



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Universitat Politècnica de València.

Máster Universitario en Inteligencia artificial, reconocimiento de formas e imagen digital.

Departamento de Sistemas Informáticos y Computación.

Trabajo Fin de Máster

**Sistema para la clasificación y reconocimiento de imágenes de ultrasonido en tiroides, basado en técnicas de aprendizaje profundo para el apoyo en el proceso de diagnóstico según la escala EU-TIRADS.**

Autor: Yhary Estefania Arias Trillos.

Tutor: Roberto Paredes Palacios.

Dir. Experimental: Juan Maroñas Molano.

Departamento: Departamento de Sistemas Informáticos y Computación.

Curso: 2021-2022

## Dedicatoria

Quiero expresar mi gratitud a Dios y a toda mi familia por estar siempre presentes, en especial a mi madre quien es mi mayor ejemplo a seguir.

A Freyman Balaguera, quien dedicó su tiempo a apoyarme y a enseñarme todo lo que ahora sé.

Mi profundo agradecimiento docente Roberto Paredes, mi director experimental Juan Maroñas, a los Doctores Federico Lubinus, Boris Marconi y la clínica Foscál. Por confiar en mí, abrirme las puertas, permitiéndome realizar todo el proceso investigativo y por el tiempo dedicado al proyecto.

“El presente documento ha sido realizado completamente por el firmante; no ha sido entregado como otro trabajo académico previo y todo el material tomado de otras fuentes ha sido convenientemente entrecomillado y citado su origen en el texto, así como referenciado en la bibliografía”

### **Trabajo de fin de máster**

Máster universitario en inteligencia artificial reconocimiento de formas e imagen digital.

**Título:** Sistema para la clasificación y reconocimiento de imágenes de ultrasonido en tiroides, basado en técnicas de aprendizaje profundo para el apoyo en el proceso de diagnóstico según la escala EU-TIRADS.

Autor: Yhary Estefania Arias Trillos.

Director: Roberto Paredes Palacios.

Director Experimental: Juan Maroñas Molano.

Colaboradores: Doc. Federico Lubinus (Clínica Foscal).

Doc. Boris Marconi (Clínica Foscal).

## Resumen

El presente trabajo de fin de máster, expone el desarrollo de un sistema que permite la clasificación de anomalías en imágenes digitales de ultrasonido de la glándula tiroidea, utilizando una prueba de tamizaje (discriminatoria) Europea, conocida como EU- TIRADS, la cual está compuesta por cinco clases y se usa para la categorización del nódulo tiroideo. Con base en la investigación realizada y en apoyo de la Clínica Foscal en Colombia, quienes proporcionaron las imágenes, se lleva a cabo un prototipo para la automatización del proceso de clasificación a través de algoritmos de aprendizaje profundo en uso de Redes Neuronales, que han sido probadas con éxito para la clasificación de otro tipo de imágenes médicas. Con 1353 imágenes, 5 clases, un modelo de red neuronal convolucional como inicio del proceso de experimentación y aplicando transfer learning con diferentes modelos pre-entrenados, que se detallan en el desarrollo del documento, se logró obtener un Accuracy de 55% y una pérdida del 1.16, teniendo en cuenta que es una clasificación de 5 clases, la probabilidad de asignar una clase correcta a una imagen es del 20%, lo que quiere decir que el modelo es 27.5% mejor que elegir al azar. En vista del proceso realizado y los resultados obtenidos con el dataset construido desde cero y el proceso que se ha llevado a cabo, se espera continuar con la investigación e implementación a futuro, teniendo como base los resultados obtenidos en los experimentos

realizados, donde se evidencia que al aumentar la data el proceso de predicción mejora significativamente, sin embargo esto también dependerá de factores como la calidad de los datos y un buen etiquetado.

**Palabras claves:** EU-TIRADS, tiroides, redes neuronales convolucionales, nódulo, ultrasonido, aprendizaje profundo, CNN, Transductor, Transfer Learning.

## Abstract.

This master's thesis presents the development of a system that allows the classification of anomalies in digital ultrasound images of the thyroid gland, using a European screening test (discriminatory), known as EU-TIRADS, which is composed of five classes and is used for the categorization of the thyroid nodule. Based on the research carried out and with the support of the Foscal Clinic in Colombia, who provided the images, a prototype is carried out for the automation of the classification process through deep learning algorithms using Neural Networks, which have been successfully tested for the classification of other types of medical images. With 1353 images, 5 classes, a convolutional neural network model as the beginning of the experimentation process and applying transfer learning with different pre-trained models, which are detailed in the development of the document, it was possible to obtain an Accuracy of 55% and a loss of 1.16, taking into account that it is a classification of 5 classes, the probability of assigning a correct class to an image is 20%, which means that the model is 27.5% better than choosing at random. In view of the process performed and the results obtained with the dataset built from scratch and the process that has been carried out, it is expected to continue with the research and implementation in the future, based on the results obtained in the experiments performed, where it is evident that by increasing the data the prediction process improves significantly, however this will also depend on factors such as data quality and good labeling.

**Keywords:** EU-TIRADS, thyroid, convolutional neural networks, nodule, ultrasound, deep learning, CNN, Transducer, Transfer Learning.

## Índice general.

1. Introducción.	1
2. Objetivos.	2
2.1. Objetivo general.	2
2.2. Objetivos específicos.	3
2.3. Razonamiento del interés.	3
2.4. Pertinencia del trabajo.	4
2.5. Hipótesis de partida.	4
2.6. Aportaciones que se esperan alcanzar.	5
3. Estado del arte.	5
3.1. Artículos.	6
3.2. Tabla de proyectos.	11
3.3. Tabla de aplicaciones	12
4. Datos.	14
5. Metodología.	16
5.1. Comprensión del negocio.	18
5.2. Comprensión de los datos.	18
5.2.1. Escala usada para la clasificación.	21
5.3. Recopilación de datos.	22
5.3.1. Pre procesamiento de los datos.	24
5.4. Modelado.	27
5.4.1. Implementación del modelo CNN.	28
5.4.2. Implementación del modelo VGG.	35
5.4.3. Implementación del modelo Xception.	38

5.4.4. Implementación del modelo ResNet.	41
5.5. Evaluación	43
5.6. Despliegue.	45
5.6.1. Identificación de requisitos.	46
5.6.2. Requerimientos funcionales.	46
5.6.3. Requerimientos no funcionales.	47
5.6.4 Herramientas.	48
6. Experimentos.	48
7. Presupuesto.	50
8. Conclusiones.	52
9. Anexos.	53
9.1. Página web para el despliegue del modelo.	53
9.2. Página web para el etiquetado de los datos.	57
10. Referencias.	58

### **Índice de tablas.**

Tabla 1. Cantidad de exámenes según el género.	4
Tabla 2. Rendimiento de la clasificación de nódulos tiroideos basada en la ecografía utilizando TIRADS.	7
Tabla 3. Mediciones cuantitativas de los resultados de segmentación utilizando todas las características.	10
Tabla 4. Tabla de resultados del pre entrenamiento, PVV Prediction of peak particle velocity, NPV Net Present Value.	11

Tabla 5. Tabla de proyectos.	12
Tabla 6. Tabla de aplicaciones.	14
Tabla 7. Cantidad de imágenes por clase, conjunto de entrenamiento.	25
Tabla 8. Cantidad de imágenes por clase, conjunto de validación.	25
Tabla 9. Cantidad de imágenes por clase, conjunto de pruebas.	26
Tabla 10. Resultados del primer entrenamiento con CNN.	32
Tabla 11. Reporte de clasificación.	33
Tabla 12. Resultados del entrenamiento con el modelo VGGT, sin congelar el modelo.	37
Tabla 13. Resultados del entrenamiento con el modelo VGGT, sin congelar el modelo.	37
Tabla 14. Resultados obtenidos del entrenamiento del modelo Xception (sin congelar la base del modelo).	40
Tabla 15. Resultados obtenidos del entrenamiento del modelo Xception (congelando la base del modelo).	40
Tabla 16. Resultados obtenidos del entrenamiento del modelo ResNet (sin congelar la base del modelo).	42
Tabla 17. Resultados obtenidos del entrenamiento del modelo ResNet (congelando la base del modelo).	43
Tabla 18. Arquitectura de cada modelo con su evaluación.	45
Tabla 19. Requerimientos funcionales.	47

Tabla 20. Requerimientos funcionales	47
Tabla 21. Requerimientos no funcionales.	47
Tabla 22. Requerimientos no funcionales.	47
Tabla 23. Experimentos, arquitectura y evaluación.	49
Tabla 24. Recursos materiales.	50
Tabla 25. Recurso humano.	50
Tabla 26. Otros recursos.	50

### **Índice de ilustraciones.**

Ilustración 1. Cantidad de imágenes por clase.	16
Ilustración 2. Metodología CRISP-DM.	17
Ilustración 3. Región anterior superficial del cuello.	19
Ilustración 4. Página web para el etiquetado de las imágenes.	23
Ilustración 5. Imágenes de Tiroides con su etiqueta.	27
Ilustración 6. Arquitectura del modelo CNN construida para la clasificación de 5 clases.	29
Ilustración 7. Gráficas de los resultados de evaluación con respecto a las épocas.	33
Ilustración 8. Arquitectura VGG16.	36

Ilustración 9. Evaluación del entrenamiento con VGG16, Accuracy vs épocas y Loss vs épocas (sin congelar el modelo).	37
Ilustración 10. Evaluación del entrenamiento con VGG16, Accuracy vs épocas y Loss vs épocas (Congelando el modelo).	37
Ilustración 11. Arquitectura de Xception.	39
Ilustración 12. Gráfica del Accuracy y Loss con respecto a las épocas.	40
Ilustración 13. Gráfica de Xception, Accuracy y Loss con respecto a las épocas.	41
Ilustración 14. Gráfica del Accuracy y Loss con respecto a las épocas, sin congelar la base del modelo.	42
Ilustración 15. Gráfica del Accuracy y Loss con respecto a las épocas, congelando la base del modelo.	43
Ilustración 16. Interfaz de usuario (Despliegue del modelo).	54
Ilustración 17. Cargar imágenes.	55
Ilustración 18. Predicción según su escala de clasificación	56
Ilustración 19. Información técnica del proyecto.	56
Ilustración 20. Página web para el etiquetado de las imágenes de tiroides.	57

## **1. Introducción.**

Según un estudio realizado por el centro de proyectos para el desarrollo de la Universidad Javeriana, actualmente el país cuenta con un promedio de 72.544 médicos, de los cuales 30.759 (42,4%) tiene estudios de posgrado registrados formalmente. Los datos indican que en Colombia se tienen 26 médicos por cada 100.000 personas, dejando en evidencia el nivel de cobertura que hay en todo el territorio nacional. En lo que respecta el área de radiología en Colombia, se han registrado 869 según la dirección de epidemiología y demografía y la dirección talento humano en salud, es decir, el 2,3% de los médicos con estudios de posgrado en Colombia.

En lo que respecta a la radiología, registran 756 especialistas de diferentes ámbitos, es decir, el 2,3% son médicos con estudios de posgrado [1]. Debido a la escasez de especialistas, la alta demanda en procedimientos para la interpretación de imágenes y el alto costo de los exámenes que se realizan, existe una creciente necesidad de métodos y sistemas de análisis de exámenes médicos que de manera ágil obtengan resultados precisos en el menor tiempo posible. En búsqueda de una solución a esta problemática, se desarrolló un sistema con modelos de Deep Learning, la construcción de un dataset de imágenes en formato DICOM, en el proceso para la medicina predictiva, con el fin de extraer y clasificar categóricamente ultrasonidos de tiroides, utilizando un prueba de tamizaje EUTIRADS, que con el uso de un algoritmo supervisado en sistemas de diagnóstico y análisis, hará importantes mejoras en la predicción de enfermedades o clasificación de patologías.

Ahora bien, con el fin de mejorar el entrenamiento y predicción, se aumentaron los datos y se pusieron a prueba 3 modelos basados en CNN, como Transfer Learning con el propósito de mejorar los niveles de sensibilidad, los cuales se refieren a la probabilidad de que una prueba identifique como persona enferma a aquél que efectivamente lo está. A su vez el tema de la especificidad, que es la probabilidad de que la una prueba realizada a un paciente, identifique como no enfermo a aquél que efectivamente no lo está.

Debido a que las imágenes recolectadas fueron clasificadas manualmente, esta tarea tomó gran parte del tiempo del desarrollo del proyecto, para agilizar el etiquetado, se llevaron a cabo varias estrategias, una de estas fue el desarrollo de una página web con una vista dinámica, donde se alojaron las imágenes junto con su etiqueta. Finalmente, el modelo se desplegó en un sistema alojado en la web para el uso exclusivo del personal médico, con políticas de privacidad, teniendo en cuenta la sensibilidad de los datos, este proyecto tiene la intención de generar un apoyo en el proceso de clasificación y diagnóstico en exámenes de TSH (hormona estimulante de la tiroides) y/o punción aspirativa con aguja fina (PAAF) en tiroides.

