8. Resultados en base a los objetivos planteados

8.1. Determinar las características asociadas a las imágenes mamográficas para detectar tumores de mama.

Para el desarrollo de esta investigación, se empezó por tener un amplio conocimiento de la literatura acerca de las características de las imágenes mamográficas. Este conocimiento fue adquirido por reuniones, libros, documentos que fueron brindados por el oncólogo y radiólogo encargado, así mismo para complementar se usaron otras fuentes como se detalla en revisión de la literatura.

Este objetivo es de mucha importancia ya que al tener la información clara y precisa tendremos la facilidad de poder desarrollar el sistema inteligente propuesto. Logrando así el cumplimiento de este objetivo por medio de una entrevista que se visualiza en el Anexo 02, que fue dirigida al especialista. Asimismo, con esta recopilación de información se obtuvo una de las limitantes sobre las complicaciones en detectar tumores de mama cuando se tiene el tipo de seno extremadamente denso que se visualiza en la figura 1.

8.2. Crear un modelo computacional para detectar tumores de mama en base a las mamografías digitales.

Primeramente, para el cumplimiento del segundo objetivo se crearon 10 modelos con diferentes filtros y capas. Este modelo fue construido a partir de las teorías de machine learning donde se encuentra en la subcategoría de clasificación, debido a que las mamografías están previamente clasificadas para el entrenamiento. Asimismo, al utilizar imágenes digitales se hace uso de las redes neuronales convolucionales que se encuentra en el campo del Deep learning. Por consiguiente, se va escoger el mejor modelo tomando como referencia el mejor porcentaje de predicción sobre la data de testeo con la data de entrenamiento. A continuación, en la Tabla V se muestra los 10 modelos.

TABLA V
COMPARACIÓN DE LOS 10 MODELOS DESARROLLADOS

<u>NOMBRE DEL MODELO</u>	<u>CANTIDAD DE CAPAS</u>	<u>NUMERO DE ÉPOCAS</u>	<u>OPTIMIZADOR</u>	<u>PORCENTAJE DE PREDICCIÓN</u>
Modelo 1	5 capas	20 épocas	adam	81.4%
Modelo 2	4 capas	20 épocas	adam	83.2%
Modelo 3	4 capas	7 épocas	adam	81.9%
Modelo 4	3 capas	7 épocas	adam	85.2%
Modelo 5	2 capas	20 épocas	adam	82.8%
Modelo 6	4 capas	12 épocas	adam	84.7%
Modelo 7	4 capas	7 épocas	adam	85.7%
Modelo 8	4 capas	20 épocas	Adam	84.5%
Modelo 9	3 capas	12 épocas	sgd	79%
Modelo 10	4 capas	7 épocas	adam	81.4%

De acuerdo con esta se eligió el modelo 7 que tiene una predicción del 85.7% siendo el más alto de todos los modelos planteados. Este modelo fue desarrollado en la etapa 5.

8.3. Validar la precisión del modelo en la detección de tumores de mama en base a las mamografías digitales.

Para el cumplimiento del tercer objetivo se desarrolló en primera instancia la validación con los datos de prueba ya previamente clasificados en mamografías sin tumores (sanas) y mamografías con tumores (anormales), ya que se encuentra dentro del aprendizaje supervisado y en la subcategoría de clasificación. Esta validación se desarrolló en la etapa 6.

Para tener una validación más aceptable del modelo se certificó con el análisis que realizó el radiólogo y el análisis que hace el sistema inteligente. Esta certificación se obtuvo por medio de un instrumento aplicado al radiólogo que se visualiza en el Anexo 03, donde se le brinda 70 mamografías digitales.

