



# Visión artificial aplicada a la clasificación de imágenes de melanomas

Grado en Ingeniería Informática

## Trabajo Fin de Grado

Autor:

Gonzalo Rodríguez Rovira

Tutor:

Ginés García Mateos

Colaborador experto:

Javier Andreu Ortego



**Facultad  
Informática  
Universidad  
Murcia**

# Visión artificial aplicada a la clasificación de imágenes de melanomas

---

## **Autor**

Gonzalo Rodríguez Rovira

## **Tutor**

Ginés García Mateos

*Departamento de Informática y Sistemas*

## **Colaborador experto**

Javier Andreu Ortego

*Médico*



Grado en Ingeniería Informática



# Preámbulo

Desde que era pequeño, he lidiado con una marca de nacimiento que me ha llevado a ir en ocasiones al dermatólogo. Esto me ha hecho darme cuenta de la importancia que tienen las enfermedades de la piel, especialmente los melanomas ya que son los más mortales con diferencia.

Estas experiencias personales me han motivado a escoger esta línea de investigación y por ende a la detección temprana de enfermedades de la piel. Además, considero que en la actualidad la piel es uno de los órganos más descuidados, con prácticas lesivas para esta como descuidar el uso de protectores solares, el exceso de exposición al sol para broncearse, cámaras de bronceado... entre muchas otras.

Asimismo, gran cantidad de la población presenta algún tipo de anomalía en la piel, la cual en algunos casos les preocupa, pero no se le da importancia ya que no suele ser nada y no acuden al médico solo para “eso”.

Por todos estos motivos y la situación actual de falta de atención médica, me han motivado a realizar este TFG en el que se busca crear una herramienta para poder ayudar a diagnosticar en una fase temprana enfermedades de la piel, teniendo en cuenta que no siempre va a ser correcto su diagnóstico y sus resultados deben ser contrastados por un médico.

Por lo tanto, la finalidad es crear una herramienta que pueda ser usada por médicos de atención primaria para tenerla de apoyo en casos de enfermedades cutáneas, además de poder ser utilizadas por particulares para tener una primera evaluación, teniendo en cuenta que nunca debe ser tomado como un diagnóstico preciso y en caso de duda acudir a un médico especializado.

# Resumen

Las enfermedades de la piel pueden ser letales si no son detectadas y tratadas a tiempo, especialmente el melanoma por su capacidad de producir metástasis rápidamente a otras partes del cuerpo. A diferencia de otros tipos de cáncer de piel, como el carcinoma basocelular y el carcinoma de células escamosas, que generalmente son más localizados y menos propensos a metastatizar, el melanoma tiende a crecer y diseminarse de manera más agresiva. Por lo que la supervivencia de los pacientes depende principalmente de una detección temprana.

Por ese motivo, el objetivo de este trabajo es la utilización de la Inteligencia Artificial para crear una herramienta tanto para los médicos de atención primaria como para el público en general, capaz de analizar imágenes con anomalías en la piel y clasificarlas dentro de malignos o benignos dando un posible diagnóstico, pero teniendo en cuenta que no siempre es exacto y debe ser consultado con un médico, ya que es simplemente una herramienta no un diagnóstico profesional.

Para desarrollar esta herramienta se ha usado un tratamiento de imágenes mediante redes neuronales convolucionales (CNN). Las imágenes con las que se han entrenado los modelos han sido obtenidas del conjunto de datos abiertos *International Skin Imaging Collaboration* (ISIC) seleccionado su versión del año 2019. Sobre esta base de datos se han realizado una serie de modificaciones en las que se han considerado malignos algunos diagnósticos que en la propia base de datos no eran considerados malignos, ya que esta solo considera malignos a los melanomas. Para este trabajo han sido considerados malignos además del diagnóstico melanoma, los diagnósticos carcinoma espinocelular y el carcinoma basocelular. La experimentación ha sido realizada en Google Colab, concretamente con su versión pro para agilizar todo el proceso de experimentación. Una vez finalizada la experimentación y la evaluación de las métricas de rendimiento y precisión se ha determinado que el modelo con mejores resultados ha sido el EfficientNet-B5. Finalmente, una vez obtenido un modelo de diagnóstico adecuado se ha creado una aplicación en Python, en la que es posible analizar de las imágenes con anomalías en la piel utilizando el modelo de diagnóstico previamente entrenado.

Los resultados de la evaluación concluyen con una herramienta que debe recibir imágenes dermatoscópicas de 512x512 píxeles. Una vez subidas a su sección de Analizar se obtendrá un diagnóstico maligno o benigno, teniendo alrededor de un 85% de tasa de acierto. Después del estudio de los metadatos y de la evaluación de resultados, se concluye que debe ser especialmente utilizada en pacientes adultos de 30 años o más, tanto en hombres como en mujeres y con tonos de piel clara.

En resumen, el objetivo de este trabajo es crear una primera versión de una herramienta de libre disposición, que pueda ser implantada en un contexto clínico en un entorno real en producción, al servicio de la comunidad médica, capaz de ayudar a diagnosticar cáncer de piel en etapas tempranas.

# Extended abstract

Skin cancer is a widespread health problem worldwide, characterized by its potential lethality and increasing incidence rates. Early detection and timely intervention are critical to mitigate its impact on patients' lives. Leveraging recent advances in artificial intelligence (AI) and deep learning, this article presents an overview of the development and implementation of an automated skin cancer diagnostic system adapted to dermoscopic images. By harnessing the power of machine learning algorithms, in particular convolutional neural networks (CNNs), the system aims to improve diagnostic accuracy, streamline clinical workflows, and ultimately improve patient outcomes.

## Introduction

Skin cancer is a major public health problem worldwide and its incidence has been steadily increasing over the past decades. Early detection and timely intervention are crucial to improve patient outcomes and reduce the morbidity and mortality associated with this disease. Traditional methods of skin cancer diagnosis rely heavily on visual inspection by qualified dermatologists, which can be time-consuming, subjective and prone to variability. To address these challenges, the integration of artificial intelligence (AI) and machine learning (ML) techniques into dermatological practice has emerged as a promising approach to improve diagnostic accuracy and efficiency.

In this context, the study focuses on the development and implementation of a machine learning-based skin cancer diagnostic tool from dermoscopic images. Dermatoscopy, also known as dermatoscopy, involves the examination of skin lesions using a handheld device equipped with a magnifying lens and a light source. Dermoscopic imaging provides a detailed view of the structural and morphologic features of skin lesions, allowing dermatologists to assess their characteristics and determine the likelihood of malignancy.

The main goal of the research is to harness the power of machine learning algorithms to analyze dermoscopic images and help healthcare professionals diagnose skin cancer more effectively. Leveraging state-of-the-art ML models trained on large datasets of annotated dermoscopic images, the aim is to develop a robust and accurate diagnostic tool capable of differentiating between benign and malignant skin lesions with high sensitivity and specificity.

Through the study, we aim to address several key challenges in the field of dermatology, including:

**Improving diagnostic accuracy:** traditional methods of skin cancer diagnosis are inherently subjective and can vary depending on the dermatologist's knowledge and experience. Leveraging machine learning capabilities, we aim to standardize the diagnostic process and improve accuracy by leveraging patterns and features extracted from large volumes of dermoscopic images.

**Facilitating early detection:** Early detection of skin cancer is crucial to improve patient outcomes and increase the likelihood of treatment success. By developing a machine learning-based diagnostic tool capable of detecting subtle signs of malignancy in dermoscopic images, we aim to facilitate early intervention and improve the prognosis of skin cancer patients.

**Improving accessibility:** Access to dermatological expertise can be limited in certain regions, especially in rural or underserved areas. By creating a user-friendly and accessible diagnostic tool that can be deployed on a variety of platforms, including desktops and mobile devices, the goal is to expand the reach of dermatologic care and provide timely support to healthcare professionals in diverse settings.

**Foster collaboration and knowledge sharing:** Machine learning-based diagnostic tools have the potential to promote collaboration and knowledge sharing among healthcare professionals by providing a standardized framework for skin cancer diagnosis. By integrating diagnostic algorithms into clinical practice, we aim to foster collaboration between dermatologists, pathologists and other healthcare professionals, ultimately improving patient care and outcomes.