## **2. Objetivos.**

### **2.1. Objetivo general.**

Construir un dataset de imágenes médicas de tiroides y desarrollar un sistema para la clasificación de las mismas, basado en la prueba de

tamizaje EU-TIRADS, a través de redes neuronales convolucionales, para brindar asistencia en el proceso de diagnóstico médico.

## **2.2. Objetivos específicos.**

1. Realizar una investigación de las anomalías presentes en tiroides y forma en la cual se genera el diagnóstico dado por el especialista.
2. Desarrollar una página web para llevar a cabo el etiquetado manual de imágenes realizado por los médicos.
3. Construir un dataset con imágenes etiquetadas en 5 clases.
4. Implementar un modelo de red neuronal convolucional para la clasificación de las imágenes.
5. Validar las técnicas usadas para la clasificación implementando técnicas de evaluación.
6. Desplegar el modelo funcional en GitHub Pages, con una interfaz visual, generando una herramienta que permita cargar una imagen y arrojar una predicción de la clase correspondiente, en apoyo al procedimiento de clasificación y diagnóstico de la glándula tiroidea.

## **2.3. Razonamiento del interés.**

El 92.6% de los casos en estudio con posible diagnóstico de nódulos tiroideos, en su mayoría han sido identificados en mujeres (ver tabla 1).

Otro de los factores importantes a mencionar es la falta de tratamientos

oportunos en este tipo de procedimientos y la importancia de realizar un seguimiento del estado del paciente de forma periódica para identificar el estado de la glándula tiroidea y lograr actuar a tiempo, debido a que el nódulo tiroideo tiende a crecer si no se médica y mejoraran hábitos de forma oportuna. (Esta investigación está basada en los datos propios que se obtuvieron para entrenar el modelo).

<b>Ultasonidos</b>	<b>Ultasonidos</b>	<b>PAAF</b>	<b>PAAF</b>
Mujer	Hombre	Mujer	Hombre
795	10	548	0

Tabla 1. Cantidad de exámenes según el género.

#### **2.4. Pertinencia del trabajo.**

Según la sociedad Americana Contra el Cáncer [2], es posible que se diagnostique un estimado de 43,800 nuevos casos de cáncer de tiroides para 2022, presente mayoritariamente en mujeres ya que según sus investigaciones las mujeres tienen 3 veces más posibilidades de padecer cáncer de tiroides que los hombres, información que se valida los datos que se tienen para esta investigación.

Para generar una mejora al proceso del diagnóstico realizado en exámenes de TSH donde se toma una muestra del nódulo por medio de una aguja fina e imágenes de ultrasonido, se requiere generar un sistema que brinde apoyo a los especialistas en radiología en la toma de decisiones con el objetivo de optimizar procesos y generar diagnósticos oportunos.

## **2.5. Hipótesis de partida.**

Con la alta demanda de exámenes TSH y PAAF en tiroides presente en la clínica Foscal (Fundación Oftalmológica de Santander), se requiere una herramienta como apoyo al diagnóstico médico, que permita generar resultados a tiempo, con el fin de realizar un control periódico al paciente. Por consiguiente, se requiere un sistema de Deep Learning para la clasificación de ultrasonidos, basado en la escala EU-TIRADS, que como base mínima del proceso de aprendizaje y predicción, alcance un Accuracy por encima del 50%, debido a la complejidad de las imágenes.

## **2.6. Aportaciones que se esperan alcanzar.**

Una herramienta de fácil acceso que facilite el etiquetado de imágenes médicas con el fin de construir un dataset con el que será entrenado un modelo de red neuronal y que finalmente se pondrá a disposición de la comunidad médica para mejorar el servicio médico y atención a cada uno de los pacientes, brindando un diagnóstico en el menor tiempo posible de manera precisa, para ello, se brindará una herramienta de uso privado para especialistas en radiología de la clínica Foscal, montado en una plataforma web al que solo tendrá acceso el área de radiología, en esta

plataforma web, se aloja el modelo listo para predecir la clase a la cual pertenece dicha imagen que se ha cargado.

### **3. Estado del arte.**

Para el desarrollo de la investigación sobre las propuestas, prototipos o artículos que se han llevado a cabo en el área de investigación que este documento propone, es decir, investigaciones con base a la clasificación de imágenes médicas de tiroides, se describen las siguientes propuestas existentes, sus avances, resultados obtenidos, y semejanzas al proyecto en curso, que pretende proponer nuevos experimentos y con propuesta innovadora, teniendo en cuenta la diferencia y originalidad.

Es pertinente dar a conocer proyectos e investigaciones que se han realizado a lo largo del tiempo en base al proyecto que se está llevando a cabo. Estas algunas investigaciones claves que han servido como inspiración:

#### **3.1. Artículos.**

**Artículo 1: “Clasificación ecográfica de los nódulos tiroideos: Una Revisión Sistemática”.**

Para el sistema de clasificación de ecografías basado en asignación de clases teniendo en cuenta nódulos tiroideos en función del riesgo de malignidad, el objetivo de esta investigación fue evaluar la literatura del rendimiento de sistemas de clasificación de

ecografías de nódulos tiroideos mediante la correlación que existe con el examen de biopsia por aspiración con aguja fina (BAAF).

Se demostró que la escala TIRADS tiene muy buena utilidad para clasificar nódulos tiroideos y se menciona los traumas que sufren algunos pacientes en el procedimiento, por consiguiente se propone centrar la investigación en cuantificar el dolor y el estrés que sufren los pacientes que son sometidos a una prueba FNAB ya que el NPV (Negative Prediction Value) del TIRADS puede compensar el impacto del VPP(Positive Prediction Value) y ayudar a reducir el BAAF que es innecesario en nódulos benignos. Por tanto el artículo propone mejorar el sistema de TIRADS realizando estudios en gran población de pacientes y así comparar el sistema TIRADS con otros sistemas de clasificación con el fin de demostrar su superioridad [3].

A continuación en la tabla 2, se observan los resultados de la investigación planteada en el artículo.

<b>Parámetro de rendimiento</b>	<b>Participantes (nódulos)</b>	<b>Mediana (%)</b>	<b>Rango (%)</b>
Sensibilidad	5360	90	70.6-97.4
Especificidad	5360	57,4	29.3-90.4
PPV (Valor predictivo positivo)	5632	49	23.3-64.3
NPV (Valor predictivo negativo)	5360	91	23.3-64.3

Tabla 2. Rendimiento de la clasificación de nódulos tiroideos basada en la ecografía utilizando TIRADS.

**Aportes:** Importancia de realizar la clasificación de ultrasonidos de tiroides con el método TIRADS, debido a su efectividad y especificidad, con el fin de evitar realizar procedimientos invasivos y prolongados como BAAF a nódulos benignos.

**Artículo 2: “Clasificaciones ecográficas de los nódulos tiroideos para la estratificación del riesgo de malignidad y la gestión clínica”.**

Este estudio se encarga de evaluar que tan efectivos son los exámenes de ecografías de nódulos tiroideos son indicadores precisos y útiles a la hora de identificar malignidad histológica. Las investigaciones realizadas han identificado pocas características en ecografías que tienden a ser frecuentes en nódulos tiroideos, estos resultados pueden ser usados como indicador un mayor riesgo de malignidad.

En el artículo propuesto se describen los principales sistemas de clasificación de nódulos tiroideos que han sido publicados en los últimos años en base a investigaciones y experimentos. Este artículo concluye que todos los sistemas de estratificación de riesgo de malignidad por ultrasonidos, demuestran una alta sensibilidad y especificidad, pero según los autores revelan que los resultados obtenidos no son suficientes para establecer cuál sería el mejor sistema para identificar malignidad y se propone realizar experimentos en un centro médico que realice este tipo de exámenes de manera habitual [4].

**Aportes:** Teniendo en cuenta que aún que no se obtuvieron resultados positivos, identificamos que la falta de datos y experimentación puede ser clave a la hora de tomar

decisiones y sacar conclusiones. Este tipo de estudios puede tomar mucho tiempo y requerir de un equipo de trabajo de experto en el área y un centro médico dispuesto a prestar servicios de investigación y experimentación.

**Artículo 3: “Una revisión sobre la caracterización del tejido del cáncer de tiroides basada en la ecografía y la clasificación automatizada”.**

Este artículo se basa en la investigación de los estudios realizados por el CAD o Diagnostico Asistido por Ordenador, para la clasificación automatizada de cáncer de tiroides tanto en nódulos benignos y malignos, donde se tienen en cuenta las diferentes características que intervienen a la hora de dar un diagnostico basados en el estudio realizado donde se utilizan técnicas estadísticas y se hace minería de datos.

Destacando las limitaciones de dichos estudios, teniendo en cuenta las técnicas ya que se utilizaron características no clínicas para la clasificación y el reporte de precisiones que obtuvieron en esta investigación fue que las técnicas de clasificación automatizada son competidoras potenciales de los exámenes FNAB(Biopsia por aspiración con aguja fina) y que podrán utilizarse como herramienta complementaria a la hora de generar un diagnostico al paciente como una segunda opinión para el especialista [5].

**Aportes:** las ventajas de este tipo de investigaciones proponen el uso de herramientas que beneficien al sector médico y a los pacientes, quienes se ven beneficiados por la eficacia, precisión y mejoras en el tiempo de diagnóstico, con la finalidad de actuar preventivamente.

**Artículo 4: “Clasificación y segmentación de la tiroides en imágenes ecográficas mediante algoritmos de aprendizaje automático”.**

La investigación de este artículo gira en torno al diagnóstico por imagen con técnicas como exámenes de ecografía y TSH que viene siendo un examen más invasivo, por consiguiente se propone dar apoyo a especialistas para que se agilicen los procesos de diagnóstico. Los exámenes de ultrasonido siguen contribuyendo significativamente a la atención y control del paciente. Lo que pretende esta investigación es desarrollar un sistema automático que clasifique las imágenes de tiroides y segmente la glándula y sus nódulos mediante algoritmos de visión por computador. En este trabajo se ponen a prueba clasificadores como Support Vector Machine y Extreme Learning Machine que tienen varias características estadísticas que potencian el proceso de clasificación, realizando una comparación con los resultados reales hechos por el radiólogo [6].

<b>Método de segmentación</b>	<b>Accuracy</b>
SVM (support vector machine)	84.78
ELM (extreme learning machine)	93.56

Tabla 3. Mediciones cuantitativas de los resultados de segmentación utilizando todas las características.

**Artículo 5: “Clasificación de nódulos tiroideos por ultrasonidos mediante una red neuronal convolucional profunda con patología quirúrgica”.**

Según la investigación propuesta, se identificó que los ultrasonidos de tiroides son propensos a interpretaciones subjetivas y variabilidades entre observadores. Este

proyecto tiene como objetivo desarrollar un sistema de clasificación de nódulos tiroideos, aplicando redes neuronales convolucionales, con exámenes de 726 pacientes para crear un modelo de Deep Learning. Después de realizar el examen de biopsia, se confirmaron 325 casos de nódulos benignos y 437 casos de malignos. Se aplica el modelo VGG16 para el transfer learning para realizar el pre procesamiento. Cabe resaltar que se hace uso de la escala TIRAD con las siguientes clases:

- Escala TIRAD: 1, 2, 3, 4a, 4b, 5.

La cual se divide en dos: Clases positivas (TI-RAD 3, 4a, 4b, 5) y clases negativas (TI-RAD 1 o 2), es decir que la clasificación terminación binaria con un dataset de 1524 imágenes. Los resultados de la siguiente Tabla 4 Obtenidos en la investigación cabe resaltar que son bastante buenos y significativos para la investigación en curso [7].

AUC	Sensibilidad	Especificidad	PPV	NPV
0.916	0.92	0.70	0.79	0.87
0.90-0.92	0.83-0.96	0.61-0.89	0.71-0.88	0.84-0.94

Tabla 4. Tabla de resultados del pre entrenamiento, PVV Prediction of peak particle velocity, NPV Net Present Value.

**Aportes:** Este proyecto planteado en el artículo anteriormente analizado, es clave en nuestra investigación ya que tiene el mismo enfoque, la clasificación de imágenes de ultrasonido en tiroides con un modelo pre entrenado pero con clasificación binaria y un porcentaje de imágenes muy parecido, en el cual se evidencian resultados bastante

favorables que cabe resaltar, ha sido inspiración para nuestro proyecto que en lugar de 2 clases, se tienen 5 clases y se maneja la escala europea EU-TIRADS.