Para evaluar la calidad de las predicciones del modelo se utilizó la matriz de confusión:

TABLA VI
MATRIZ DE CONFUSIÓN [32]

		<u>PREDICCIÓN DEL MODELO</u>	
		Mamografías con tumores (enfermo)	Mamografías sin tumores (sano)
<u>PREDICCIÓN DEL RADIOLOGO</u>	Mamografías con tumores (enfermo)	VP = 29	FN = 6
	Mamografías sin tumores (sano)	FP = 2	VN = 33

Para determinar las métricas de rendimiento del modelo se trabajó con las siguientes formulas [32]:

- **Porcentaje de exactitud:**

$$Exactitud = \frac{VP + VN}{VP + FN + FP + VN}$$

$$Exactitud = \frac{29 + 33}{29 + 6 + 2 + 33}$$

$$Exactitud = \frac{62}{70} = \mathbf{88.5\%}$$

- **Porcentaje de precisión:**

$$Precisión = \frac{VP}{VP + FP}$$

$$Precisión = \frac{29}{29 + 2}$$

$$Precisión = \frac{29}{31} = \mathbf{93.5\%}$$

Con los resultados obtenidos de la matriz se dedujo que el modelo tiene una exactitud del 88.5% de predecir una mamografía con tumor y sin tumor respecto al análisis realizado por el experto. Asimismo, también se obtuvo una precisión del 93.5%, debido a que se tiene predicciones correctamente clasificada entre el total de las predicciones predichas.

8.4 Crear una aplicación para la detección de tumores de mama en base a las mamografías digitales.

Para la creación de la aplicación se utilizó las fases de la metodología RUP que están distribuidas en cada etapa desarrollada como, por ejemplo, el análisis de requerimientos

se detalla en la etapa 1. El análisis y diseño de la aplicación web local se desarrolló en la etapa 3. Al culminar con el desarrollo de la aplicación uno de los detalles importante es unir el modelo con el sistema, esto se detalló en el diseño de la arquitectura. Finalmente, los resultados obtenidos del modelo se implementaron en el sistema inteligente esto se detalló en la etapa 7. Asimismo, las pruebas de la funcionalidad del software se detallan en la misma etapa y la sección de la fase de transición de RUP.

Para complementar el cumplimiento de este objetivo se elaboró un instrumento de validación respecto a la funcionalidad del sistema inteligente. Esto se visualiza en el Anexo 04, donde se le aplicó a un docente de la carrera de ingeniería de sistemas obteniendo una validación aceptable de todos los requerimientos establecidos.

A partir de los resultados obtenidos y entrando a la parte de la discusión, se puede ver que al determinar las características asociadas a las imágenes mamográficas para detectar tumores de mama apoyados de un especialista en análisis mamográfico ha permitido filtrar los factores de riesgo más importantes para el desarrollo del sistema.

Uno de los factores dentro de esta investigación se trata de la dificultad de detectar tumores de mama cuando se tiene un tipo de seno extremadamente denso, por lo que [9], al tener este problema, realiza un trabajo con más profundidad ya que no hace uso de mamografías digitales sino de imágenes de tejido mamario. Esto ayuda a tener una vista más profunda del análisis de la mama y poder resolver este problema. Sin embargo, al no tener este tipo de imágenes en la presente institución, se tuvo que realizar el trabajo teniendo en cuenta este factor.

Al crear el modelo computacional para detectar tumores de mama se ha hecho utilizando las redes neuronales convolucionales y la construcción de la red se elaboró desde cero. Teniendo que elegir el modelo más óptimo de los 10 elaborados. De acuerdo con [6], utiliza el mismo método para la construcción del modelo, pero en este caso de tratarse de una competencia propuesta por The International Skin Imaging Collaboration (ISIC) 2018, se eligió entrenar el modelo utilizando redes pre entrenadas las cuales son: AlexNet, VGG y ResNet. Esto es debido por cuestiones de tiempo, ya que le tomaría mucho tiempo desarrollar una nueva. En tal sentido, se resalta que un modelo creado desde cero contribuye a la investigación, pero si se trata de realizar trabajos de suma importancia como esta competencia es necesario utilizar una red que ya fue validada por otros expertos.