In summary, our study represents a significant advancement in the field of dermatology by leveraging machine learning techniques to develop a sophisticated diagnostic tool for skin cancer. Through our research efforts, we aim to empower healthcare professionals with innovative tools and technologies that enhance diagnostic accuracy, promote early detection, and improve patient outcomes in the fight against skin cancer.

## **Data collection and preprocessing**

Data collection and preprocessing are fundamental steps in the development of any machine learning model, especially in medical applications where data quality and integrity are paramount. In the study, a large dataset of dermoscopic images obtained from the International Skin Imaging Collaboration (ISIC) database, a widely recognized repository of dermatological images for research purposes, was used.

The ISIC dataset contains a wide range of dermoscopic images covering various skin lesions, including benign and malignant lesions. These images are annotated with labels indicating the presence or absence of skin cancer, providing essential training data for our machine learning models.

To facilitate data preprocessing and improve computational efficiency, we adopted the use of TFRecords, a binary file format optimized for TensorFlow, a popular deep learning framework. TFRecords enable efficient storage and retrieval of large volumes of data by serializing and storing image data in binary format. This preprocessing step involved resizing the dermoscopic images to a standardized resolution of 256x256, 384x384, 512x512 and 768x768 pixels and normalizing the pixel values to ensure consistency and comparability between different images.

In addition, data augmentation techniques were applied to improve the diversity and robustness of our training data. Data augmentation consists of applying random transformations to the input images, such as rotation, translation, scaling, and flipping, to generate augmented versions of the original images. These augmented images introduce variability into the training dataset, thus reducing overfitting and improving the generalization ability of our machine learning models.

In addition, metadata associated with the dermoscopic images, including patient demographics, lesion characteristics, and clinical annotations, were leveraged to enrich the training dataset and provide additional context for the machine learning algorithms. By incorporating metadata into this study, we aimed to capture valuable information that could potentially influence diagnostic results and improve overall system performance.

Overall, data collection and preprocessing play a crucial role in the development of our machine learning-based skin cancer diagnostic tool. By retaining a high-quality dataset, standardizing image resolution, leveraging TFRecords for efficient data storage, and incorporating data augmentation and metadata enrichment techniques, the robustness, reliability, and generalizability of our models are ensured, ultimately improving their diagnostic accuracy and clinical utility.

## **Model development**

The most important piece of this project lies in the development of a robust and accurate machine learning model for skin cancer diagnosis. Taking advantage of advances in deep learning, we employ a convolutional neural network (CNN) architecture adapted to handle dermoscopic images with high fidelity and accuracy.

**Model architecture selection:** After extensive experimentation and evaluation, we chose the EfficientNet-B5 architecture as the diagnostic model architecture. EfficientNet-B5 represents a state-of-the-art CNN architecture recognized for its superior performance in image classification tasks while maintaining computational efficiency. Its depth and complexity allow it to capture intricate patterns and features present in dermoscopic images, essential for accurate diagnosis.

**Fine tuning and hyperparameter optimization:** To adapt the pre-trained EfficientNet-B5 architecture to our domain-specific task, we performed fine tuning by unfreezing and retraining certain layers while keeping others frozen to avoid overfitting. In addition, hyperparameters such as learning rate, image resolution, epochs, initial bias, and dropout were meticulously tuned through experimentation and validation to maximize model performance and convergence speed.

**Loss function selection:** The choice of an appropriate loss function is critical to effectively train a machine learning model. For this binary classification task (benign vs. malignant), we employ the binary cross-entropy loss function, which measures the discrepancy between the predicted and true labels. Binary cross-entropy loss is well suited to binary classification problems and facilitates optimization by gradient descent during training.

**Evaluation metrics:** To evaluate the performance and generalization capability of the model, standard evaluation metrics such as accuracy, recall, area under the accuracy-recall curve (AUC-PR) and area under the receiver operating characteristic curve (AUC-ROC) were used. But giving more weight to the recovery metric as one of the main objectives is to reduce false negatives. These metrics provide a complete picture of the predictive capabilities of the model, including its ability to correctly classify both benign and malignant cases while minimizing false positives and false negatives.

**Validation and cross-validation:** Rigorous validation procedures were employed to assess the robustness of the model's performance and mitigate overfitting. By splitting the dataset into multiple folds and training the model on different subsets while evaluating on the remaining data, we obtained more reliable estimates of performance metrics and model generalization.

Through meticulous model development and optimization, we were able to create a highly accurate and reliable diagnostic tool capable of discerning benign and malignant skin lesions with adequate accuracy. The combination of state-of-the-art architecture, transfer learning, fine tuning and comprehensive evaluation methodologies resulted in a machine learning model poised to make a significant impact in the field of dermatology by assisting physicians in the early and accurate detection of skin cancer.

## **Design and implementation of the application**

To finalize the project development, an intuitive and easy-to-use application designed to streamline the skin cancer diagnosis process using the trained machine learning model was developed. The application was designed with a focus on accessibility, functionality and efficiency, catering to both healthcare professionals and individuals concerned about the health of their skin.

**User interface design:** Using Python 3 and Qt Designer, a visually appealing and ergonomic user interface (UI) was created to enhance the user experience and facilitate smooth interaction with the application. The UI was carefully designed with clear navigation and intuitive controls, ensuring ease of use for users of all technical levels.

### **Application sections:**

- **Information:** as the home screen of the application, the information section provides users with essential information about the application's purpose, usage instructions, and limitations. In addition, users can access a downloadable file containing fifty labeled images for testing, which encourages hands-on participation and learning.

- **Analysis:** This section is the core functionality of the application, allowing users to upload dermoscopic images for analysis via the "Diagnose" button. The interface dynamically updates to display the diagnostic result (benign or malignant) along with a visual indicator (green for benign, red for malignant). The loaded image is also displayed as a reference, allowing users to visually confirm the analysis.



- **History:** Here, users can view a chronological record of the eight most recent images analyzed, including their file names and corresponding diagnoses. The history section provides users with a convenient way to track previous diagnoses and monitor trends over time, facilitating longitudinal evaluation of skin lesions.

**Real-time diagnostics:** The application leverages the trained machine learning model to perform real-time diagnostics of the uploaded images, enabling fast and accurate assessment of skin lesions. By seamlessly integrating the model into the application's backend, users benefit from rapid feedback on the benignity or malignancy of lesions, allowing them to make informed decisions regarding further evaluation or treatment.

**Visual information:** To enhance user understanding and confidence in the diagnosis, the application provides visual information in the form of color-coded indicators (green for benign, red for malignant) that accompany the diagnostic result. This visual indication allows users to quickly determine the nature of the lesion without relying solely on textual information, thus encouraging effective decision making.

**Scalability and accessibility:** The application has been designed with scalability and accessibility in mind, allowing for seamless integration of future enhancements and features. In addition, the application's compatibility with various operating systems and devices ensures broad accessibility, allowing users to conveniently access the diagnostic tool from their preferred platform.

By meticulously designing and implementing the application with a focus on user experience, functionality and accessibility, a powerful diagnostic tool has been created that allows users to assess skin lesions accurately and efficiently. The app's intuitive interface, real-time diagnostic capabilities and comprehensive functions make it a valuable asset in the field of dermatology, facilitating early detection and intervention in cases of skin cancer.

## **Assessment and future perspectives**

The developed system represents an example of significant progress in the automation of skin cancer diagnosis and offers potential benefits for both healthcare professionals and patients. Future lines of research include leveraging metadata in the creation and training of the diagnostic model to improve diagnostic specificity, distinguishing between different types of malignant lesions, especially melanomas, integrating multiple models for ensemble learning, the refinement of image preprocessing techniques to address common imperfections, another more ambitious idea is to create a large national database of dermoscopic images to continuously enrich the model training data and another simpler idea is to deploy it as a web application as it will improve accessibility and facilitate its widespread adoption in clinical settings.

## **Conclusions**

In conclusion, the automated skin cancer diagnostic system presented in this project exemplifies the potential of AI technology in healthcare as satisfactory results have been achieved. By combining state-of-the-art deep learning techniques with an intuitive application design, the system offers a viable solution to improve early detection rates and patient outcomes.