### 3.2. Tabla de proyectos.

Tabla de proyectos			
Nombre	Tema	Método	Resultados
Clasificación de imágenes hiperespectrales basada en la extracción de características espectrales espaciales.	Algoritmo de clasificación hiperespectral basado en la extracción de características espectrales espaciales.	Machine learning, IFRF Support Vector Machine (SVM).	Precisión de clasificación global mejorada y coeficientes Kappa que los algoritmos tradicionales y de última generación.
Using Nonnegative Matrix Factorization with Projected Gradient for Hyperspectral Images Feature Extraction.	Función de extracción de grandes volúmenes de imágenes de teledetección.	PGNMF (técnica de partición del anillo y la factorización de matriz no negativa de la gradiente proyectada).	Extracción de características de imágenes hiperespectrales mediante el uso de PGNMF.
Proyecto de clasificación usando soft computing	Clasificación de estos proyectos basados en la lógica difusa	Lógica difusa.	Clasificación de proyectos de software basado en el diagrama de actividad

Seguimiento de objetos visuales guiado por acción de DL por refuerzo.	Algoritmos de seguimiento visual.	ADNet, arquitectura CNN profunda.	El rastreador propuesto que controla por un ADNet que persigue iterativamente el objeto mediante acciones secuenciales.
---	-----------------------------------	-----------------------------------	---

Tabla 5. Tabla de proyectos.

### 3.3. Tabla de aplicaciones

Tabla de aplicaciones			
Nombre	Empresa	En que consiste	Aportes
Amazon Rekognition	Amazon	Facilita la incorporación del análisis de imágenes y videos a sus aplicaciones.	Uso de la tecnología que maneja esto para la clasificación de imágenes médicas.
Las AMI de aprendizaje profundo de AWS	Amazon	Aprendizaje automático la infraestructura y las herramientas necesarias para agilizar las tareas de aprendizaje profundo en la nube a cualquier escala.	Desarrollo de aprendizaje automático con diferentes tecnologías.
Amazon Polly	Amazon	Servicio que convierte texto en habla realista.	Reconocimiento de información que se le genere dando una salida óptima.
Amazon Translate	Amazon	Servicio de traducción automática neuronal que ofrece	Procesamiento de múltiples datos.

		traducción de idiomas.	
Watson	IBM	Servicio de traducción automática neuronal que ofrece traducción de idiomas.	Procesamiento de múltiples datos.
Cloud Vision	Google	Clasifica las imágenes rápidamente en miles de categorías, detecta objetos y caras determinados dentro de dichas imágenes y es capaz de leer las palabras impresas.	Procesamiento de datos en imágenes.
Watson Health	IBM	Proporcionar una atención eficaz y gestionar la salud de su población.	Aprendizaje automático, procesamiento de datos.

Tabla 6. Tabla de aplicaciones.

#### 4. Datos.

La clínica Foscal de la ciudad de Bucaramanga en Colombia y en colaboración el doctor Federico Lubinus y el Doctor Boris Marconi quienes participan activamente del proyecto, han proporcionado las imágenes para la construcción desde cero del datase en formato DICOM y JPG. Actualmente se tienen 1353 imágenes almacenadas en un disco duro de 1 Tera. Para cada lote de imágenes que se recolectó, fue necesario ir cada

cierto tiempo a la clínica y solicitar los permisos necesarios para ingresar a la sala donde se realizan los exámenes y tomar del ordenador las imágenes por semanas o por meses según la disponibilidad. Actualmente se cuenta con más de 5.000 imágenes de las cuales solo se logró clasificar 1353.

Es importante mencionar que es posible obtener más imagen 'es pero no se puede garantizar su clasificación debido al tiempo de cada uno de los colaboradores del proyecto, por ende vale la pena que entidades públicas o privadas quieran unirse a este proyecto y obtener una muy buena data y no solo las entidades, sino especialistas en radiología. El etiquetado de los datos se hizo de forma manual por medio de una herramienta web que se desarrolló desde cero para facilitar el etiquetado de imágenes que realizaron los doctores, fue una de las formas más efectivas para ejecutar esta tarea.

Después de construir el data set y realizar el procedimiento de etiquetado en **5 clases** discriminatorias basadas en la escala europea EU-TIRADS que fue propuesta por el doctor Federico Lubinus quién tiene años de experiencia en el área de radiología y fue director de la asociación Colombiana de radiología, con base en su experiencia, sugirieron usar dicha escala para el proceso de etiquetado.

Para el proceso de entrenamiento se tienen:

- 971 imágenes para entrenamiento.
- 243 imágenes de validación.
- 139 imágenes de prueba.



Ilustración 1. Cantidad de imágenes por clase.

## 5. Metodología.

Por la naturaleza de los datos y los resultados obtenidos, se aplica una metodología cuantitativa, que nos permite evaluar el rendimiento de los modelos aplicados y su efectividad, obteniendo datos medibles, en un proceso experimental, el cual es llevado a cabo desde la extracción y construcción del dataset, hasta el proceso de evaluación y despliegue.

Como se menciona en los capítulos 1 y 2, el problema que se quiere abordar, son las demoras en la asignación del diagnóstico de un paciente, después de un procedimiento TSH y PAAF, ya que es de suma importancia actuar a tiempo, evitando complicaciones y gastos extras tanto al paciente como al centro de salud. Esto por lo general es así debido a la demanda de exámenes de este tipo, que son más frecuentes en mujeres.

La metodología de desarrollo, que mejor se adapta a este tipo de problemas, es CRISP-DM, que plantea un modelo de procesos para la minería de datos. En la siguiente figura, podemos observar su arquitectura.

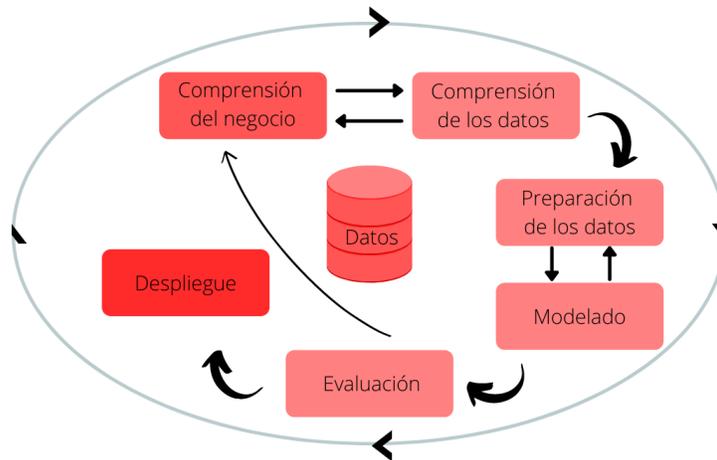


Ilustración 2. Metodología CRISP-DM.

CRISP-DM incluye un modelo con una estructura compuesta de seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior, en este caso nos regresamos a varias de las etapas descritas a continuación, para poder revisar, ajustar e incluso agregar más datos, por lo que la sucesión de fases no tiene por qué ser ordenada desde la primera hasta la última. En la figura 2 se puede observar las fases en las que se divide CRISP-DM y las posibles secuencias a seguir entre ellas.

A continuación se describen cada una de las fases en las que se divide CRISP-DM, aplicadas al proyecto de investigación y desarrollo para un sistema de clasificación en 5 clases, de imágenes de ultrasonido.

### **5.1. Comprensión del negocio.**

Es en esta etapa de la metodología en la que se plantea la herramienta que permitirá desarrollar cada uno de los objetivos planteados al inicio del documento.

Se identificó la necesidad de una herramienta alojada en la web, que permita la clasificación de ultrasonidos en tiroides, en tiempo real, que le permitirá al especialista tomar una decisión, acompañado de un sistema de Deep Learning, que apoya el diagnóstico.

### **5.2. Comprensión de los datos.**

Como primera etapa del proceso, para adquirir las imágenes médicas, estas se extraen de un eco grafo que si bien es la herramienta que permiten explorar el interior del cuello por medio de ondas sonoras de alta frecuencia. Genera imágenes de la glándula tiroidea por medio de ultrasonidos, siendo este un procedimiento seguro e indoloro. Este examen de ultrasonido o ecografía, utiliza un transductor y con el gel que se coloca directamente sobre la piel, permite desplazarlo por las siguientes zonas:

- Región parotídea.

- Región supra hioidea.
- Región infra hioidea.
- Región carotidea.

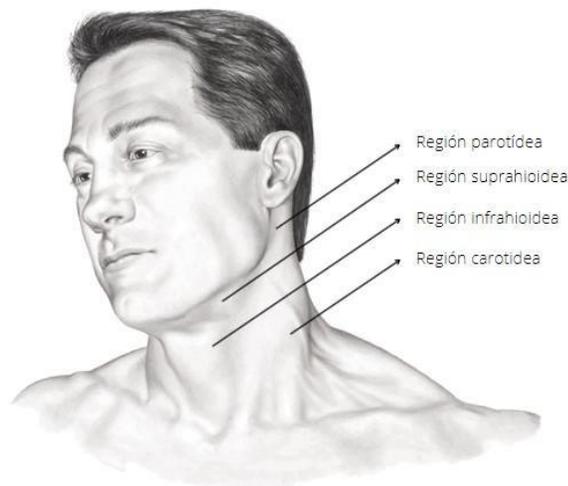


Ilustración 3. Región anterior superficial del cuello.

Fuente: <http://dea.unsj.edu.ar/biologia2/cuello.pdf>

Este tipo de procedimientos indoloros se realizan para descartar posibles anomalías que, por no hacer un diagnóstico oportuno, terminen en cáncer de tiroides. Según la asociación América de cáncer, cada año ha aumentado un 0.6% la tasa de mortalidad del cáncer de tiroides [9].

Comúnmente se suele diagnosticar en edades más tempranas y las mujeres tienen 3 veces más probabilidad de padecer cáncer de tiroides que los hombres. Para dar un

contexto más profundo de la investigación es necesario tener claro ciertos conceptos que se abordarán en el desarrollo del documento y la metodología planteada.

Cuando se ejecuta un procedimiento como PAAF, se extrae una muestra del nódulo, introduciendo una aguja fina en la zona específica que nos muestra el ecógrafo, por ello mientras se hace la extracción de dicha muestra, también se van haciendo tomas del nódulo con el transductor.

Las complicaciones que se identifican en este tipo de exámenes, tienen a afectar la producción de hormonas, generando cambios en el ritmo de actividad de nuestro cuerpo.

Las hormonas tiroideas son de importancia crítica para el crecimiento y el desarrollo fetales, regulan la frecuencia cardíaca y la función contráctil (una de las funciones de las proteínas encargada del movimiento de los músculos), influyen en la motilidad gastrointestinal, la función respiratoria y la eliminación renal del agua y también modulan el gasto energético del organismo, el peso y el metabolismo [10].

La función que cumple esta glándula, afecta todo nuestro metabolismo, por ende, es importante identificar a tiempo que tipo de bocio presenta el paciente.

Debido a que las imágenes se capturan en tiempo real, pueden mostrar la estructura y el movimiento de los órganos internos del cuerpo. También pueden mostrar la sangre fluyendo por los vasos sanguíneos, pero esto solo se obtiene con un visor especial para

el formato en el que se generan dichas imágenes, en este caso usamos “Lite” un visor de archivos DICOM. Este formato nos permite observar con la más alta calidad las diferentes tomas que se obtuvieron en un examen y toda la meta data que contienen dichas imágenes, con los datos personales de los pacientes, que para este proyecto se eliminaron, como protección a la información privada y anonimizando todas las imágenes adquiridas.

### **5.2.1. Escala usada para la clasificación.**

Para tener un concepto más claro sobre que son los archivos DICOM, según Clinic-Cloud [8]. "DICOM es un protocolo estándar de comunicación entre sistemas de información y a la vez un formato de almacenamiento de imágenes médicas". Una imagen médica por sí misma no aporta suficiente información, ya que hace falta la lectura de la misma, hecha por el médico. Para que sea correctamente interpretada, es necesario que vaya acompañada de las dimensiones o áreas del objeto a analizar o mapas de calor sobre la imagen, lo que facilita el estudio. El bocio puede presentarse como maligno o benigno, según sea el caso, es decir que está asociado a la superproducción de la hormona tiroidea.

Un diagnóstico médico después de dicho análisis puede estar compuesto de la siguiente forma (análisis benigno): Los siguientes datos son un ejemplo de un paciente real, claramente, su información personal no se registra, con el fin de proteger sus datos.

**Ejemplo de diagnóstico 1:** “Se observa glándula tiroides globalmente agrandada con presencia de un quiste coloide en cada lóbulo.”

**Ejemplo de diagnóstico 2:** “TIRADS 2. Bocio coloide, tiroiditis, nódulo hiperplástico. No identifico adenopatías ni masas cervicales.”