Al determinar la validez de la precisión del modelo en predecir si hay tumores de mama o no, se realizó una comparativa por medio de los resultados obtenidos del radiólogo. Esta comparación se ve reflejado en la Tabla XXIV, que se obtuvo una precisión del 93.5% y un acierto del 88.5% alineadas a los resultados que dio el experto. Esto quiere decir que el modelo puede ser un buen apoyo para el análisis mamográfico. Respecto a estos resultados se hace una comparativa con [7], que también hace una investigación relacionada con ecografías mamarias. Es aquí donde se evidencia que el modelo desarrollado tiene una precisión del 62.3%, por lo cual esto es debido a que el mismo autor manifiesta la carencia de no poseer una gran base de datos que contenga ecografías mamarias. En este sentido, se recalca la importancia de tener una gran cantidad de imágenes para el entrenamiento, ya que para llegar a una alta precisión es un requisito indispensable tener la data apropiada y robusta. Para apoyar esta teoría, en la investigación de [10], se trabajó con 10000 imágenes mamográficas obtenidas por la entidad pública EsSalud. Esta gran base de datos se utilizó para el entrenamiento del modelo en la identificación de microcalcificaciones en la mama. Asimismo, al culminar con el modelo se obtuvo una precisión del 100% y una exactitud de 91%. Ante estos resultados se valida la teoría de trabajar con una gran base de datos para el entrenamiento.

Por último, para la creación de una aplicación se desarrolló fusionando las etapas del modelo de machine learning con las fases de la metodología RUP, donde la última se utilizó para el desarrollo de todo el sistema web local que fue programado usando el lenguaje php. Logrando así obtener una interfaz que sirve para interacción del modelo con el radiólogo y además cumpliendo con los requerimientos establecidos por experto. En las investigaciones de [6], [9] y [10] solo llegan a la creación del modelo, por lo cual dentro de sus recomendaciones hace énfasis de la implementación de sus modelos en un aplicativo ya sea móvil o web. Respecto a la investigación que si concluye con el desarrollo de un aplicativo se encuentra en el trabajo de [8]. Este autor implementó un aplicativo móvil haciendo uso del modelo creado para tener una interacción amigable y final para el usuario, quien será el responsable del uso del sistema inteligente. Teniendo como base estas investigaciones se optó a desarrollar un aplicativo web local ya que estará alojada dentro de la entidad pública y solamente es utilizado por los responsables del área de oncología y el radiólogo.

Conclusiones

Primero, al determinar las características asociadas de una mamografía por medio de entrevistas y reuniones con el experto se logró obtener una base sólida de conocimiento. Así mismo, para complementar este conocimiento se utilizaron fuentes externas como primer pilar, ayudando a la construcción del sistema inteligente.

Segundo, Se logró crear un modelo computacional desde cero haciendo uso de las redes neuronales convolucionales, logrando así clasificar las mamografías que presentar tumores y las que no (sanas) con una precisión del 85.7% en base a la data de entrenamiento con la data de testeo.

Por consiguiente, se logró validar el modelo en primera instancia con la data de validación por medio de la matriz de confusión, obtenido un 80% de exactitud y un 86.2% de precisión. En segunda instancia para tener una validación más exacta se recurrió a mostrar 70 mamografías digitales para que fueran analizadas por el experto (radiólogo) y así comparar los resultados con el modelo. Obteniendo así un 88.5% de exactitud y un 93.5% de precisión de casos acertados. Por lo que se concluye que el sistema inteligente brinda resultados fiables.

Por último, se logró implementar el modelo en un aplicativo web local haciendo uso de la metodología RUP que se desarrolló en cada etapa del modelo de Machine Learning. Asimismo, a parte del modelo se cumplieron con los requerimientos establecido por normativa del colegio de médicos del Peru y dados por el experto.

Recomendaciones

Para futuros trabajos, se recomienda mejorar el modelo planteado ya sea aumentando las capas, el tamaño del filtro o usando otro optimizador para así obtener una precisión más exacta para la detección de tumores de mama.

También, se recomienda entrenar el modelo con más imágenes ya que en esta investigación por motivos de la pandemia se utilizaron bases de datos libres. Por lo cual al tener una fuente de base de datos grande de mamografías actuales el entrenamiento sería más óptimo.

Asimismo, se recomienda mejorar el problema que tiene el modelo en reconocer los tumores de mama cuando se tienen el tipo de seno extremadamente denso ya que esta categoría para un radiólogo con poca experiencia se le es difícil detectar anomalías en la mama.