# Índice

<b>Preámbulo.....</b>	<b>1</b>
<b>Resumen.....</b>	<b>2</b>
<b>Extended abstract .....</b>	<b>3</b>
<b>Índice de figuras.....</b>	<b>10</b>
<b>Índice de tabla.....</b>	<b>11</b>
<b>1. Introducción.....</b>	<b>12</b>
1.1 Anatomía y fisiología de la piel .....	13
1.2 Dermatoscopia .....	14
1.3 Situación actual .....	15
1.4 Impacto.....	16
<b>2. Estado del arte.....</b>	<b>18</b>
2.1 Aprendizaje automático .....	18
2.2 Redes neuronales artificiales .....	19
2.3 Redes neuronales convolucionales.....	23
2.4 EfficientNet.....	26
2.5 Evaluación del modelo.....	27
2.6 Detección fototipo de piel.....	29
2.7 Aprendizaje automático aplicado al cáncer de piel.....	30
<b>3. Análisis de objetivos y metodología.....</b>	<b>32</b>
3.1 Metodología.....	33
3.2 Materiales y recursos .....	34
3.2.1 Entorno de desarrollo Google Colab.....	35
3.2.2 Análisis de la base de datos ISIC.....	35
<b>4. Diseño y resolución del trabajo realizado .....</b>	<b>37</b>
4.1 Creación del modelo de diagnóstico.....	37
4.1.1 Enriquecimiento y preparación ISIC-2019.....	37
4.1.2 Análisis ISIC-2019 .....	39
4.1.3 Preparación y carga de datos .....	43
4.1.4 Experimentación .....	45
4.1.5 Entrenamiento del modelo óptimo.....	48
4.1.6 Evaluación del modelo óptimo .....	49
4.2 Aplicación de diagnóstico.....	52
<b>5. Conclusiones y vías futuras .....</b>	<b>56</b>
<b>Bibliografía .....</b>	<b>58</b>
<b>Lista de Acrónimos y Abreviaturas .....</b>	<b>61</b>

# Índice de figuras

<b>Figura 1:</b> Capas de la piel y cánceres de piel. ....	13
<b>Figura 2:</b> Dermatoscopio smartphone y smartphone de mano. ....	14
<b>Figura 3:</b> Número estimado de nuevos casos de melanoma de piel y muertes en todo el mundo en 2020. ....	15
<b>Figura 4:</b> Esquema de un perceptrón. ....	20
<b>Figura 5:</b> (a) Conjunto representable, es clasificable con una sola recta (b) conjunto que no es representable. ...	21
<b>Figura 6:</b> Red neuronal de un solo perceptrón. ....	22
<b>Figura 7:</b> Red neuronal monocapa. ....	22
<b>Figura 8:</b> Red neuronal multicapa. ....	23
<b>Figura 9:</b> Aplicación capa convolucional. ....	24
<b>Figura 10:</b> Aplicación paso a paso de una convolución. ....	24
<b>Figura 11:</b> Aplicación capa de agrupación. ....	25
<b>Figura 12:</b> Aplicación de la agrupación máxima o max pooling. ....	25
<b>Figura 13:</b> Representación realista del proceso de convolución. ....	26
<b>Figura 14:</b> Arquitectura del modelo EfficientNet-B0. ....	27
<b>Figura 15:</b> Matriz de confusión. ....	28
<b>Figura 16:</b> Fototipo y clasificación usando el ITA. ....	30
<b>Figura 17:</b> Mosaico de imágenes de ISIC. ....	31
<b>Figura 18:</b> Grafica de metadatos ISIC-2019. ....	39
<b>Figura 19:</b> Cantidad/porcentaje de casos malignos en el dataset ISIC-2019. ....	41
<b>Figura 20:</b> Cantidad/porcentaje de casos benignos en el dataset ISIC-2019. ....	41
<b>Figura 21:</b> Cantidad/porcentaje de casos respecto al sexo en el dataset ISIC-2019. ....	42
<b>Figura 22:</b> Cantidad/porcentaje de casos respecto a los grupos de edad en el dataset ISIC-2019. ....	42
<b>Figura 23:</b> Cantidad/Porcentaje casos respecto ITA ISIC-2019. ....	43
<b>Figura 24:</b> Imágenes cargadas TFRecords. ....	44
<b>Figura 25:</b> Métricas de rendimiento sobre los conjuntos de entrenamiento y validación. ....	49
<b>Figura 26:</b> Matriz de dispersión modelo óptimo. ....	50
<b>Figura 27:</b> Distribución del diagnóstico real en falsos negativos. ....	50
<b>Figura 28:</b> Distribución del diagnóstico real en falsos positivos. ....	51
<b>Figura 29:</b> Distribución de los diagnósticos correctos frente a incorrectos en función de características. ....	52
<b>Figura 30:</b> Sección Información aplicación. ....	53
<b>Figura 31:</b> Sección Analizar aplicación. ....	54
<b>Figura 32:</b> Sección Analizar funcionamiento. ....	54
<b>Figura 33:</b> Sección Historial funcionamiento. ....	55

# Índice de tabla

<b>Tabla 1:</b> Distribución de diagnósticos de enfermedades de la piel en la base de datos ISIC. ....	34
<b>Tabla 2:</b> Metadatos originales ISIC-2019. ....	38
<b>Tabla 3:</b> Metadatos modificados ISIC-2019. ....	39
<b>Tabla 4:</b> Distribución de los tipos de diagnósticos en el dataset ISIC-2019 modificado. ....	40
<b>Tabla 5:</b> Distribución de los diagnósticos maligno/benigno en el dataset ISIC-2019 modificado. ....	40
<b>Tabla 6:</b> Distribución de los conjuntos para el entrenamiento y prueba de modelos. ....	44
<b>Tabla 7:</b> Experimentos Learning Rate.....	46
<b>Tabla 8:</b> Experimentos Learning Rate y Sesgo inicial. ....	46
<b>Tabla 9:</b> Experimentos Dropout. ....	47
<b>Tabla 10:</b> Experimentos Versiones del modelo. ....	47
<b>Tabla 11:</b> Experimentos Épocas.....	48
<b>Tabla 12:</b> Experimentos Resolución. ....	48
<b>Tabla 13:</b> Hiperparametros y métricas finales modelo óptimo sobre conjunto de validación. ....	49
<b>Tabla 14:</b> Métricas finales modelo óptimo sobre conjunto de prueba ....	50

# 1. Introducción

En la actualidad el cáncer es la principal causa de muertes. Según los datos de la Organización Mundial de la Salud (OMS) en 2020 fallecieron casi 10 millones de personas por esta enfermedad [1], es decir, casi una de cada seis personas falleció por cáncer. Dentro de todos los tipos de cánceres que existen, el melanoma es el 5º tumor más frecuente en hombres y el 6º en mujeres [2]. Los distintos tipos de cáncer de piel representan cerca del 50% de los tumores que se detectan en el mundo [3]. Por eso se debe considerar que existen más tipos de cánceres de piel además del melanoma, como puede ser el carcinoma espinocelular, al cual corresponden el 10-30% de cánceres cutáneos, el más común que es el carcinoma basocelular que representa el 80-90% de cánceres cutáneos [4].

En los últimos años los distintos tipos de cáncer de piel son más comunes y al mismo tiempo se presentan en pacientes más jóvenes. Por ese motivo es muy importante la prevención y detección temprana de estos tipos de cánceres. Considerando que normalmente el primer diagnóstico es realizado por un médico general que no es experto en este tipo de patologías, se puede apreciar la importancia del desarrollo de una herramienta para ayudar en la identificación y clasificación correcta de estas anomalías de la piel, ya que puede ayudar a un diagnóstico precoz de la enfermedad y en consecuencia disminuir la mortalidad.