**Ejemplo de diagnóstico 3:** “TIRADS 3. Atipia de significado indeterminado. Se debe repetir la PAAF a un intervalo apropiado ya que un 20-25% de estas lesiones se volverán a clasificar en una segunda punción con Atipias de significado indeterminado, con un riesgo de ser una lesión maligna en un 5-10%”.

Después un análisis ya sea indoloro como lo es un ultrasonido de tiroides o el procedimiento PAAF que suele ser invasivo, aparte de la extracción de la muestra del módulo, también se tiene una serie de imágenes en formato DICOM.

Este tipo de archivos médicos generan ciertas ventajas sobre formatos JPG o PNG, ya que contienen meta data que permite evaluar a profundidad cada detalle y como ventaja dar un informe a detalle y preciso, para generar un diagnóstico correcto. Un archivo DICOM puede contener una imagen y esta puede tener múltiples “fotogramas” (frames), que contienen diferente toma en diferentes ángulos, permitiendo hacer un análisis a profundidad, teniendo en cuenta varios aspectos según la perspectiva. Incluso en la figura 3 pudimos observar las zonas en las que es más común hacer las tomas de imágenes que se mencionó anteriormente.

### 5.3. Recopilación de datos.

Después del procedimiento que se le realiza al paciente, el siguiente paso es la extracción de los datos para realizar su respectivo análisis médico. Es en este punto, donde se lleva a cabo, la entrega de las imágenes por parte de los médicos, que han sido colaboradores del proyecto.

Desde el 2020, se han ido recopilando archivos de imágenes DICOM, extraídas del ecógrafo, donde se realizan procedimientos indoloros, como es la toma de fotogramas mediante el transductor o exámenes de biopsias para diagnosticar nódulos tiroideos con evidencia de malignidad.

Estas imágenes tienen una particularidad, debido a que contienen información sensible del paciente y algunas características sobre puestas, que permiten detallar el objeto a identificar. Para facilitar el trabajo de etiquetado a los especialistas, se desarrolló una página web bastante simple hecha con el framework Django de Python montando en el servidor de Heroku, que de forma gratuita nos permite desplegar una aplicación web, a la que se puede acceder a través del siguiente enlace: <https://freython-projects.herokuapp.com/>

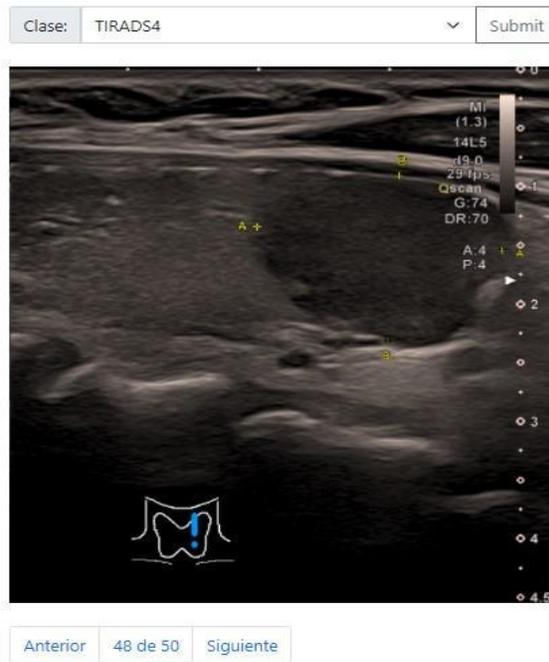


Ilustración 4. Página web para el etiquetado de las imágenes.

Tiene un diseño simple y amigable el cual consta de un menú de opciones en el que se encuentran las 5 etiquetas con las que se realizará la clasificación y los botones para guardar y siguiente. Cada semana se subían un mínimo de 50 imágenes, para finalmente reunir 1353 imágenes etiquetadas.

### 5.3.1. Pre procesamiento de los datos.

1. Se obtienen las imágenes originales con un tamaño de 996 de alto por 726 píxeles de ancho, clasificadas manualmente por los radiólogos, archivadas en carpetas con la clase correspondiente.
2. Se cambia los formatos de las imágenes de .DICOM a JPG con el fin de trabajar con un formato estándar y de fácil implementación en el modelo de red neuronal.
3. Las imágenes obtenidas como .JPG se recortan para eliminar distracciones al momento de entrenar y eliminar datos personales de los pacientes. Este procedimiento se hace de forma manual debido a que cada imagen varía en tamaño, forma de la toma y posición.
4. Se hace exploración de la data con el fin de identificar la cantidad de datos que se disponen para cada clase, la dimensión y el tipo de imagen, en este caso, se trabaja con 5 clases de tipo RGB. Se redujimos el tamaño a cada imagen para estandarizarlas, con un alto de 490 píxeles y un ancho de 411 píxeles.
5. Verificamos la calidad de los datos debido al cambio de formato.

Con las imágenes listas, procedemos a verificar que las clases estén desbalanceadas, e identificamos que dado que el conjunto de datos que contiene más imágenes es la clase TIRADS 4, por ello es importante equilibrar cada clase antes de continuar con el

proceso de entrenamiento. Los pesos de las imágenes para todas las clases de TIRADS, se ajustaron para que TIRADS 1 y 2 tenga un mayor impacto en la red neuronal haciendo un balance a las clases con casos benignos, esto nos ayuda a evitar posibles sesgos u overfitting.

Por lo tanto se tienen las siguientes cantidades de datos por clase, distribuidos por Train, Validation y Test.

Train:

TIRADS 1	TIRADS 2	TIRADS 3	TIRADS 4	TIRADS 5
122	242	160	309	138

Tabla 7. Cantidad de imágenes por clase, conjunto de entrenamiento.

Validation:

TIRADS 1	TIRADS 2	TIRADS 3	TIRADS 4	TIRADS 5
31	60	40	78	34

Tabla 8. Cantidad de imágenes por clase, conjunto de validación.

Test:

TIRADS 1	TIRADS 2	TIRADS 3	TIRADS 4	TIRADS 5
18	35	23	43	20

Tabla 9. Cantidad de imágenes por clase, conjunto de prueba.

Teniendo en cuenta que nuestros datos constan de 1353 imágenes en total, lo que significa que no es lo suficientemente grande como para evitar el overfitting al ajustar una CNN y aún más cuando se trabaja con una clasificación de 5 clases, que para estos casos es importante tener un número significativo por clase, lo que no se logró debido que es proceso solo lo puede realizar un experto y todo esto toma mucho tiempo. Por lo tanto, las muestras se aumentaron para mejorar la capacidad de predicción de nuestro modelo de Deep Learning. En el proceso de aumentar los datos, se tuvo en cuenta cada detalle de las imágenes porque los nódulos en los que se evidencia una serie de puntos blancos pueden significar calcificación y finalmente un factor importante a tener en cuenta en nódulos cancerígenos.

Se aplica el método ImageDataGenerator que se encarga de hacerle transformaciones y aumentar las imágenes, aplicando una rotación a cada una, seguido a esto, desplazamos la imágenes hacia la izquierda o hacia la derecha (desplazamientos horizontales) con el método "width shift range", en cuanto a "height shift range" funciona igual que "width shift range" pero cambia verticalmente (hacia arriba o hacia abajo) haciendo que nuestro objetivo se traslade según lo indicado, las imágenes también pasaron por un proceso de distorsión a lo largo de un eje, principalmente para crear o rectificar los ángulos de percepción. Por lo general, se usa para aumentar las imágenes para que las computadoras puedan ver como los humanos ven las cosas desde diferentes ángulos esto ayudará a que el modelo no se aprenda las respuestas de

memoria. También se realizaron cortes horizontales y con “brightness range” jugamos con el brillo de la imagen y finalmente aplicamos “fill mode” que identifica los puntos más cercanos y que son los de mayor importancia sobre nuestro conjunto de datos.

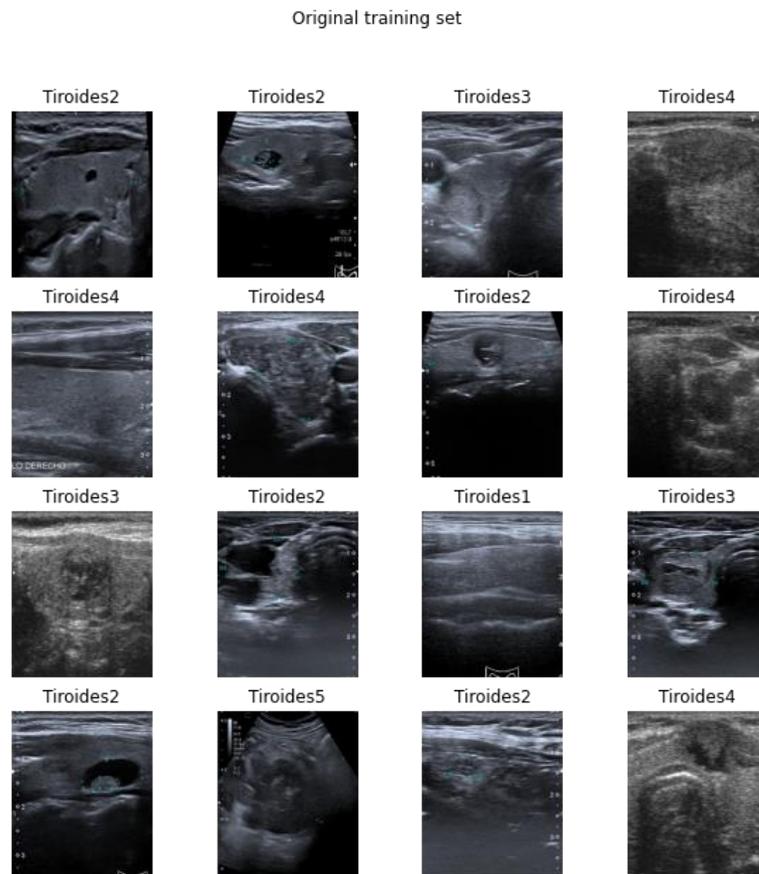


Ilustración 5. Imágenes de Tiroides con su etiqueta.

#### 5.4. Modelado.

Teniendo un pre procesado de la data y su balanceo, se da inicio a la construcción del modelo del entrenamiento.

Se dispone de un tensor de tamaño (490, 411, 3), es decir 490 de alto, 411 de ancho de las imágenes y 3 que indica el canal de colores, en este caso tenemos imágenes RGB.

#### **5.4.1. Implementación del modelo CNN.**

Procedemos a construir la arquitectura de nuestro modelo de red neuronal para luego dar inicio del proceso de entrenamiento, por consiguiente, tomamos como base del entrenamiento el poderoso modelo **CNN** (red neuronal convolucional) que consiste en un modelo basado en el funcionamiento de las neuronas específicamente en la corteza visual del cerebro. Técnicamente tiene como base el perceptrón multicapa, según su estructura trabaja sobre matrices bidimensionales, siendo muy efectivas para tareas de visión por computador, tanto para la clasificación y segmentación entre otras tareas.

A continuación se hace una descripción grafica de la arquitectura construida para este proceso de clasificación de 5 clases discriminatorias:

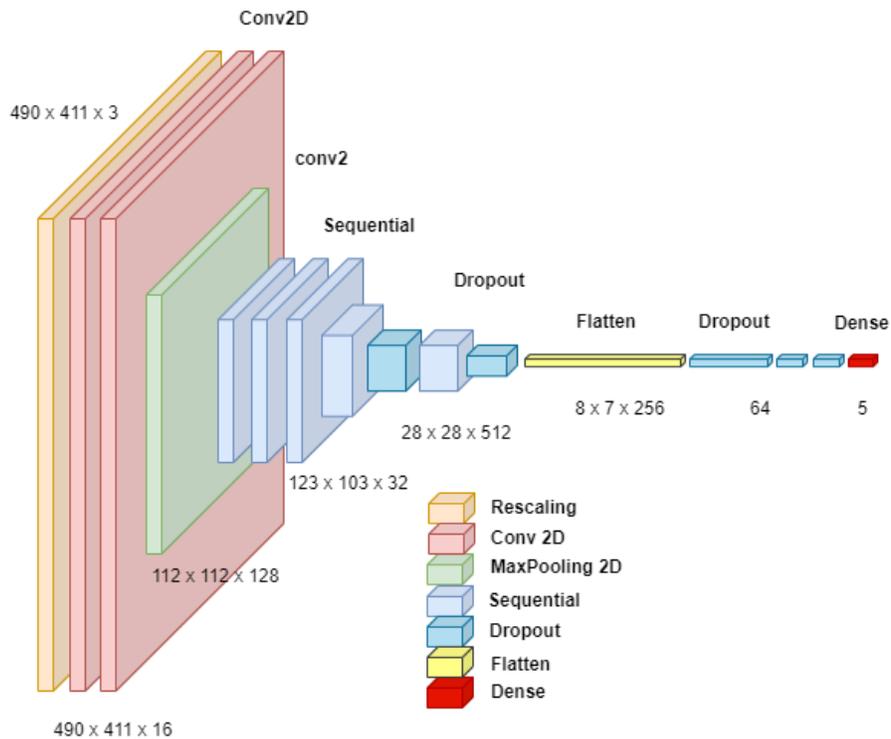


Ilustración 6. Arquitectura del modelo CNN construida para la clasificación de 5 clases.