Por último, en cuanto al desarrollo del sistema se recomienda agregar un módulo de citas. Esto podría apoyar en la atención médica que se le brinda al paciente como también tener un seguimiento de los controles médicos que son realizados en el área de oncología.

Referencias

- [1] F. Ahmad, N. A. M. Isa, M. H. M. Noor, y Z. Hussain, «Intelligent Breast Cancer Diagnosis Using Hybrid GA-ANN», en *2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks*, Madrid, Spain, jun. 2013, pp. 9-12. doi: 10.1109/CICSYN.2013.67.
- [2] J. D. Sanchez y <https://www.facebook.com/pahowho>, «OPS/OMS | Cáncer de mama», *Pan American Health Organization / World Health Organization*. <https://cutt.ly/znwK7eD> (accedido may 24, 2021).
- [3] «Cáncer de mama - Estadísticas», *Cancer.Net*, jun. 21, 2012. <https://cutt.ly/3nwK6lr> (accedido may 24, 2021).
- [4] «Plan preventivo.pdf». Accedido: may 24, 2021. [En línea]. Disponible en: <http://bvs.minsa.gob.pe/local/MINSA/4234.pdf>
- [5] D. L. Clavijo Rodríguez, M. Bernal Valencia, y J. F. Silva, «Sistema inteligente de reconocimiento de enfermedad coronaria (isquemia)», *Arch. Med. Manizales*, vol. 12, pp. 54-64, jun. 2012, doi: 10.30554/archmed.12.0.1565.2006.
- [6] C. Pérez Lorenzo, «Detección precoz de cáncer de piel en imágenes basado en redes convolucionales», Tesis de grado, Universidad Autónoma de Madrid, Madrid, Spain, 2019. Accedido: abr. 02, 2021. [En línea]. Disponible en: <https://cutt.ly/bnwK11X>
- [7] H. Chanampe, S. Aciar, G. Carrascosa, y A. Lorefice, «Modelo de Redes Neuronales Convolucionales Profundas para la Clasificación de Lesiones en Ecografías Mamarias», p. 6.
- [8] C. A. García Cortegano, «Sistema inteligente móvil basado en redes neuronales convolucionales para el diagnóstico previo de enfermedades renales», Tesis doctoral, Universidad Nacional Federico Villarreal, Lima, 2020. Accedido: abr. 02, 2021. [En línea]. Disponible en: <https://cutt.ly/OnwZAZf>
- [9] K. L. Yacolca Huamán, «Estudio de Atipia celular utilizando redes neuronales convolucionales: aplicación en tejidos de cáncer de mama», Tesis de pregrado, Pontificia Universidad Católica del Perú, Lima, 2020. Accedido: abr. 03, 2021. [En línea]. Disponible en: <https://cutt.ly/9nwXlyu>
- [10] R. Quesquen Liza, «Utilización de algoritmos para la identificación automática de microcalcificaciones en imágenes digitales de mamografía», Tesis de pregrado, Universidad Señor de Sipan, Pimentel, 2020. Accedido: abr. 04, 2021. [En línea]. Disponible en: <https://bit.ly/3woLUMR>
- [11] F. Roche, *Hablemos del Cáncer de mama*, Activos de Comunicación Visual. Madrid, Spain: ACV, 2011. Accedido: abr. 15, 2021. [En línea]. Disponible en: <https://bit.ly/3gmZJUX>
- [12] V. Paulina Neira, «Densidad mamaria y riesgo de cáncer mamario», *Rev. Médica Clínica Las Condes*, vol. 24, n.º 1, pp. 122-130, ene. 2013, doi: 10.1016/S0716-8640(13)70137-8.
- [13] «Senos densos: Respuestas a las preguntas más frecuentes - Instituto Nacional del Cáncer», feb. 16, 2018. <https://bit.ly/3gqejLC> (accedido may 24, 2021).
- [14] L. Aibar *et al.*, «Clasificación radiológica y manejo de las lesiones mamarias», *Clínica E Investig. En Ginecol. Obstet.*, vol. 38, n.º 4, pp. 141-149, jul. 2011, doi: 10.1016/j.gine.2010.10.016.
- [15] «manual-operativo-de-uso-de-mamografia.pdf». Accedido: may 24, 2021. [En línea]. Disponible en: <https://bancos.salud.gob.ar/sites/default/files/2018-10/0000000013cnt-10-manual-operativo-de-uso-de-mamografia.pdf>
- [16] F. Chollet, *Deep learning with Python*. Shelter Island, New York: Manning Publications Co, 2018.