Para la detección y diagnóstico de los distintos tipos de enfermedades cutáneas, los médicos emplean la dermatoscopia, que es una técnica de diagnóstico no invasiva que permite la observación detallada de las estructuras de la epidermis y la dermis mediante un sistema de amplificación de la imagen y una iluminación especializada, que elimina las distorsiones causadas por la reflexión y refracción de la luz en la superficie de la piel. Esto revela patrones de pigmentación y vascularización que no son visibles a simple vista. Esta técnica se emplea de manera rutinaria en el diagnóstico del melanoma, ya que se ha demostrado que mejora significativamente la precisión del diagnóstico en comparación con la observación a simple vista.

Es importante tener en cuenta que las diferencias visuales entre un melanoma y una lesión benigna de la piel en las imágenes dermatoscópicas pueden ser muy sutiles y casi imperceptibles, incluso para dermatólogos con experiencia. Por esta razón, una herramienta que brinde un diagnóstico de apoyo en casos de riesgo de cáncer de piel basándose en este tipo de imágenes sería una herramienta valiosa para apoyar a los expertos y a médicos no expertos en este tipo de patologías.

Para la realización de este TFG se ha contado con el apoyo del doctor Javier Andreu, médico de familia en Orihuela, con formación e intereses de investigación en dermatología.

## 1.1 Anatomía y fisiología de la piel

La piel es el órgano más extenso del cuerpo humano y, a su vez, forma su cubierta externa. Desempeña funciones esenciales, como la protección contra lesiones e infecciones, la regulación de la temperatura corporal y la separación del organismo del entorno externo entre otras. Actúa como una barrera vital que nos resguarda de factores externos, garantizando la integridad y el equilibrio del cuerpo.

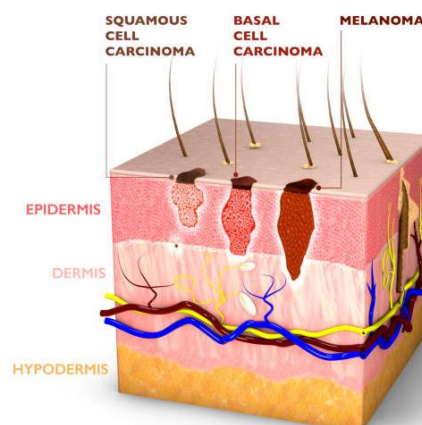
La piel está compuesta por tres capas distintas, como se puede ver en la figura 1, cada una con sus propias estructuras y funciones específicas:

**Epidermis:** La capa más externa y delgada de la piel. En ella se encuentran tres tipos principales de células. Las células escamosas, también conocidas como la capa córnea, se sitúan en la superficie de la piel y se renuevan constantemente. Las células basales se encuentran debajo de las células escamosas y son responsables de la regeneración celular. Por último, los melanocitos, distribuidos en toda la epidermis, son los encargados de darle a la piel su tonalidad característica.

**Dermis:** La capa intermedia de la piel y contiene una amplia variedad de estructuras, como vasos sanguíneos, linfáticos, folículos capilares, glándulas sudoríparas y sebáceas, nervios y fibras de colágeno. También alberga los receptores del dolor y del tacto, lo que permite la percepción sensorial.

**Hipodermis:** La capa más profunda de la piel y está compuesta principalmente por tejido conectivo, colágeno y adipocitos. Su función principal es la de proteger contra lesiones y mantener la temperatura corporal. Además, contiene células que pueden iniciar el proceso de reparación de las capas más superficiales en caso de lesión.

Los distintos tipos de cáncer de piel generalmente se originan en la capa más externa, es decir, en la epidermis. Sin embargo, si no se detectan y se tratan a tiempo, los melanomas pueden crecer y propagarse a capas más profundas de la piel [5].



**Figura 1:** Capas de la piel y cánceres de piel.

**Fuente:** [6].

## 1.2 Dermatoscopia

La dermatoscopia, también conocida como microscopía de epiluminiscencia o dermoscopia, es una técnica no invasiva utilizada para diagnosticar lesiones cutáneas, tanto benignas como malignas. Su aplicación mejora la precisión diagnóstica de lesiones hiperpigmentadas y facilita el diagnóstico temprano de lesiones malignas. Además, la fiabilidad del diagnóstico de un dermatólogo experto con la visualización a simple vista puede oscilar entre el 70% y el 85%. El uso de la dermatoscopia puede aumentar esta fiabilidad hasta un 92% [7].

Existen diversos tipos de dermatoscopios, que van desde portátiles hasta estereomicroscopios, algunos integrados con cámaras fotográficas o sistemas informáticos para evaluación y clasificación de lesiones. Aunque hay otros métodos no invasivos con mayor resolución, la dermatoscopia es preferida en la actualidad debido a su rapidez, precisión y facilidad de uso incluso por personal no especializado [8].

El dermatoscopio es esencialmente una lupa con una fuente de iluminación integrada, compuesto por una lente acromática de aproximadamente 10 aumentos, una fuente de luz y una fuente de alimentación. La penetración de la luz varía según su longitud de onda. Las luces halógenas amarillas pueden alterar el color de las lesiones, mientras que las luces Light Emitting Diode (LED) proporcionan una intensa luz blanca. Al atravesar la piel, la luz experimenta varios fenómenos como reflexión, refracción, dispersión y absorción. Por ejemplo, en pieles secas, la mayor parte de la luz se refleja en la capa córnea, mientras que, en pieles hidratadas y grasas, la luz puede penetrar hasta las capas más profundas de la dermis. Para una dermatoscopia efectiva, es crucial minimizar la reflexión, refracción y dispersión, favoreciendo la absorción de la luz para visualizar las capas dérmicas más profundas.

Lo más común es utilizar un dermatoscopio de mano que se puede conectar a un smartphone o una tablet. Algunos ejemplos de estos se pueden ver en la figura 2. Para obtener imágenes de calidad, es fundamental contar con un dermatoscopio que tenga una fuente de luz potente, una óptica de alta calidad y un visor lo suficientemente amplio y luminoso. Estas características garantizan la captura de imágenes nítidas y detalladas, lo que es esencial para un diagnóstico preciso.



**Figura 2:** Dermatoscopio smartphone y smartphone de mano.

**Fuente:** [9][10]

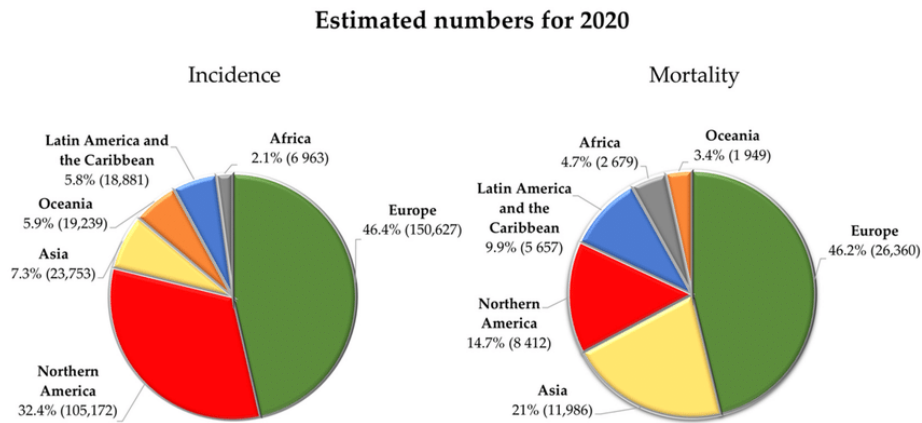


## 1.3 Situación actual

El cáncer de piel es una afección cada vez más frecuente en nuestra sociedad y representa uno de los principales desafíos de salud mundial.

A nivel mundial, entre el 1% y el 7% de los nuevos casos de cáncer primario diagnosticados corresponden a cáncer de piel, y hasta el 7% de las muertes por cáncer son atribuibles a este tipo de cáncer. Las tasas de incidencia y mortalidad varían considerablemente de un país a otro, pero en todos ellos han experimentado un aumento anual de alrededor del 3% en los últimos años.

Dentro de todos los tipos de cáncer de piel, el melanoma destaca como el más peligroso, ya tiene la mayor tasa de mortalidad. Según los datos publicados por la Organización Mundial de la Salud (OMS) para el año 2020, muestran una incidencia global de 324.635 casos de melanoma y una mortalidad de 57.043 casos [11].