Con el fin especificar el modelo de Red Neuronal Convolutacional con el que se iniciará el proceso de entrenamiento, se tiene la siguiente arquitectura: El tensor de entrada del modelo tiene la forma 490, 411, 3, esto significa que tenemos como entrada imágenes de tamaño 490 de alto, 411 de ancho y un canal RGB (3). Se declara la función Conv\_block que tiene como entrada los filtros que se refiere a la cantidad de filtros que se le aplican al conjunto de datos en formato entero, este método está compuesto por 2 capas convolucionales, un Batch Normalization y un Maxpoolig con la función de activación ReLu.

Seguido a crear las capas de convoluciones con sus filtros, construimos un método dense block, que viene siendo el bloque de capa dense que recibe las unidades de filtros y la tasa de Dropout o salida, esta bloque se compone de una capa dense con una función de activación ReLu, una capa de Batch Normalization y el Dropout al que le pasamos la tasa de Dropout en porcentaje.

Finalmente, construimos el modelo con todo el conjunto de bloques que acabamos de desarrollar, este método se encarga de ensamblar todo. Para este proyecto de clasificación en 5 clases se construyó con las siguientes especificaciones:

- Capa de Rescaling, escalando las imágenes entre 0 y 1.
- Dos capas de convolución con 16 filtros.
- Una capa de MaxPooling.
- 8 bloques de convolución con filtros 32, 64, 64, 256, 512, 128, 64 respectivamente (recordemos que estos bloques de convolución están conformados por 2 capas de convolución, un Batch Normalization y un MaxPooling con dos filtros).
- 2 capas Dropout que nos permiten reducir el Overfitting.
- Una capa de Flatten.
- Capa Dense con 5 unidades de salida y una función de activación Sigmoid ya que tenemos una clasificación “categorical” o “multi class”.

En cuanto a las estrategias de alcance o Scope, le pasamos el modelo construido, las métricas de evaluación que en este caso son 3:

- Accuracy.
- Precision.
- Recall.

La imagen anterior es la arquitectura creada para esta tarea en específico que consta de una clasificación de 5 clases teniendo 1353 imágenes en total, que para cada clase, por lo tanto el dataset se divide en 3 carpetas, una para el entrenamiento, una para validación y una para prueba, se tienen:

Compilamos nuestro modelo aplicando el optimizador Adam, el método de pérdida Categorical Cross Entropy y le pasamos las métricas de evaluación.

Tenemos 60 pasos por épocas y 15 pasos para validación, también es importante guardar nuestros checkpoints en un archivo formato h5.

Para tener una idea más clara sobre los pasos por épocas y como saber cuántos son los pasos correctos dependiendo de la cantidad de datos. La siguiente formula nos muestra la manera correcta para indicar los pasos exactos a nuestro modelo:

$$Steps = \frac{train\ image\ count}{batch\ size}$$

Pasos para el proceso de validación:

$$Steps = \frac{validation\ image\ count}{batch\ size}$$

Cabe recordar que no siempre se obtiene los pasos con esta fórmula, pero es una forma de experimentar y fue una de las maneras en las que se hizo pruebas para finalmente escoger 60 pasos por época.

Es importante definir el método “Early Stopping” que nos da la ventaja de parar el entrenamiento si es que no presenta mejoría, es decir que los resultados después de 10 pasos no mejora, se para la ejecución. Así evitamos el consumo de recursos innecesarios y si no está aprendiendo el modelo lo mejor es parar el entrenamiento.

Finalmente, procedemos a entrenar nuestro modeló con 100 épocas de entrenamiento, proceso en el cual se obtiene como resultado: Parámetros a tener en cuenta:

- AUTOTUNE = tf.data.experimental.AUTOTUNE
- BATCH SIZE = 16 \* strategy.num replicas in sync
- EPOCHS = 100.

Resultados obtenidos:

Época	Accuracy	Loss	Precision	Recall
9	0.3381	1.5773	0.2423	0.6259

Tabla 10. Resultados del primer entrenamiento con CNN.

Generamos las gráficas de los resultados obtenidos según la Precision, Recall, Accuracy y Loss:

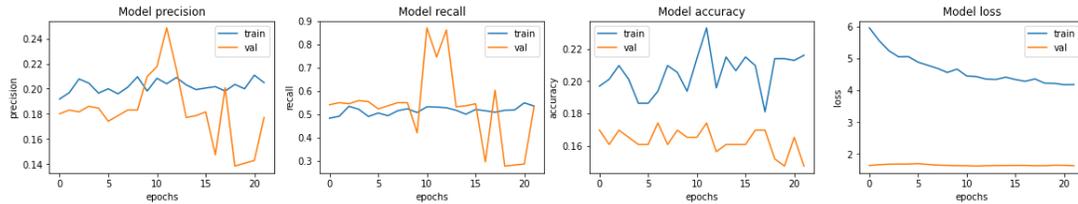


Ilustración 7. Gráficas de los resultados de evaluación con respecto a las épocas.

Luego generamos el grafico de la matriz de confusión con la finalidad de visualizar el desempeño de nuestro algoritmo supervisado y observar los casos fallidos, es decir las equivocaciones que tuvo el modelo a la hora de predecir. Continuación se adjunta la tabla de resultados después de entrenar nuestro modelo con una CNN.

Reporte de la clasificación según todas las variables de evaluación:

Clases	Precisión	Recall	F1-Score	Support
0	0	0	0	18
1	0.43	0.17	0.24	35
2	0.3	0.48	0.37	23
3	0.38	0.67	0.48	43
4	0.14	0.05	0.07	20
Accuracy	0	0	0.34	139
Macro AVG	0.25	0.27	0.23	139
Weighted AVG	0.29	0.34	0.28	139

Tabla 11. Reporte de clasificación.

Como podemos ver en el anterior fragmento donde se plasman los resultados obtenidos del entrenamiento con CNN, no se logran obtener buenos resultados, ya que al evaluarlo con 139 imágenes, presenta 92 desaciertos, un poco más de la mitad de los datos que se le pasan al modelo para testear, con esto podemos deducir que la cantidad de imágenes que se tienen no son suficientes para una clasificación multi-clase, por ello fue necesario aplicar el método Transfer Learning, que consiste en:

Transferir conocimiento, literalmente, adquirido en otro entrenamiento con otro conjunto de datos, resolviendo un problema en específico y que funciona para resolver otros tipos de problemas. Es una de las mejores técnicas para mejorar procesos de aprendizaje en los que no se han obtenido buenos resultados, con la finalidad de mejorarlos.

Es importante tener buenos recursos de máquina para este tipo de entrenamientos como buena capacidad de RAM y una GPU, teniendo en cuenta esto, fue necesario adquirir una cuenta pro de **Google Colab**, para mejorar la calidad del rendimiento y reducir el tiempo. Algunas de las ventajas son:

- GPU más rápidas
- Más memoria
- Tiempos de ejecución más largos

Como primeras experimentaciones aplicamos Fine-Tuning al modelo, ajustando los hiperparámetros, se elige el modelo con la finalidad de realizar pruebas según su

eficiencia. Se realizan varios experimentos y se reentrena el modelo, con un modelo de red neuronal convolucional. El objetivo de estos experimentos es mejorar los porcentajes de predicción teniendo en cuenta que se tiene poca data, sin alterarla y evitando el overfitting. Medimos su rendimiento con el conjunto de prueba ("test") para estimar el error y perdida. Guardamos el modelo entrenado como model.h5, esto nos permite desplegarlo en la página web que se desarrolló para mostrarlo al usuario de forma funcional. La plataforma está hecha con Angular y tiene en funcionamiento el modelo. En el sitio se permite subir un archivo (imagen), se aplica la función model predict, esto nos arroja la clase a la que pertenece dicha imagen.

Después de guardar el modelo comenzamos aplicar Transfer Learning, como primer modelo para entrenar junto con datos diferentes a los que se tenían, se escoge iniciar con:

#### **5.4.2. Implementación del modelo VGG.**

VGG16 es una arquitectura basada en el modelo de red neuronal convolucional y altamente utilizada para ImageNet. Su base fundamental es el reconocimiento de imágenes a gran escala, con 16 capas como su nombre nos lo indica, ha logrado el 92,7% con ImageNet que viene siendo un conjunto de datos con más de 14 millones de imágenes que pertenecen a 1000 clases. VGG16 es muy utilizado para clasificación de imágenes de Deep Learning y un modelo de fácil implementación. [6] Su arquitectura es la siguiente:

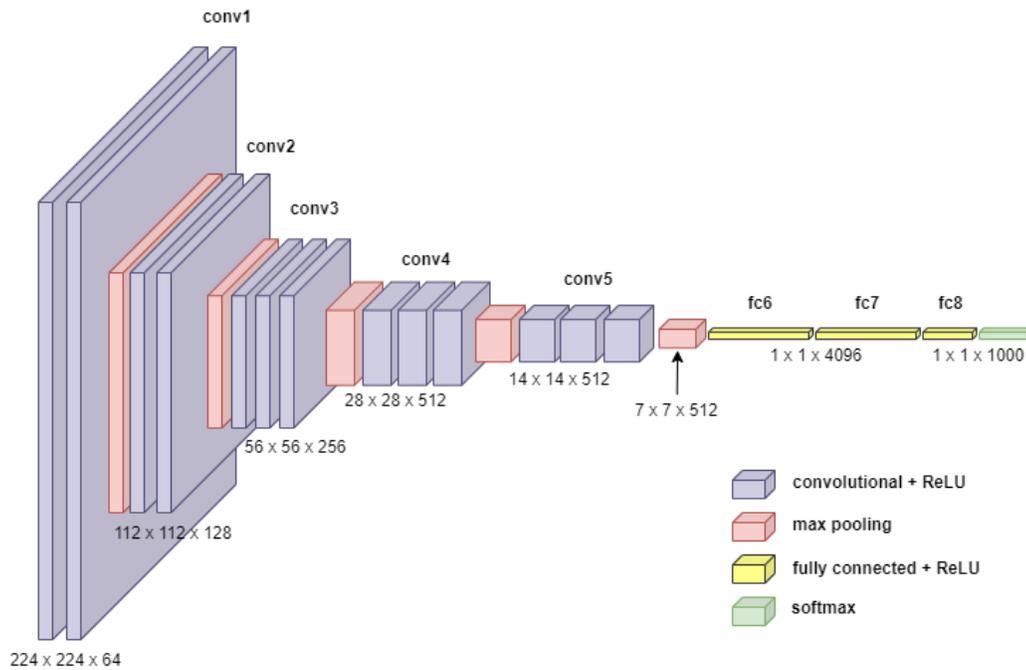


Ilustración 8. Arquitectura VGG16.

Teniendo en cuenta que VGG16 es una arquitectura especial para clasificación en imágenes aplicamos esta arquitectura nuestra primera prueba para Transfer Learning.

Descargamos el modelo VGG16 con los pesos del “ImageNet”, para este experimento final no congelamos la base del modelo, pero si declaramos los parámetros y variables necesarias para construir la arquitectura sin olvidar que la salida de nuestro modelo es de dimensión 5 ya que tenemos 5 clases.

Se obtienen los siguientes resultados después de entrenar por 8 épocas ya que se congeló el entrenamiento debido a que no mejoraba.

Época	Accuracy	Loss	val_loss	val_accuracy
8	0.4611	3.5652	1.5998	0.2857

Tabla 12. Resultados del entrenamiento con el modelo VGGT, sin congelar el modelo.

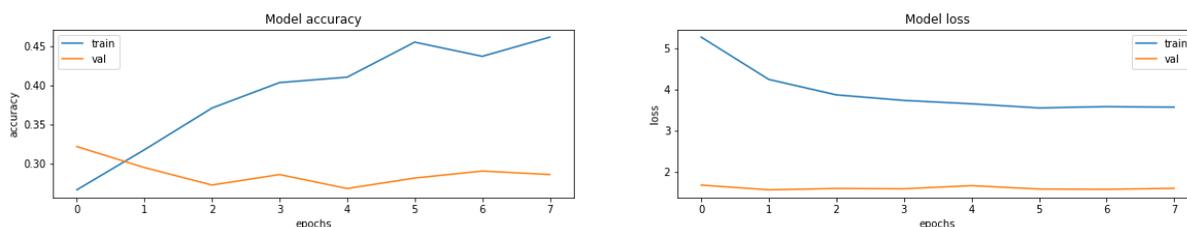


Ilustración 9. Evaluación del entrenamiento con VGG16, Accuracy vs épocas y Loss vs épocas (sin congelar el modelo).