- [17] P. Harrington, *Machine learning in action*. Shelter Island, N.Y: Manning Publications Co, 2012. [En línea]. Disponible en: <https://bit.ly/3pDL8sF>
- [18] J. Hurwitz, «Machine Learning For Dummies®, IBM Limited Edition», p. 75, 2018.
- [19] Y. LeCun, Y. Bengio, y G. Hinton, «Deep learning», *Nature*, vol. 521, n.º 7553, pp. 436-444, may 2015, doi: 10.1038/nature14539.
- [20] L. Á. Santamaría Colula, «Reconocimiento de patrones en imágenes médicas por medio de redes neuronales convolucionales», Tesis de Maestría, Benemérita Universidad Autónoma de Puebla, Mexico, 2019. Accedido: abr. 15, 2021. [En línea]. Disponible en: <https://bit.ly/3pFqesZ>
- [21] J. Wu, «Introduction to Convolutional Neural Networks», p. 31, may 2017.
- [22] I. A. on August 22 y 2017, «A Friendly Introduction to Convolutional Neural Networks». <https://bit.ly/3g77dN8> (accedido may 25, 2021).
- [23] N. Buduma y N. Locascio, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, Inc., 2017.
- [24] «codigo_etica.pdf». Accedido: may 25, 2021. [En línea]. Disponible en: <https://www.cmp.org.pe/wp-content/uploads/2021/01/CODIGO-DE-ETICA.pdf>
- [25] «The ACM Code of Ethics arose from the experiences, values and aspirations of computing professionals around the world, and captures the conscience of the profession. It affirms an obligation of computing professionals to use their skills for the benefit of society.» <https://bit.ly/3iBcq1d> (accedido may 25, 2021).
- [26] «GDA_CreadoresDeSoftware.pdf». Accedido: may 25, 2021. [En línea]. Disponible en: https://www.indecopi.gob.pe/documents/20182/143803/GDA_CreadoresDeSoftware.pdf
- [27] «IEEE Código de Ética». <https://bit.ly/3va0Ahj> (accedido may 25, 2021).
- [28] «The mini-MIAS database of mammograms». <https://bit.ly/3wbyi7x> (accedido may 25, 2021).
- [29] «USF Digital Mammography Home Page». <https://bit.ly/3isLYH7> (accedido may 25, 2021).
- [30] «15 Librerías de Python para Machine Learning - IArtificial.net». <https://bit.ly/3v66IqF> (accedido may 25, 2021).
- [31] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization», *ArXiv14126980 Cs*, ene. 2017, Accedido: may 25, 2021. [En línea]. Disponible en: <http://arxiv.org/abs/1412.6980>
- [32] A. Botana López *et al.*, *Avances en Informática y Automática. Decimocuarto workshop*. 2020. Accedido: may 09, 2021. [En línea]. Disponible en: <https://gredos.usal.es/handle/10366/144706>

Anexos

Anexo 01. Guía de análisis documental

<p>Título de la investigación:</p> <p>SISTEMA INTELIGENTE PARA APOYAR AL ANÁLISIS MAMOGRAFICO EN LA DETECCIÓN DE TUMORES DE MAMA FEMENINO ENTRE LAS EDADES DE 40 A 60 AÑOS EN EL HOSPITAL "LAS MERCEDES" DE CHICLAYO</p>
--

Categoría	Clasificación		
1. Procedencia	Artículo en revista científica		De Revisión
			Resultado de Investigación
			Editorial
			Revista indexada
			Revista no indexada
	Libro		Completo
			Capítulo de Libro
	Otros		Ponencia en congreso, simposio o seminario
			Documento electrónico
			Revista en la Internet
			Material audiovisual
			Documento Legal
			Material no publicado
	Tesis – Trabajo de Grado		Otro:
		Pregrado	
		Maestría	
Lugar de procedencia	Ciudad		
	Editorial	nombre	
2. Fuentes y Fecha	Número de fuentes referidas		
	Año de publicación		
3. Datos del(os) autor(es)	Autor 1		
	Autor 2		
	Autor 3		
	Autor 4		