**Figura 3:** Número estimado de nuevos casos de melanoma de piel y muertes en todo el mundo en 2020.

**Fuente:** [12]

Ambos estudios también revelan diferencias en la incidencia y morbilidad del melanoma según el sexo y la edad. Se observa que los hombres están más afectados por este tipo de tumor que las mujeres, especialmente en los grupos de edad mayores de 70 años [11][12].

En conclusión, el melanoma tiene una incidencia significativa, especialmente en países como Australia, Nueva Zelanda y regiones de Europa, y afecta principalmente a la población masculina mayor de 70 años. La disparidad en la incidencia en Australia en comparación con otras regiones se atribuye a la mayor prevalencia de personas con piel clara, altos índices de radiación ultravioleta en el ambiente y una cultura arraigada de bronceado. Además, se ha documentado que afecta más a los hombres debido a un menor uso de protectores solares y una menor frecuencia de exámenes de detección cutánea.

## 1.4 Impacto

En primer lugar, debemos tener en cuenta que, dada la naturaleza de un Trabajo de Fin de Grado (TFG), las actividades desarrolladas se encuentran limitadas en tiempo, ya que se realizan dentro de un semestre de curso universitario. Además, se cuenta con recursos humanos limitados, siendo un único estudiante con el apoyo de sus tutores. En términos de recursos económicos, se utiliza principalmente recursos computacionales de bajo coste para llevar a cabo el proyecto como es Google Colab. Por lo tanto, considerando las restricciones existentes, este trabajo puede ser visto como una “prueba piloto” de este concepto, es decir, un proyecto que, a pesar de contar con recursos limitados, tiene como objetivo demostrar la viabilidad de una idea.

La Inteligencia Artificial (IA) se utiliza cada vez más en problemas que pueden tener un impacto importante en la vida de las personas, como puede ser este proyecto en el que se centra en el ámbito médico buscando crear una herramienta para la detección temprana de melanomas, carcinomas espinocelulares o carcinomas basocelulares. Este ámbito la IA presenta varias ventajas frente a los humanos, como puede ser su capacidad para trabajar con gran volumen de datos (imágenes).

Si analizamos el impacto positivo resultante de una herramienta capaz de detectar enfermedades de riesgo en la piel, podemos concluir que una detección temprana de estas enfermedades aumenta de una manera significativa la calidad de vida de los pacientes, al tiempo que reduce también de forma considerable el consumo de medicamentos y también reduce los gastos médicos, con el consiguiente ahorro de recursos. Además de la posibilidad de reducir considerablemente la tasa de mortalidad de estas enfermedades y en consecuencia salvar vidas.

En cuanto al impacto ético-social, aun siendo los datos anónimos, un proyecto de esta índole debe abordar el desafío inherente al tratamiento de información personal, como las imágenes dermatoscópicas y sus metadatos asociados. Es crucial garantizar que estos datos sean utilizados exclusivamente para los fines especificados a los pacientes al recopilar dicha información, evitando cualquier otro uso no autorizado.

A pesar de la importancia del impacto en aspectos de sostenibilidad y ético-social, posiblemente el impacto en diversidad sea el aspecto más significativo a considerar. Varias investigaciones han señalado preocupaciones sobre el riesgo de sesgos no intencionados en estos modelos, los cuales podrían afectar de manera injusta a individuos pertenecientes a ciertos grupos.

Como ejemplo de posible discriminación en el ámbito de la herramienta, se puede considerar la información proporcionada por la Academia Americana de Dermatología [13]. Según esta fuente, aunque las personas de piel clara tienen un mayor riesgo de desarrollar cáncer de piel, la tasa de mortalidad entre las personas afroamericanas es considerablemente más alta. Se estima que la tasa de supervivencia a cinco años es del 73% para las personas afroamericanas, en comparación con el 90% para los estadounidenses blancos.

Es importante tener en cuenta que el modelo se basa principalmente en cómo aparecen

las lesiones cutáneas en segmentos mayoritarios de la población. Existe el riesgo de que no pueda predecir correctamente lesiones en pacientes pertenecientes a segmentos minoritarios por falta de casos en su base de datos. Por lo tanto, es crucial evaluar los sesgos presentes en los datos de entrenamiento y en el modelo que puedan afectar al color de la piel, sexo, edad u otros factores. Esto es fundamental para garantizar la precisión y equidad de la herramienta.

## 2.Estado del arte

### 2.1 Aprendizaje automático

El aprendizaje automático, también conocido como *machine learning*, es una disciplina dentro de la Inteligencia Artificial que se centra en el estudio y desarrollo de algoritmos informáticos capaces de mejorar automáticamente a partir de la experiencia adquirida durante su entrenamiento.

Según Nilsson [14], el aprendizaje automático implica que una máquina cambie su estructura, programa o datos de manera que su rendimiento mejore. Esto implica que la máquina adquiera capacidades de Inteligencia Artificial que le permitan realizar tareas como reconocimiento, diagnóstico, control robótico, planificación y predicción, entre otras. En el aprendizaje automático, los sistemas mejoran su rendimiento al modificar su estructura de forma automática o autónoma, utilizando experiencias previas o conocimientos almacenados. A diferencia de la programación convencional, en el aprendizaje automático no se proporciona un programa explícito para resolver una tarea. En su lugar, el sistema se autoajusta para mejorar su desempeño en función de la retroalimentación recibida, utilizando ejemplos o recompensas en lugar de secuencias de instrucciones programadas por el desarrollador.

La definición clásica sobre el concepto de aprendizaje en el contexto del aprendizaje automático fue dada en 1997 por el profesor Tom M. Mitchell [15]:

*“Diremos que un programa informático aprende de la experiencia  $E$  respecto de alguna tarea  $T$  y de alguna medida de rendimiento  $P$ , si su rendimiento en  $T$ , medido por  $P$ , mejora con la experiencia  $E$ ”.*

La tarea  $T$  en este contexto se puede definir como la correspondencia entre una entrada, definida por una serie de características, y una salida determinada. Respecto a las medidas de rendimiento  $P$ , se define como la métrica utilizada para evaluar los resultados obtenidos. Estas varían según la problemática tratada. La experiencia  $E$ , es el conjunto de datos utilizados para entrenar el modelo, a través de los cuales un algoritmo de aprendizaje automático puede aprender a realizar la tarea  $T$ . Se pueden establecer varios tipos de experiencia, siendo los principales:

**Aprendizaje supervisado:** En este enfoque, el conjunto de entrenamiento contiene la salida para cada uno de sus elementos. Si la salida es continua, se denomina tarea de regresión, mientras que si es discreta se trata de una tarea de clasificación.

**Aprendizaje no supervisado:** En este enfoque, el conjunto de datos de entrenamiento no incluye una salida específica, es decir, son datos sin etiquetar. El algoritmo debe inferir conocimiento a partir de estos datos, generalmente observando la estructura en su representación.

**Aprendizaje semisupervisado:** En este enfoque, el etiquetado de datos suele ser un proceso costoso, por lo que es común tener muchas instancias sin etiquetar y pocas

instancias etiquetadas en una organización. Los algoritmos de aprendizaje semisupervisados están diseñados para manejar datos que están parcialmente etiquetados.

**Aprendizaje reforzado:** En este enfoque, el sistema solo recibe datos sin etiquetar. Pero a través de interacciones con un entorno, el sistema recibe retroalimentación positiva o negativa sobre su desempeño y se adapta según esta retroalimentación.

Para nuestro caso de estudio, podemos decir que la tarea  $T$  es la correspondencia a la entrada que será una matriz con información de los píxeles de la imagen dermatoscópica con una salida determinada que dará el diagnóstico de la imagen diciendo si es maligna o benigna. Respecto a la medida de rendimiento  $P$  serían las métricas empleadas para evaluar al modelo. Para este proyecto ha sido empleado un aprendizaje supervisado para entrenar el modelo de diagnóstico.