Para el siguiente paso y experimento, congelamos la base del modelo con las mismas especificaciones anteriores y se obtuvieron los siguientes resultados:

Época	Accuracy	Loss	val_loss	val_accuracy
13	0.5580	2.8828	1.4766	0.4196

Tabla 13. Resultados del entrenamiento con el modelo VGGT, sin congelar el modelo.

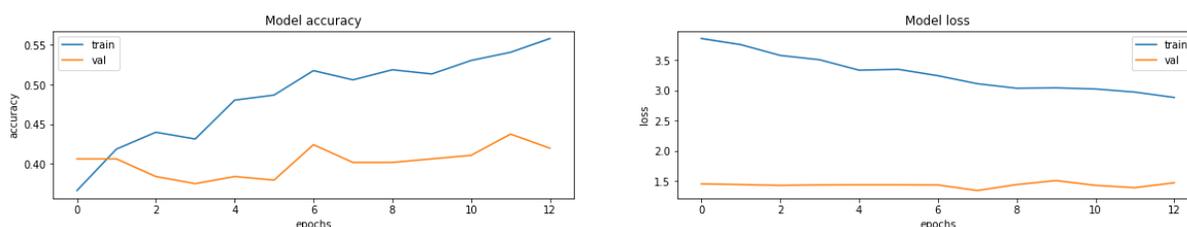


Ilustración 10. Evaluación del entrenamiento con VGG16, Accuracy vs épocas y Loss vs épocas (Congelando el modelo).

### **5.4.3. Implementación del modelo Xception.**

Otra de las técnicas que aplicamos como Transfer Learning con el objetivo de mejorar los resultados, teniendo en cuenta que nuestro dataset no es lo suficientemente robusto para una clasificación multi class, por ende se decide hacer pruebas con Xception, según su artículo de investigación publicado en Arxiv, Cornell University, una convolución separable en profundidad puede entenderse como un módulo de inicio con número máximo de torres o escalones, esto lleva a los investigadores a proponer una nueva arquitectura de red neuronal convolucional profunda inspirada en Inception, donde la estructura de Inception se ha reemplazado por convoluciones separables en profundidad.

Esta arquitectura supera ligeramente a Inception V3 entrenado con el conjunto de datos ImageNet, ya que Xception tiene la ventaja que hace un uso eficiente de los parámetros del modelo puesto que en estructura tiene la misma cantidad de parámetros que Inception V3.

Su arquitectura en cuanto al proceso de convolución profunda en Xception es la siguiente:

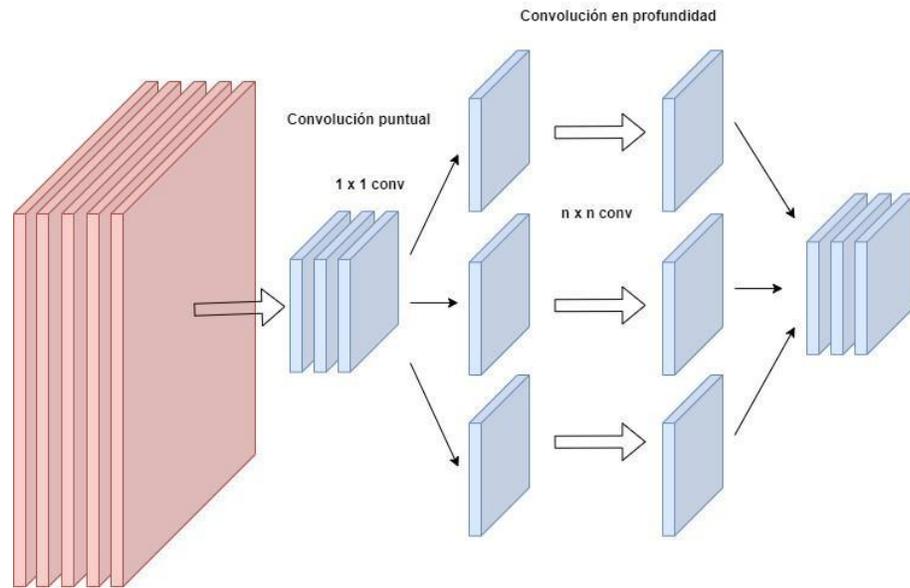


Ilustración 11. Arquitectura de Xception.

Descargamos el dataset “imagenet” y cargamos el modelo Xception, y pasamos a entrenar pero antes es necesario hacer ciertos ajustes:

- Llamamos al método de Keras, Clear session para que limpie la sesión del entrenamiento anterior con el modelo CNN.
- Congelamos la base del modelo para controlar la forma en que se actualizan los pesos y esto tiende a acelerar el entrenamiento y así trabajar con los pesos que ya se tienen.
- Aplicamos Early Stopping por si el entrenamiento no mejorar después de 6 iteraciones.
- Redujimos la tasa de aprendizaje para aplicar Fine-Tuning.
- Guardamos los checkpoints del entrenamiento.

Como resultados del entrenamiento se obtuvo:

Época	Accuracy	Loss	val_loss	val_accuracy
26	0.4995	3.1475	1.3032	0.4958

Tabla 14. Resultados obtenidos del entrenamiento del modelo Xception (sin congelar la base del modelo).

Graficas del Accuracy y Loss contra las épocas de entrenamiento:

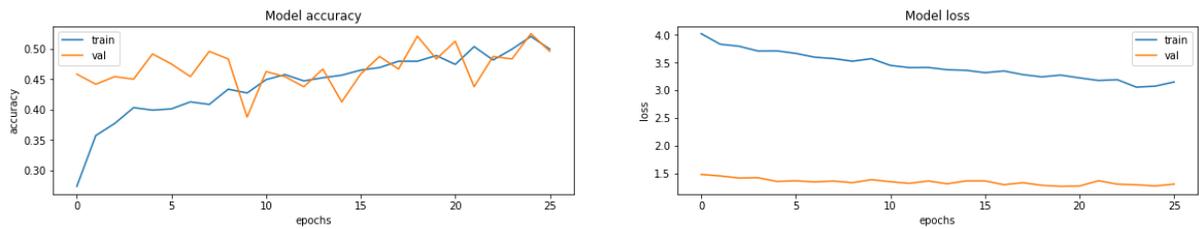


Ilustración 12. Gráfica del Accuracy y Loss con respecto a las épocas.

Luego aplicamos Fine-Tuning a nuestro modelo Xception, congelamos de nuevo la base del modelo y guardamos los checkpoints, entrenamos, y esta vez se obtuvo los siguientes resultados:

Época	Accuracy	Loss
9	0.5540	1.1644

Tabla 15. Resultados obtenidos del entrenamiento del modelo Xception (congelando la base del modelo).

Aquí podemos observar como logramos mejorar el Accuracy y que el Loss sigue sin disminuir, quiere decir que algo anda mal en nuestro modelo o datos.

Los resultados obtenidos para Accuracy y Loss con respecto a las épocas de entrenamiento, son:

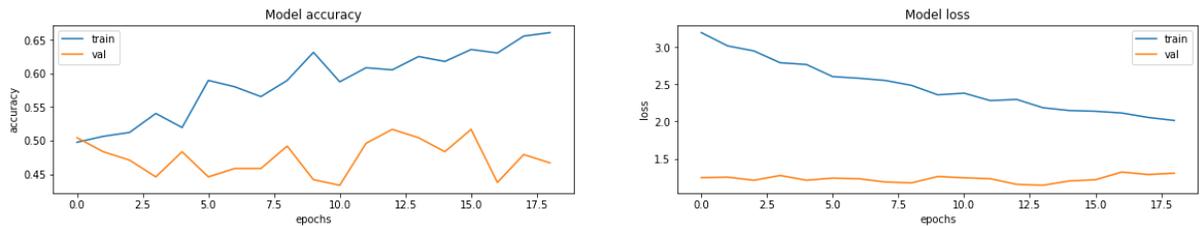


Ilustración 13. Gráfica de Xception, Accuracy y Loss con respecto a las épocas.

Continuando la serie de mejoras al proyecto, aplicamos otro modelo para Transfer Learning, como bien es conocido el modelo ResNet152 V2.

#### 5.4.4. Implementación del modelo ResNet.

Este modelo consiste en Deep Learning para el reconocimiento de imágenes, está inspirado en VGG-19 en la que le agregan la conexión de acceso directo, esto se refiere a que estos bloques o conexiones luego convierten la arquitectura en la red residual, como es tan potente y ha demostrado dar muy buenos resultados en cuanto a clasificaciones de imágenes, se aplicó a este proyecto con el fin de mejorar los resultados obtenidos con respecto a los modelos anteriores.

Cargamos el modelo ResNet512V2, sus pesos pre entrenados con el data set ImageNet y esta vez no congelamos la base del modelo. De nuevo aplicamos las métricas de

evaluación y el early stopping, configuramos nuestro learning rate, finalmente procedemos a entrenar, los resultados obtenidos se muestran a continuación:

Época	Accuracy	Loss	val_loss	val_acc
23	0.5141	1.3135	1.3135	0.4375

Tabla 16. Resultados obtenidos del entrenamiento del modelo ResNet (sin congelar la base del modelo).

Grafica de los resultados obtenidos:

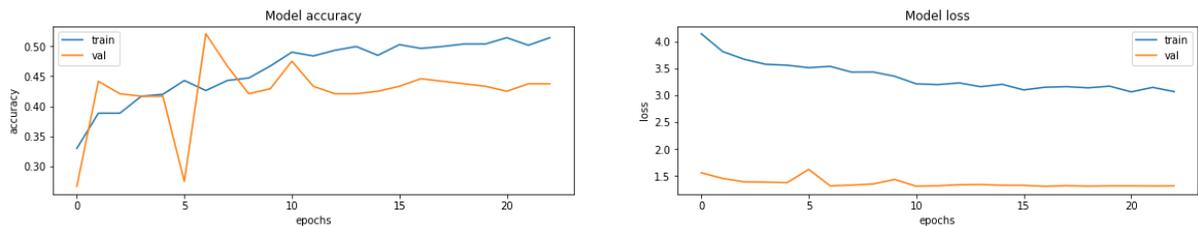


Ilustración 14. Gráfica del Accuracy y Loss con respecto a las épocas, sin congelar la base del modelo.

Podemos observar como el loss creció más de un 30% del experimento anterior, lo que significa que no son buenos resultados, es necesario volver a realizar otro experimento.

Para este último proceso entrenamos el modelo con las mismas características que usamos anteriormente y aplicamos el congelar el modelo.

Esta vez con tan solo 13 épocas logramos bajar el Loss y aumentar el Accuracy a un 55%.

Época	Accuracy	Loss
13	0.55404	1.1644

Tabla 17. Resultados obtenidos del entrenamiento del modelo ResNet (congelando la base del modelo).

Grafica de rendimiento del modelo según la época, evaluando el Accuracy y el Loss:

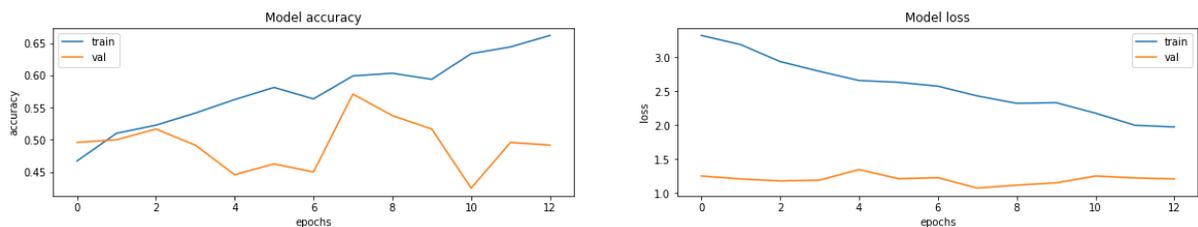


Ilustración 15. Gráfica del Accuracy y Loss con respecto a las épocas, congelando la base del modelo.

Como conclusión del proceso realizado durante todo el desarrollo del modelo es que, a base de tener pocas imágenes ha sido un trabajo bastante complejo lograr mejorar las métricas de evaluación y la precisión de cada modelo aplicado teniendo en cuenta que es una clasificación de 5 clases lo que hace que sea decisivo e importante tener una gran cantidad de datos etiquetados, se estima que alrededor de unos 5000 datos sería una buena cantidad para entrenar y mejorar los resultados, así realizar buenas predicciones.

## 5.5. Evaluación

Como resultado de cada modelo que se implementó en el desarrollo del proyecto, se elabora una tabla con la arquitectura y los resultados obtenidos. Todas estas pruebas con los diferentes modelos, fue necesario buscar alternativas debido a que se contó con un dataset muy pequeño para la clasificación de 5 clases.