Anexo 02. Guía de entrevista

Entrevista 1

Dirigido: Al Oncólogo Marco Calderón Espil

1. **¿Qué complicaciones se presentan para realizar el diagnóstico del cáncer de mama?**
2. **¿Cuánto tiempo demora en la entrega de resultados mamográficos?**
3. **Al realizar el diagnóstico que tiene en cuenta**
4. **¿Qué síntomas más frecuentes presenta un paciente que tiene cáncer de mama?**
5. **¿Al detectar que un paciente presente esas características de tener cáncer de mama, cual es el procedimiento para detectar al 100% esta enfermedad?**
6. **¿Usted conoce o utiliza algún software para apoyar al diagnóstico del cáncer de mama aparte de realizar la mamografía?**

Entrevista 2

Dirigido: Al radiólogo del área de oncología

1. **¿A qué se debe el por qué usted no está cumpliendo su función de analizar las mamografías de los pacientes?**
2. **¿Qué problema puede presentar al analizar una mamografía?**
3. **¿Con que instrumento usa para categorizar el cáncer de mama?**

Anexo 03. Lista de cotejo

SISTEMA INTELIGENTE PARA APOYAR AL ANÁLISIS MAMOGRAFICO EN LA DETECCIÓN DE TUMORES DE MAMA FEMENINO ENTRE LAS EDADES DE 40 A 60 AÑOS EN EL HOSPITAL "LAS MERCEDES" DE CHICLAYO

El presente documento se realiza con el objetivo de obtener una validación aprobatoria de resultados obtenidos de las mamografías que analiza el sistema inteligente vs con el analisis que realiza el radiólogo en el hospital "Las Mercedes".

Este instrumento sirve para medir el porcentaje de exactitud y precisión del sistema inteligente. Para ello se le entrega 70 imágenes que serán analizadas por el radiólogo encargado.

INSTRUMENTO DE VALIDACIÓN DE MAMOGRAFÍA DIGITALES

Nombre: _____

Apellido: _____

Fecha: _____

Cargo: _____

CMP: _____

Mamografía	Diagnóstico del experto		Diagnóstico del sistema inteligente	
	Mamografía que NO presenta tumores (SANA)	Mamografía que presenta tumores (ANORMAL)	Mamografía que NO presenta tumores (SANA)	Mamografía que presenta tumores (ANORMAL)
01				
02				
03				
04				
05				
06				
07				
08				
09				
10				
11				
12				
13				

14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				

62				
63				
64				
65				
66				
67				
68				
69				
70				

Leyenda:

		Valores predichos	
		Sí	No
Valores reales	Sí	TP	FN
	No	FP	TN

Tabla 1: Matriz de confusión de ejemplo.

Índice de Exactitud

- Excelente = >100% (70)
- Aceptable= Entre 80%-99% (Entre 56 a 69)
- Deficiente= Menor de 80% (Menor de 56)

Índice de Precisión

- Excelente = >100% (70)
- Regular= Entre 80%-99% (Entre 56 a 69)
- Deficiente= Menor de 80% (Menor de 56)

Firma del radiólogo

Para complementar este instrumento de validación se elaboró un formulario en Google para adquirir el análisis correspondiente del experto de las 70 mamografías.

Instrumento de validación de mamografía digitales

El presente documento se realiza con el objetivo de obtener una validación aprobatoria de resultados obtenidos de las mamografías que analiza el sistema inteligente vs con el análisis que realiza el radiólogo del hospital "Las Mercedes".