Para comenzar, es necesario establecer cómo se compararán las predicciones realizadas por el modelo con respecto a la salida deseada, lo cual se lleva a cabo mediante la aplicación de una función de pérdida, la cual varía según el tipo de problema. El objetivo del proceso de entrenamiento es minimizar esta función de pérdida a lo largo de múltiples iteraciones (conocidas como épocas), ajustando cada vez más los parámetros del algoritmo utilizado. Durante las distintas pruebas realizadas durante la experimentación, se deben ajustar los hiperparámetros. Los hiperparámetros son configuraciones externas al modelo de predicción que controlan el proceso de entrenamiento y afectan la capacidad del modelo para aprender de los datos. A diferencia de los parámetros del modelo, que son aprendidos durante el entrenamiento, los hiperparámetros se establecen antes de que comience el proceso de entrenamiento y se mantienen constantes durante el mismo.

La modificación que ocurre en cada iteración para aproximarse a este mínimo se rige por un hiperparámetro denominado *Learning Rate* (LR) conocido como tasa de aprendizaje. Un learning rate elevado produce una tasa de aprendizaje elevada puede acelerar el proceso de aprendizaje, pero existe el riesgo de no alcanzar el mínimo global de la función de pérdida. Por otro lado, una tasa de aprendizaje reducida puede llevar a que el entrenamiento se estanque en un mínimo local y no logre acercarse al mínimo global deseado.

Otro aspecto crucial para considerar durante la fase de entrenamiento es el riesgo de sobreajuste. Este fenómeno ocurre cuando el modelo, después de ser entrenado, es capaz de hacer predicciones mucho mejores sobre los datos de entrenamiento que sobre datos nuevos. En este escenario, se dice que el modelo ha memorizado los datos de entrenamiento, pero no ha logrado generalizar la solución al problema planteado, ya que simplemente se ha aprendido los datos de entrenamiento. Suele ocurrir cuando el número de épocas es excesivo y se puede detectar cuando durante varias épocas rendimiento del modelo empeora.

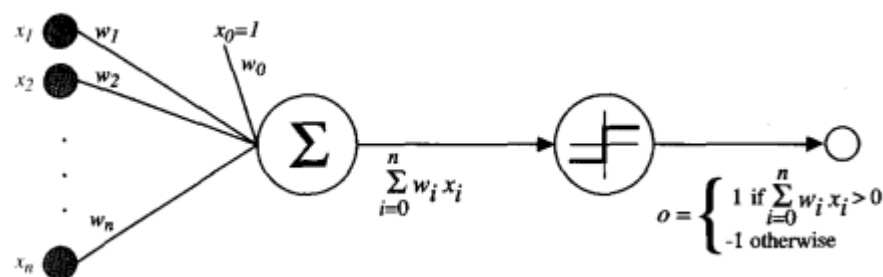
## 2.2 Redes neuronales artificiales

El estudio de las Redes Neuronales Artificiales (RNAs) ha sido inspirado en parte por la observación de sistemas biológicos de aprendizaje. Están contruidos por redes muy complejas de neuronas interconectadas. Las redes neuronales artificiales están formadas por

un conjunto densamente interconectado de unidades simples, donde cada unidad toma una serie de entradas de valores reales (posiblemente las salidas de otras unidades) y produce una sola salida de valor real (que puede convertirse en la entrada de muchas otras unidades). Este tipo de algoritmos son apropiados para resolver problemas donde el conocimiento puede ser impreciso o puede variar en el tiempo.

Para desarrollar una idea sobre esta analogía, consideremos algunos hechos de la neurobiología. Por ejemplo, se estima que el cerebro humano contiene una red interconectada densamente de aproximadamente  $10^{11}$  neuronas, cada una conectada, en promedio, a  $10^4$  otras. La actividad neuronal se excita o inhibe típicamente a través de conexiones con otras neuronas. Los tiempos de conmutación neuronal más rápidos se conocen en el orden de los  $10^{-3}$  segundos, bastante lentos en comparación con las velocidades de conmutación de las computadoras de  $10^{-10}$  segundos. Sin embargo, los humanos son capaces de tomar decisiones sorprendentemente complejas, sorprendentemente rápido [16].

La unidad básica de las RNA son los perceptrones, comúnmente llamadas neuronas por la analogía. En la figura 4 se encuentra la representación según el profesor Tom M. Mitchell. Un perceptrón toma un vector de entradas de valores reales, calcula una combinación lineal de estas entradas y luego produce un 1 si el resultado es mayor que algún umbral y -1 en caso contrario.



**Figura 4:**Esquema de un perceptrón.

**Fuente:** [15]

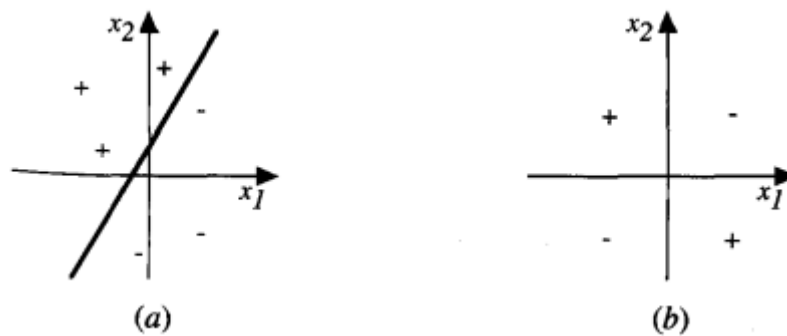
En este diagrama se representa una neurona  $i$  que recibe un conjunto de entradas  $X = \{x_1, x_2, \dots, x_n\}$ . Cada elemento del conjunto de entradas está ponderado por un conjunto de valores  $W = \{w_1, w_2, w_n\}$ . Estos valores  $\{w_j\}$  representan la importancia o peso asignado al valor de entrada  $x_j$  que llega a la neurona  $i$  desde una neurona  $j$  conectada a ella.

La neurona toma el conjunto de entradas  $X$  que llegan a ella y les aplica una función de entrada. El valor resultante sirve como entrada a una función de activación que genera un valor de salida  $y$ . Este valor, a su vez, puede servir como entrada a otras neuronas conectadas con ella, o puede ser el valor de salida de toda la red si no hay más neuronas conectadas después de ella.

Un solo perceptrón puede usarse para representar muchas funciones booleanas primitivas. Por ejemplo, si asumimos valores booleanos de 1 (verdadero) y -1 (falso), entonces una forma de usar un perceptrón de dos entradas para implementar la función AND es configurar

los pesos  $w_0 = -3$ ,  $w_1 = w_2 = 0.5$ . Este perceptrón puede representar la función OR si se altera el umbral a  $w_0 = -0.3$ . De hecho, AND y OR pueden verse como casos especiales de funciones m-de-n, es decir, funciones donde al menos m de las n entradas al perceptrón deben ser verdaderas. La función OR corresponde a  $m = 1$  y la función AND a  $m = n$ . Cualquier función m-de-n se puede representar fácilmente usando un perceptrón configurando todos los pesos de entrada al mismo valor (por ejemplo, 0.5) y luego estableciendo el umbral  $w_0$  en consecuencia.

Los perceptrones pueden representar todas las funciones booleanas primitivas AND, OR, NAND (NOT AND), y NOR (NOT OR). Sin embargo, algunas funciones booleanas no pueden ser representadas por un solo perceptrón, como la función XOR cuyo valor es 1 si y solo si  $x_1 \neq x_2$ , como se puede apreciar en la figura 5. Se debe tener en cuenta que el conjunto de ejemplos de entrenamiento linealmente no separables mostrado en la figura 5 (b) corresponde a esta función XOR. Para poder representar este tipo de operaciones se necesitan una red más elaborada, cada función booleana puede ser representada por alguna red de perceptrones de solo dos niveles de profundidad, en la que las entradas se alimentan a múltiples unidades y las salidas de estas unidades se ingresan a una segunda etapa final



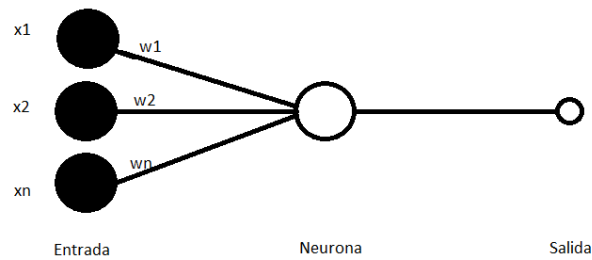
**Figura 5:** (a) Conjunto representable, es clasificable con una sola recta (b) conjunto que no es representable.