Con la implementación de los modelos anteriores se puede concluir que tuvimos una mejora significativa en los resultados del Accuracy aplicando las diferentes arquitecturas debido el modelo CNN no tuvo el mejor rendimiento después de varios experimentos y ajuste que se le aplicaron. Se evidencia mejoras en cada entrenamiento que se realizó con los diferentes modelos planteados, es decir pasamos de una pérdida o loss de 4.1 a un loss de 1.97 que se obtuvo con la arquitectura ResNet. Finalmente, el modelo ResNet pre entrenado con ImageNet logra alcanzar un accuracy de 66%. Para lograr estos resultados se practicaron varios experimentos que se detallan en la siguiente sección.

La siguiente tabla representa un resumen claro de cada modelo, su arquitectura y los resultados obtenidos.

Modelo	Arquitectura	Accuracy	Loss
CNN	18 capas de convolución, 10 capas de batch normalization, 8 capas de Max Pooling, 6 capas de Dropout, 3 capas Dense, una capa Flatten, una última capa Dense con un tamaño de salida 5, 30	0.1655	1.6285

	pasos por época, 7 pasos de validación, 100 épocas para entrenamiento y optimizador Adam.		
VGG 16 + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo NO congelada	No congelamos la base del modelo, cargamos el modelo, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.46	3.56
VGG 16 + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo congelada	Congelamos la base del modelo, cargamos el modelo pre-entrenado VGG16, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.4460	1.4307
Xception + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo NO congelada	No congelamos la base del modelo, cargamos el modelo pre-entrenado Xception con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.4590	3.5212
Xception + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo congelada	Congelamos la base del modelo, cargamos el modelo pre-entrenado Xception con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.5180	1.161
ResNet + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo NO congelada	No congelamos la base del modelo, cargamos el modelo pre-entrenado ResNet con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.05141	3.0673

ResNet + Fine Tuning (Transfer Learning - ImageNet) - Base del modelo congelada	Congelamos la base del modelo, cargamos el modelo pre-entrenado ResNet con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.5540	1.1644
---	--	--------	--------

Tabla 18. Arquitectura de cada modelo con su evaluación.

## 5.6. Despliegue.

En esta sección del documento, se describe el proceso de desarrollo del software para el prototipo de clasificación y su despliegue. A continuación, los pasos para la construcción de la página web como herramienta para el área de radiología en apoyo al análisis del diagnóstico médico en anomalías de la glándula tiroidea.

### 5.6.1. Identificación de requisitos.

Como se planteó al inicio del documento, se desarrolló un software como apoyo a especialistas en radiología, que genera la clasificación de imágenes de ultrasonido. Este sistema permitirá apoyar el diagnóstico oportuno que elabora el doctor encargado, en el menor tiempo posible, siendo esto de gran beneficio para los pacientes ya que se podrá actuar de forma oportuna.

Este software tiene una interfaz de usuario web, lo que dará una mejor experiencia para realizar evaluaciones hombre - máquina. Este proceso se hará cada que se agreguen

más imágenes al dataset. Con ello mejorar el modelo hasta el punto de implementarlo en centros de salud.

### 5.6.2. Requerimientos funcionales.

<b>Requerimiento #RF01</b>	<b>Cargar una imagen desde el ordenador.</b>
Descripción	El software debe dar la opción de cargar un archivo en formato JPG.
Responsable	Usuario del software (Radiólogos).
Entradas	Imagen en formato JPG.

Tabla 19. Requerimientos funcionales.

<b>Requerimiento #RF02</b>	<b>Computar la clase correspondiente a la imagen según la escala EU-TIRADS.</b>
Descripción	El software generará un resultado basado en las características de la imagen cargada al sitio web, esta imagen pasará por el modelo ya entrenado para predecir su clase.
Responsable	Modelo de Deep Learning.
Entrada	Mapa de características.
Resultado	Obtener un resultado de la clasificación de la imagen según la escala EU-TIRADS.

Tabla 20. Requerimientos funcionales

<b>Requerimiento #RFN01</b>	<b>Plataforma web responsiva</b>
Descripción	La plataforma web permitirá la visualización de esta en cualquier dispositivo.
Responsable	Usuario del software (Radiólogos).
Resultado	Visibilidad en cualquier dispositivo

Tabla 21. Requerimientos no funcionales.

### 5.6.3. Requerimientos no funcionales.

<b>Requerimiento #RFN02.</b>	<b>Plataforma web amigable.</b>
Descripción	Plataforma web de fácil acceso con interfaz principal a su funcionalidad específica.

Tabla 22. Requerimientos no funcionales.

### 5.6.4 Herramientas.

Se hizo uso de las siguientes herramientas para el desarrollo del software y el modelo de Deep Learning.

- Tensorflow 2.
- Tensorflow JS.
- Python 3.8.
- React JS.
- GitHub Page.
- Despliegue de modelos.
- Servidor para alojar la aplicación web (Heroku).

## 6. Experimentos.

Experimento	Modelo	Arquitectura	Accuracy	Loss
EXP #1	CNN	Rescaling, 3 capas de convolución, 3 capas de Max Pooling, una capa de Flatten, 2 capas Dense, 100 épocas y un optimizador Adam.	1.0	1.38
EXP #2	CNN	12 capas de convolución, 7 capas de Batch Normalization, 4 caps de Max Pooling, 5 capas de Dropout, 1 capa Flatten, 4 capas Dense y una optimizador Adam.	0.28	1.57
EXP #3	CNN	18 capas de convolución, 10 capas de batch normalization, 8 capas de Max Pooling, 6 capas de Dropout, 3 capas Dense, una capa Flatten, una última capa Dense con un tamaño de salida 5, 30 pasos por época, 7 pasos de validación, 100 épocas para entrenamiento y optimizador Adam.	0.16	1.62
EXP #4	VGG16	Congelamos la base del modelo, cargamos el modelo pre-entrenado VGG16, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.44	1.43
EXP #5	Xception	Congelamos la base del modelo, cargamos el modelo pre-entrenado Xception con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores, optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5	0.51	1.16
EXP #6	ResNet	Congelamos la base del modelo, cargamos el modelo pre-entrenado ResNet con ImageNet, base_model se ejecutando en modo de inferencia, pasando training=False, convertimos las características del shape base_model.output_shape[1:] en vectores,	0.55	1.16

		optimizador Adam, 100 épocas y una capa Dense con una salida de tamaño 5		
--	--	--	--	--

Tabla 23. Experimentos, arquitectura y evaluación.

Nos damos cuenta que el Loss no varía mucho, tenemos resultados desde 1.16 con los modelos Xception y ResNet a 1.62 con CNN, es decir que tenemos una pérdida bastante alta, más del 100% siendo un tema crucial para tratar, en el que se propone aumentar los datos pero sin ningún método, simplemente agregando más imágenes originales al dataset. Posiblemente nos permitirá mejorar poco a poco los resultados del Accuracy y disminuir el Loss.

Es evidente que el mejor resultado de Accuracy obtenido durante los entrenamientos con los diferentes modelos propuestos, basados en redes neuronales convolucionales, y Transfer Learning, ha sido el modelo ResNet, pre-entrenado con ImageNet quién ha generado el más alto Accuracy con respecto a los demás, con un 55%.

## 7. Presupuesto.

La siguiente información presenta el presupuesto a detalle que ha sido necesario para el desarrollo del proyecto. En las siguientes tablas se representa el presupuesto relativo a los recursos materiales, a los recursos humanos y a otros recursos.

Recursos materiales	Unidades	USD/ud	Total
Ordenador HP Pavilion 16RAM, TG Radeon	1	1.059	1.059
Disco duro 1T	1	55	56
Cuenta Google Colab Pro	8	10	80

Cuenta de Heroku	12	25	300
Google Drive	12	10	120

Tabla 24. Recursos materiales.

Recurso humano	Unidades	USD/hora	Costo/mes
Ingeniero informático	1	12	2.880
Radiólogo	2	32	7.680

Tabla 25. Recurso humano.

Recursos (otros)	Unidades	USD/día	Costo/mes
Transporte al hospital	300	3	900

Tabla 26. Otros recursos.

Debido a la escasez de datos con los que se entrenó el modelo y a pesar de haber aumentado la data con la función Data Augmentation, no se registran mejoras esperadas, por el contrario, si se aumentan las épocas los resultados del Accuracy comienzan a variar, inicialmente da un porcentaje de 20% pero a medida que pasan las épocas de entrenamiento el porcentaje de Accuracy tiende a disminuir o aumentar un 50% y finalmente 60% y no aumenta más. No es viable aumentar las épocas porque los porcentajes de Accuracy no superan el 30% después de 50 épocas.

El problema que presenta el modelo desarrollado junto con el conjunto de datos con el que se entrenó es que la red comienza a sobre ajustarse y ambos fenómenos ocurren al mismo tiempo. La red está comenzando a aprender patrones solo relevantes para el conjunto de entrenamiento y no muy buenos para la generalización, lo que lleva al

fenómeno 2, algunas imágenes del conjunto de validación se predicen realmente mal, con un efecto amplificado por la “asimetría de pérdida”.

Sin embargo, al mismo tiempo sigue aprendiendo algunos patrones que son útiles para la generalización (fenómeno uno, “buen aprendizaje”) a medida que cada vez se clasifican correctamente más imágenes.

Otro posible problema es que las imágenes (ultrasonidos) no hayan sido clasificados correctamente debido a que este proceso es con aprendizaje supervisado, es decir que la clasificación depende de un experto.

Este etiquetado manual requiere de tiempo y al menos uno o dos especialistas en área, puesto que de ahí dependen los buenos resultados.

## **8. Conclusiones.**

Como primera conclusión se verifica la efectividad del uso de redes neuronales convolucionales, según la investigación realizada, para la clasificación de imágenes de cualquier tipo, incluyendo las imágenes médicas que por su alto nivel de detalle y complejidad, pueden representar un reto para otros tipos de modelos de aprendizaje computacional. El gran desafío a sortear cuando se trabaja la clasificación de imágenes médicas, está en obtener un dataset con muestras suficientes, para cada una de las categorías, validado por expertos en el tema de investigación a tratar y del cual se puedan extraer las características principales necesarias para el aprendizaje. Se necesita

una colaboración fuerte entre las disciplinas de la radiología e ingenieros informáticos, encargados del desarrollo del modelo, para construir bases de información relevantes que se puedan explotar a través del data mining y el machine learning.

Para trabajos futuros se recomienda, obtener una mayor cantidad de imágenes por clase, ya que en este proyecto, se trabajó con un total de 1353 imágenes y obtuvimos un 55% de Accuracy y una pérdida de 1.16, lo que significa que aún no puede ser implementado en producción ya que no arrojó buenos resultados. Sin embargo, del proyecto resaltan aspectos como: no fue necesario entrenar con más de 50 épocas, lo cual facilita la experimentación; a pesar de no conseguir un dataset robusto, un 55% de Accuracy para la clasificación de 5 clases, es bastante prometedor ya que tenemos una probabilidad del 27% de escoger la clase correcta con respecto a escogerla al azar.

Otro factor importante, es realizar pruebas hombre vs maquina con la finalidad de ajustar cuando sea necesario el modelo y llegar a obtener mejores resultados. Para este trabajo de investigación e implementación se tenía una cantidad mínima de imágenes que fue necesario aumentar para mejorar los resultados, pero no se logrará un avance en el tema hasta no obtener un dataset con las etiquetas correspondientes que se dan de forma manual, supervisadas y verificadas por el especialista en radiología. Debido a la complejidad del proyecto en cuanto al dataset para entrenar, se propone seguir contribuyendo para mejorar el modelo, con la finalidad de ofrecer una buena herramienta que de soporte en la toma de decisiones en diagnósticos enfocados a problemas que puede presentar la glándula tiroides.

## **9. Anexos.**

### **9.1. Página web para el despliegue del modelo.**

Se quiso desarrollar una interfaz web que permitiera el uso del modelo de clasificación de ultrasonidos de tiroides, para los usuarios o los médicos encargados de evaluar el rendimiento, con la finalidad de obtener una herramienta amigable.

Esta plataforma se montó con GitHub Page y se desplegó el modelo con Tensorflow JS. A continuación, se muestran las interfaces de la plataforma con sus funcionalidades:

Vista home con la descripción del proyecto y el menú de opciones, que se refiere a la información técnica y los colaboradores.