Indicación: Marque una de las opciones que corresponda a la mamografía analizada


Nombre *

Texto de respuesta corta

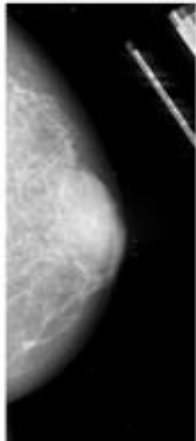
Apellidos *

Texto de respuesta corta

Fecha: *

Mes, día, año 


Mamografía 01 *



Mamografía que NO presenta tumores (SANA)

Mamografía que presenta tumores (ANORMAL)


Anexo 04. Instrumento de validación de la funcionalidad del sistema

	SISTEMA INTELIGENTE PARA APOYAR AL ANÁLISIS MAMOGRAFICO EN LA DETECCIÓN DE TUMORES DE MAMA FEMENINO ENTRE LAS EDADES DE 40 A 60 AÑOS EN EL HOSPITAL "LAS MERCEDES" DE CHICLAYO	Fecha de emisión:	
		Versión:	1.0
		Núm de pag:	1

Documento para medir la funcionalidad	
Elaborado por:	José Cristhian Ramón Cueva
Revisado por:	Ing. Chavarry Chancay Mariana
Aprobador por:	Ing. Chavarry Chancay Mariana
Objetivo de la ficha	
Calificar la funcionalidad del sistema inteligente web local	
Procedimiento	
El experto debe evaluar la funcionalidad del Sistema inteligente web local	
Nombre del experto:	Ing. Zelada Valdivieso Hector Miguel
Indicación:	Califique la funcionalidad del Sistema Inteligente web local
Criterios de evaluación:	Si: Se marca cuando el requerimiento indicado se esté cumpliendo en un 100% No: Se marca cuando el requerimiento indicado no se está cumpliendo en su totalidad

Para el módulo de inicio de sesión, responda la siguiente pregunta			
nº	Pregunta	Criterio	
		Si	No
1	¿Se validó el inicio de sesión? Ejemplo: Que solo permita el acceso al administrador y radiólogo, para ello debe de ingresar su Nick y clave.	x	
Para el módulo usuario, responda la siguiente pregunta			
nº	Pregunta	Criterio	
		Si	No
2	¿Se validó el registro de usuario ingresando todos los campos requeridos? Esto se encuentra en el menú del administrador, ítem usuario	x	
3	¿Se guarda los datos editados del usuario correctamente? Esto se encuentra en el menú del administrador, haciendo clic al nombre del usuario	x	
4	¿Se lista los datos guardados del usuario? Esto se encuentra en el menú del administrador, ítem usuario	x	
5	¿Permite un bloqueo manual del usuario seleccionado, cambiándole al icono del candado de color verde a rojo? Esto se encuentra en el menú del administrador, ítem usuario	x	
6	¿Permite activar al usuario seleccionado, cambiándole al icono del candado de color rojo a verde? Esto se encuentra en el menú del administrador, ítem usuario, dentro del listado	x	
7	¿Permite cambiar de cargo a un usuario? Esto se encuentra en el menú del administrador, ítem usuario, en el listado del usuario de la parte inferior	x	
8	¿Permite buscar al usuario ingresando su nombre en la opción de búsqueda? Esto se encuentra en el menú del administrador, ítem usuario, en el listado del usuario de la parte inferior	x	
Para el módulo paciente, responda la siguiente pregunta			
nº	Pregunta	Criterio	
		Si	No