**Fuente:** [18]

En conclusión, los perceptrones básicos solo pueden representar regiones lineales, por lo que solo son útiles cuando los problemas han de ser sencillos y tiene poca aplicabilidad en el mundo real. Debido a eso, la arquitectura dependerá de la clase de problema a resolver y de la calidad de los resultados que se deseen, teniendo en cuenta el coste computacional en función a la complejidad de la estructura.

En general podemos agrupar las arquitecturas en tres conjuntos:

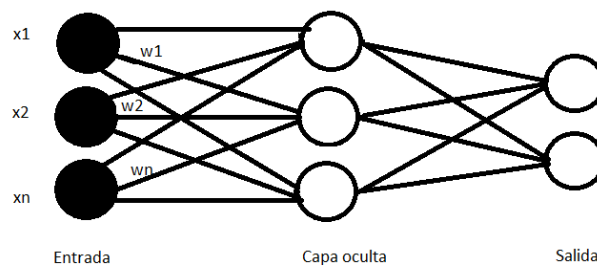
**Red con neurona única:** Esta es la configuración más básica, donde una sola neurona está conectada a un conjunto de entradas y produce una única salida. Este tipo de red puede realizar tareas simples, como funciones de regresión no lineales. En la figura 6 se puede apreciar un ejemplo.



**Figura 6:** Red neuronal de un solo perceptrón.

**Fuente:** Propia.

**Red monocapa:** Esta configuración, consta de una única capa de entrada, una capa oculta (capa que se encuentra entre la capa de entrada y la capa de salida de la red neuronal) de procesamiento con un número variable de neuronas, y una capa de salida con una o varias neuronas. Este tipo de redes pueden hacer predicciones sobre patrones de entrada más complejos y de mayor dimensionalidad en comparación con las redes de una sola neurona. En la figura 7 se puede apreciar un ejemplo.

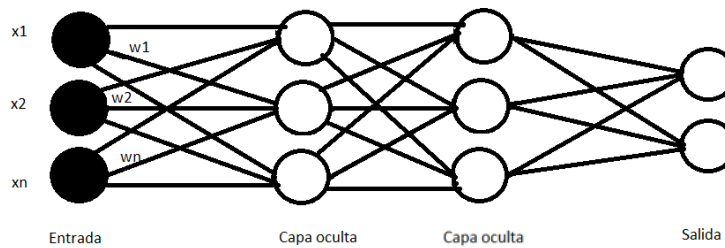


**Figura 7:** Red neuronal monocapa.

**Fuente:** Propia.

**Red multicapa:** Esta es la configuración se basa en la arquitectura de una sola capa. Sobre esta es posible incrementar el número de capas ocultas, lo que aumenta la complejidad de la red. Estas redes se utilizan para hacer predicciones sobre patrones aún más complejos, pero conllevan un mayor riesgo de sobreajuste. En la figura 8 se puede apreciar un ejemplo.





**Figura 8:** Red neuronal multicapa.

**Fuente:** Propia.

## 2.3 Redes neuronales convolucionales

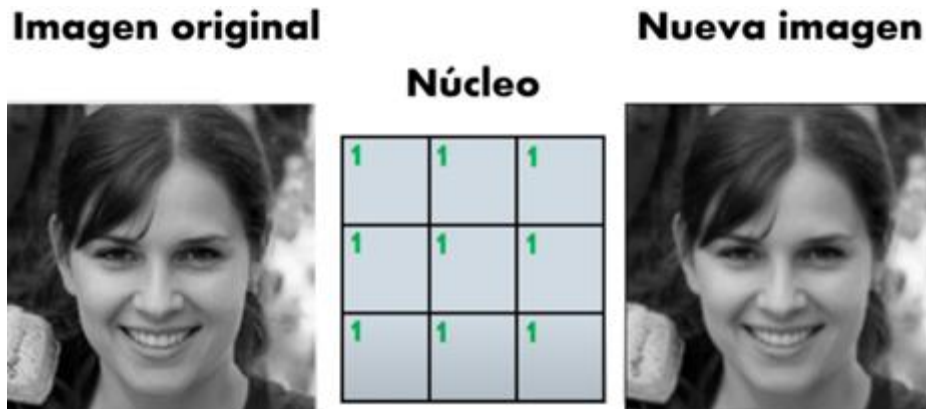
Las redes neuronales convencionales consideran cada píxel de la imagen como una característica individual e inconexa del resto de píxeles, lo cual no es apropiado ya que los píxeles por sí solos no representan características significativas. Esto puede resultar en la incapacidad de reconocer correctamente una imagen si es ligeramente distinta de las imágenes de entrenamiento. En cambio, para abordar este problema, es necesario extraer características de las imágenes y trabajar con ellas, como hacen las redes neuronales convolucionales (CNN) por sí solas.

Por ejemplo, si una red neuronal convencional recibe una imagen de 1000x1000 píxeles con tres canales de color, se requeriría un vector de 3 millones de dimensiones para almacenar esta imagen. Si consideramos una red neuronal con una capa oculta de 1000 neuronas, la dimensión de la primera matriz sería de  $1000 \times 3 \times 10^6 = 3 \times 10^9$  posiciones, lo que equivale al número de parámetros necesarios para configurar una red neuronal completamente conectada.

En contraste, las redes neuronales convolucionales son capaces de extraer características de los píxeles de manera más eficiente. Estas redes trabajan directamente con la imagen, extrayendo características en lugar de tratar cada píxel como una característica independiente.

Para conseguir la extracción de características, las CNN emplean dos tipos de capas extras:

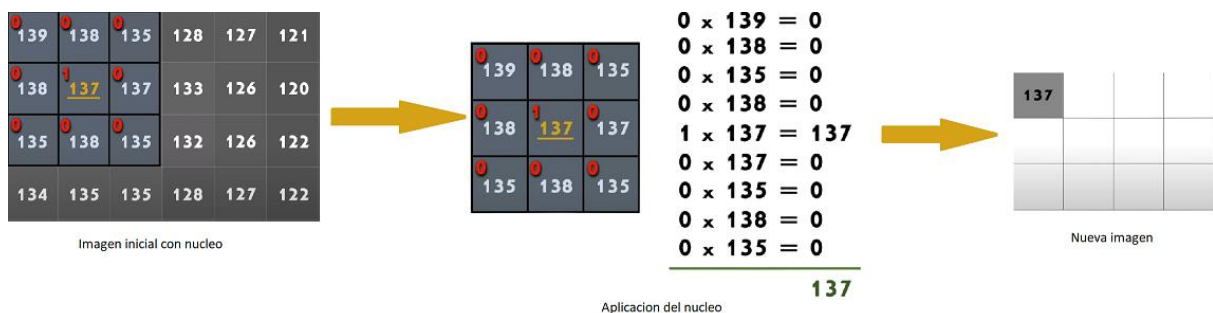
**Capa de convolución:** Es una parte fundamental que se encarga de extraer características de la imagen de entrada, principalmente ejes. Esta capa aplica una serie de filtros (*kernels*) a la imagen de entrada utilizando operaciones de convolución. Cada filtro detecta patrones específicos en la imagen, como bordes, texturas o formas, convirtiendo así la imagen de entrada en una representación más abstracta de las características relevantes. En la figura 9 se puede ver qué sucede al aplicarle un kernel de todos 1, logrando un ligero desenfoque.



**Figura 9:** Aplicación capa convolucional.

**Fuente:** [16]

Durante la convolución, el filtro se desliza a lo largo de la imagen de entrada y realiza multiplicaciones píxel a píxel entre los valores de la imagen y los pesos del filtro en cada ubicación. Estas operaciones de multiplicación se suman para producir un único valor en la imagen de salida, que representa la activación del filtro en esa ubicación específica de la imagen, todo ese proceso se puede apreciar en la figura 10.