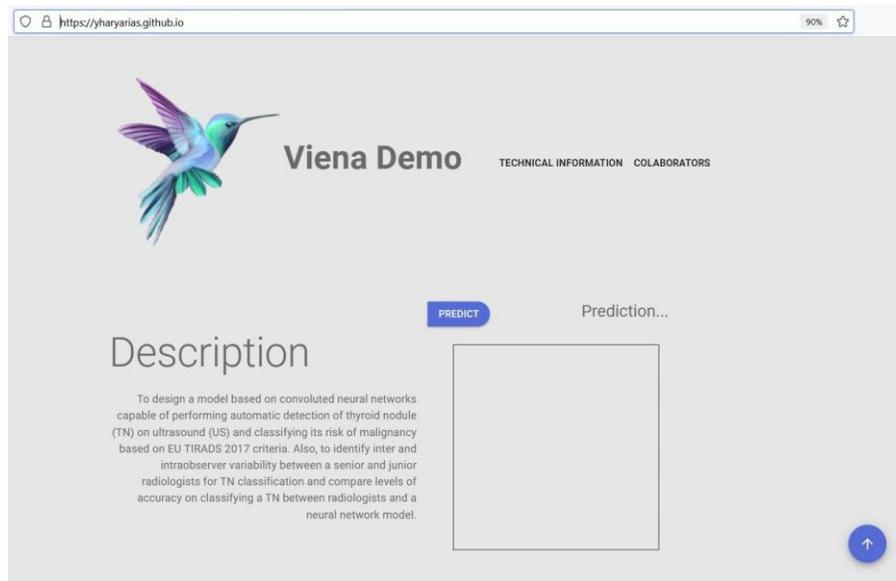


Ilustración 16. Interfaz de usuario (Despliegue del modelo).

Botón con la función de predicción “predica” donde se carga la imagen y se muestra gráficamente, finalmente arroja el resultado de la clasificación según la prueba de tamizaje EU-TIRADS.

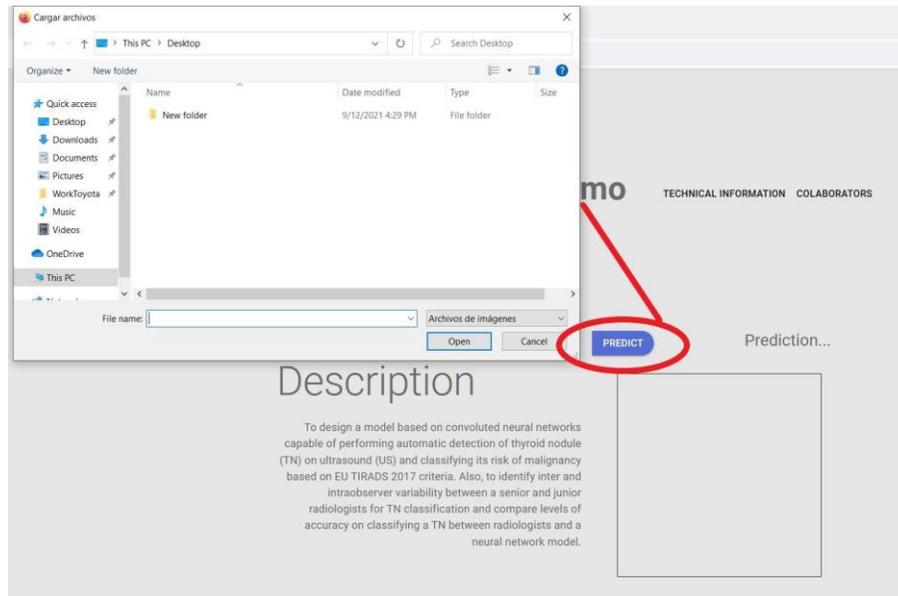


Ilustración 17. Cargue de imágenes.

1. Una vez cargada la imagen, el modelo evalúa el mapa de características es decir la imagen de ultrasonido y esta pasa por la función `model.predict()`, generando el resultado según la escala de TIRADS.
2. Interfaz con la información técnica usada para el desarrollo del proyecto



# Viena Demo

TECHNICAL INFORMATION COLLABORATORS

PREDICT TIRADS5

## Description

To design a model based on convoluted neural networks capable of performing automatic detection of thyroid nodule (TN) on ultrasound (US) and classifying its risk of malignancy based on EU TIRADS 2017 criteria. Also, to identify inter and intraobserver variability between a senior and junior radiologists for TN classification and compare levels of accuracy on classifying a TN between radiologists and a neural network model.

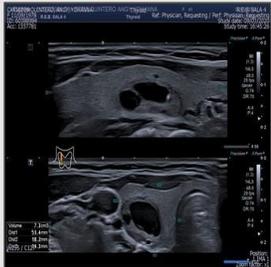


Ilustración 18. Predicción según su escala de clasificación

## TECHNICAL INFORMATION

The prediction model was built using a Convolutional Neural Network Architecture made in Python with Tensorflow and Keras. The model make a process to extract the most important characteristics from medical thyroid ultrasounds classifying them into few classes of TIRADS-EU.



TensorFlow



Keras



Python



Ilustración 19. Información técnica del proyecto.

## 9.2. Página web para el etiquetado de los datos.

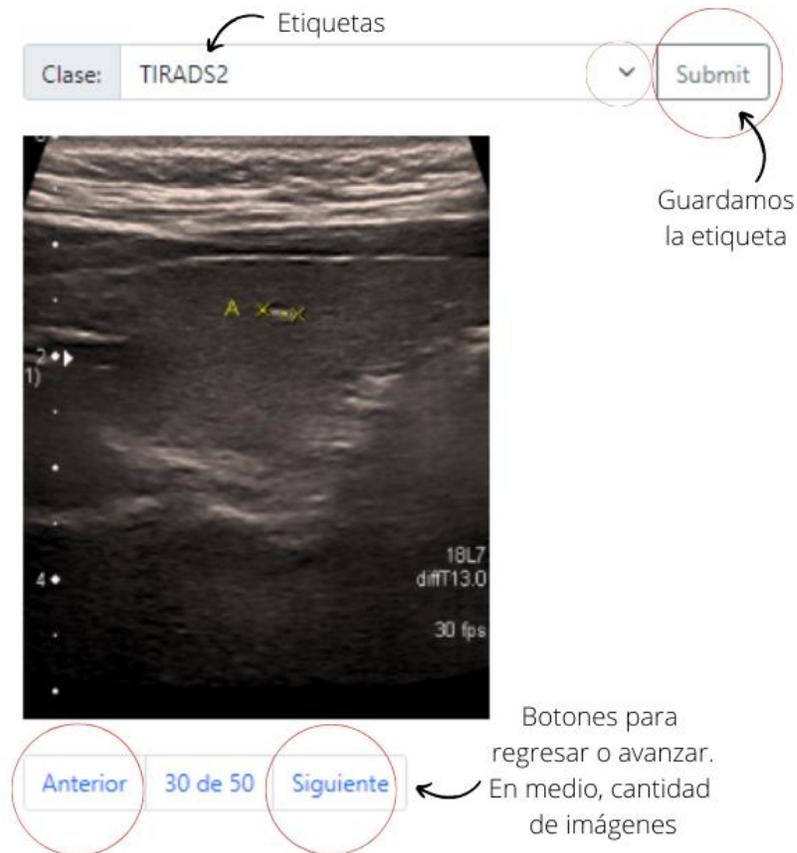


Ilustración 20. Página web para el etiquetado de las imágenes de tiroides.

La función principal de esta página web es facilitar el trabajo de etiquetado a los especialistas. Por ello se desarrolló una página web hecha con el framework Django de Python montando en el servidor de Heroku, que de forma gratuita, que permite desplegar una aplicación web, a la que se puede acceder a través del siguiente enlace: <https://freython-projects.herokuapp.com/>.

## 10. Referencias.

- [1] S.A.S., E., 2022. *Por cada 1.000 habitantes en Colombia, hay alrededor de 1,5 médicos generales.* [online] Diario La República.
- [2] Cancer.org. 2022. *Estadísticas importantes sobre el cáncer de tiroides.*
- [3] Mistry, R. (2020, 11 maart). *Ultrasound Classification of Thyroid Nodules: A Systematic Review.* Cureus. Geraadpleegd.
- [4] Ultrasound imaging classifications of thyroid nodules for malignancy risk stratification and clinical management: state of the art. (2019, 8 septiembre).
- [5] G. Swapna, U. R. A. (2014, 7 augustus). A Review on Ultrasound-Based Thyroid Cancer Tissue Characterization and Automated Classification. Journals.
- [6] V.S. Sharnitha, D. S. (2011). Thyroid classification and segmentation in ultrasound images using machine learning algorithms. IEEE.
- [7] J.F.S. (2014, 5 maart). TI-RADS classification of thyroid nodules based on a score modified according to ultrasound criteria for malignancy. Webcir.
- [8] Hershman, J. M. (2020, oktober). Introducción a la glándula tiroidea. Manual MSD versión para público general.
- [9] Department of Computer Science University of Waikato. (1999). Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper. New Zealand: waikato
- [10] Exámenes PET CT y RM: ¿Qué son y cómo usarlos en la medicina diagnóstica? (2015).
- [11] Medicina nuclear: SPECT y PET en tumores primarios del Sistema Nervioso — NeuroWikia. (2010).
- [12] Northside Radiology Associates. (2016). MIBG scintiscan. Atlanta: Editorial team.

- [13] University of Washington School of Medicine. (2016). Gammagrafía de la tiroides. Seattle: Editorial Director y A.D.A.M.
- [14] Working paper series. (2000). Correlation - based feature selection for discrete and numeric class Machine Learning. New Zealand.
- [15] Cloud AutoML - Custom Machine Learning Models — AutoML — Google Cloud. (2018).
- [16] Convolutional Neural Networks for Visual Recognition. (2018).
- [17] Journal of Physics: Conference Series. (2016). Recent development of feature extraction and classification multispectral/hyperspectral images: a systematic literature review.
- [18] Purves, D., Augustine, G., Fitzpatrick, D., Katz, L., LaMantia, A., McNamara, J., Williams, S. (2018). Types of Eye Movements and Their Functions.
- [19] Chang, C., Member, S. (2010). Thyroid Segmentation and Volume Estimation in Ultrasound Images.
- [20] Raicea, R. (2018). Want to know how Deep Learning works? Here's a quick guide for everyone.
- [21] School of Electronics and Control Engineering Chang'an University (2017). Hyperspectral Image Classification Based on SpectralSpatial Feature Extraction. Xi'an, China.
- [22] School of Electronics and Control Engineering. (2013). Using Nonnegative Matrix Factorization with Projected Gradient for Hyperspectral Images Feature Extraction.
- [23] Department of Computer Science California State Polytechnic University. (2016). Oculomotor Plant Feature Extraction from Human Saccadic Eye Movements. Pomona, USA.
- [24] Bharathidasan College of Arts Science. (2016). Fusion of Big Data and Neural Networks for Predicting Thyroid. Tamilnadu, India.

- [25] Bharathidasan College of Arts Science. (2016). Fusion of Big Data and Neural Networks for Predicting Thyroid. Tamilnadu, India.
- [26] Bernard, O., Lalande, A., Zotti, C., Cervenansky, F. (2018). Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis.
- [27] School of Informatics and Computing, Indiana University. (2015). Temporal Pattern and Association Discovery of Diagnosis Codes using Deep Learning. IN, USA.
- [28] What is Watson. (2018). Retrieved from <https://www.ibm.com/watson/about/index.html> ©2018 IEEE.
- (2018). Delta Univ. for Science and Technology. Gamasa City, Egypt.
- [29] National Chung Hsing University. (2009). Optimal Grouping by using Genetic Algorithm and Support Vector Machines.
- [30] IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. (2010). Thyroid Segmentation and Volume Estimation in Ultrasound Images.
- [31] IEEE. (2017). Tiroid Kanserinde Bilgisayarlı Tomografi Temelli Yeni Oznitelikler Computerized Tomography Based Novel Features in Thyroid Cancer.
- [32] IEEE. (2009). Computer-Aided Diagnosis of Thyroid Malignancy Using an Artificial Immune System Classification Algorithm.
- [33] AIMBE. (2012). Automated Benign Malignant Thyroid Lesion Characterization and Classification in 3D ContrastEnhanced Ultrasound. IEEE.
- [34] The department of thyroid surgery. (2012). Discussion about misdiagnosed reasons and reoperation of thyroid cancer. Changchun city, China.
- [35] Department of Electronics and Computer Engineering, Gifu University Yanagido. (1992). Neural Network Approach for the ComputerAided Diagnosis of Coronary Artery Diseases in Nuclear Medicine.

- [36] Shandong Provincial Key Laboratory of Computer Networks. (2016). Web Identification Image Recognition Based on Deep Learning. Jinan, China.
- [37] College of Computer Science and Technology. (2012). Interface Schema Matching with the Machine Learning for Deep Web. Harbin, P.R. China.
- [38] Towards Data Science. (2019). Metrics to Evaluate your Machine Learning Algorithm.