9	¿Permite validar que no haya duplicidad con el número de historia y DNI del paciente? Esto se encuentra en el menú del administrador, ítem paciente	x	
10	¿Registra correctamente al paciente ingresando todos los campos requeridos? Esto se encuentra en el menú del administrador, ítem paciente, sub-ítem registrar	x	
11	¿Se guarda los datos editados del paciente correctamente? Esto se encuentra en el menú del administrador, ítem paciente, en la lista de pacientes	x	
12	¿Se lista los datos guardados del paciente? Esto se encuentra en el menú del administrador, ítem paciente, sub-ítem registrar en la parte inferior	x	
13	¿Permite buscar al paciente ingresando su nombre, DNI, todos los campos que están en la lista, en la opción de búsqueda? Esto se encuentra en el menú del administrador, ítem paciente, en el listado del paciente de la parte inferior	x	
14	¿Filtra los pacientes pendientes que estén listos para cargar sus mamografías? Esto se encuentra en menú del administrador, ítem paciente, sub-ítem pacientes pendientes	x	
15	¿Registra de manera correcta las mamografías digitales del paciente? Esto se encuentra en menú del administrador, ítem paciente, sub-ítem pacientes pendientes, columna cargar	x	
16	¿Solo permite registrar mamografías con formato JPG o PNG? Esto se encuentra en menú del administrador, ítem paciente, sub-ítem pacientes pendientes, columna cargar	x	
17	¿Permite eliminar una mamografía si en caso no le pertenece al paciente? Esto se encuentra en menú del administrador, ítem paciente, sub-ítem pacientes pendientes, columna cargar	x	
18	¿Filtra los Pacientes Listos para ser evaluados respecto a sus mamografías registradas? Esto se encuentra en menú del administrador, ítem paciente, sub-ítem pacientes listos	x	
19	¿Permite realizar una copia de seguridad de la base de datos? Esto se encuentra en menú del administrador, ítem Backup	x	
20	¿Se visualiza un dashboard con gráficos estadísticos? Esto se encuentra en menú del administrador, ítem Dashboard	x	
Para el módulo evaluaciones, responda la siguiente pregunta			
nº	Pregunta	Criterio	
		Sí	No
21	¿Filtra los pacientes a evaluar? Esto se encuentra en menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes a evaluar	x	
22	¿El sistema retorna los resultados de cada mamografía evaluada? Esto se encuentra en menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes a evaluar, dando clic en la columna Evaluar	x	
23	¿Registra correctamente la Evaluación ingresando todos los campos requeridos? Esto se encuentra en el menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes a evaluar, dando clic en la columna Evaluar	x	
24	¿Filtra los pacientes evaluados? Esto se encuentra en menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes evaluados	x	
25	¿Permite filtrar por fechas a los pacientes evaluados? Esto se encuentra en menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes evaluados	x	
26	¿Permite buscar los pacientes ingresando, todos los campos que aparecen en la tabla, en la opción de búsqueda? Esto se encuentra en el menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes evaluados	x	
27	¿Permite exportar el reporte de los resultados de cada paciente? Esto se encuentra en el menú del radiólogo, ítem Evaluaciones, sub-ítem pacientes evaluados	x	

Observaciones			
Sugerencia: Mejorar la presentación de los resultados. Siendo una de las funcionalidades más importantes, debería mostrarse de una manera más clara y que resalte la evaluación (pronóstico) que hace el mismo sistema.			
RESULTADOS:			
Mamografía 1: 99.199% de NO Tener TUMORES			
Mamografía 2: 93.056% de NO Tener TUMORES			
Mamografía 3: 99.999% de NO Tener TUMORES			
Mamografía 4: 99.688% de NO Tener TUMORES			
Yo como experto, doy conformidad de que se evaluó la presente ficha de manera fidedigna.			
Firma del experto:		Firma del tesista	

Anexo 05. Constancia de aprobación del producto acreditable de la entidad donde se ejecutó la tesis



GOBIERNO REGIONAL DE LAMBAYEQUE
Gerencia Regional de Salud
HOSPITAL "LAS MERCEDES"
CHICLAYO



CONSTANCIA DE APROBACIÓN DE PRODUCTO

ACREDITABLE

HOSPITAL REGIONAL DOCENTE
LAS MERCEDES

Por medio de la presente certificamos la aprobación de la aplicación web local presentada por el Estudiante **JOSE CRISTHIAN RAMON CUEVA** de la carrera profesional de Ingeniería de Sistemas y Computación de la **"UNIVERSIDAD CATOLICA SANTO TORIBIO DE MOGRIVEJO"**, como producto acreditable de su trabajo de investigación de fin de grado. Cumpliendo todos los requisitos planteados.

Se expide la presente constancia a petición del interesado para los fines que crea convenientes.

Atentamente:

DR. ROQUE FERNÁNDEZ MUNDACA
Medico Radiólogo
CMP: 16374 RNE: 7685

Tradición que protege tu salud!!!

Chiclayo, 14 de mayo de 2021