**Figura 10:** Aplicación paso a paso de una convolución.

**Fuente:** Propia.

La capa de convolución puede tener varios filtros, lo que permite detectar múltiples características en paralelo. Se debe tener en cuenta que los pesos de los núcleos son generados por la propia CNN. Además, la salida de cada filtro se pasa a través de una función de activación no lineal, como *Rectified Linear Unit* (ReLU), para introducir no linealidades en el modelo.

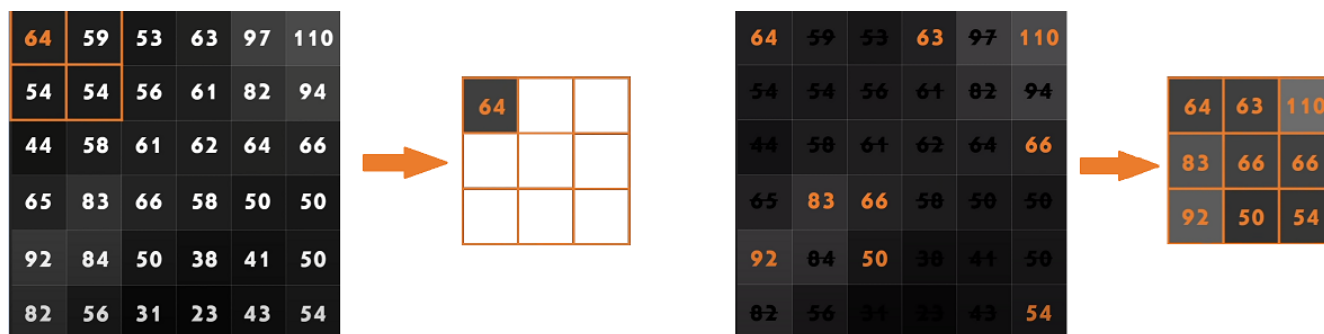
**Capa de agrupación:** Esta capa recibe las imágenes una vez han pasado por la capa de convolución. Sus dos objetivos principales son reducir el tamaño de la imagen y resaltar las características importantes. En la figura 11 se puede apreciar cómo se aplica cuatro veces.



**Figura 11:** Aplicación capa de agrupación.

**Fuente:** Propia.

La capa de agrupación actúa de manera independiente en regiones locales de la imagen y realiza una operación específica, como la suma, promedio o máximo, sobre estos grupos de píxeles. La operación más común es la agrupación máxima (*max pooling*), donde se selecciona el valor máximo de cada región local. Por ejemplo, en un *max pooling* de tamaño 2x2, se divide la imagen en cuadrados de 2x2 píxeles y se toma el valor máximo de cada cuadrado para crear una nueva imagen de salida con una resolución reducida. Todo este proceso que se realiza sobre los píxeles de la imagen se puede apreciar en la figura 12.



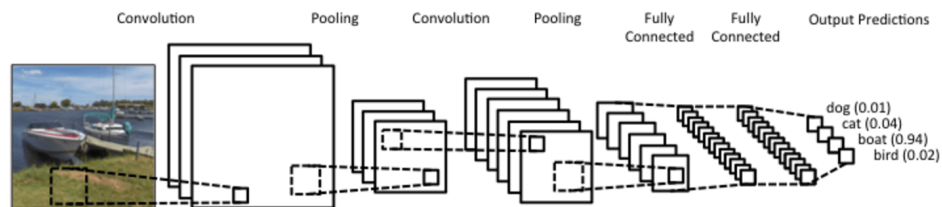
**Figura 12:** Aplicación de la agrupación máxima o *max pooling*.

**Fuente:** Propia.

En las CNN reales, es común tener múltiples capas de convolución y agrupación colocadas en secuencia, una detrás de otra. Esta estructura de capas repetidas permite que la red aprenda representaciones cada vez más complejas y abstractas de las características presentes en las imágenes.

Por lo general, estas capas se organizan en bloques repetitivos, donde cada bloque consiste en una o más capas de convolución seguidas por una capa de agrupación. Estos bloques se repiten varias veces en la arquitectura de la red.

La presencia de múltiples capas de convolución y agrupación permite que la red aprenda características de diferentes niveles de abstracción en la imagen. Las primeras capas de convolución pueden aprender características simples, como bordes o texturas, mientras que las capas más profundas pueden aprender características más complejas y abstractas, como formas o patrones más grandes. En la figura 13 se puede ver una representación general de cómo actúan las CNN.



**Figura 13:** Representación realista del proceso de convolución.

**Fuente:** [17]

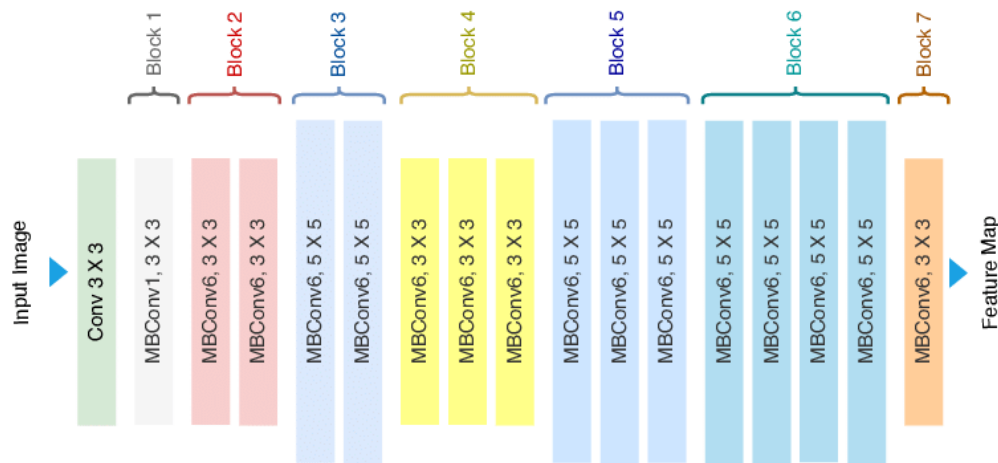
## 2.4 EfficientNet

EfficientNet es una familia de arquitecturas de redes neuronales convolucionales diseñadas para lograr un equilibrio óptimo entre precisión y eficiencia computacional. Estas redes se destacan por su capacidad para lograr un alto rendimiento en tareas de visión por computadora mientras mantienen un bajo coste computacional, lo que las hace adecuadas para su implementación en dispositivos con recursos limitados. EfficientNet fue propuesto en el año 2019 por Google AI en el artículo EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks para la International Conference on Machine Learning [18].

La eficiencia de EfficientNet se logra mediante el uso de un enfoque escalable que optimiza simultáneamente el ancho, la profundidad y la resolución de la red. Esto se logra mediante un proceso de búsqueda sistemática que ajusta automáticamente estos tres factores para encontrar la combinación óptima que maximice el rendimiento dado un límite de recursos computacionales.

En lugar de diseñar manualmente arquitecturas específicas para diferentes tamaños de modelos, como pequeño, mediano o grande, EfficientNet utiliza un enfoque de escalado compuesto que ajusta automáticamente el ancho, la profundidad y la resolución de la red para adaptarse a diferentes restricciones de recursos. Esto permite que EfficientNet logre un rendimiento competitivo en una variedad de tamaños de modelo. En la figura 14 se puede apreciar la arquitectura de EfficientNet-B0.

Gracias a ese principio, se creó la arquitectura EfficientNet, cuya modelo base es el modelo EfficientNet-B0, a partir de la cual se puede escalar hasta el modelo EfficientNet-B7.



**Figura 14:** Arquitectura del modelo EfficientNet-Bo.

**Fuente:** [19]

## 2.5 Evaluación del modelo

Las métricas de evaluación del rendimiento son esenciales para obtener un modelo óptimo en la clasificación. Por lo general, se divide el conjunto de datos en tres subconjuntos de manera aleatoria, asegurándose de mantener una proporción similar de categorías a las del conjunto original, es decir intentando que todos tengan una distribución de casos malignos y benignos similar.

Los tres subconjuntos en los que se divide son:

**Conjunto de entrenamiento:** Este conjunto se emplea para entrenar el modelo. Este utiliza este conjunto de datos para ajustar sus parámetros o pesos de manera que pueda aprender la relación entre las entradas y las salidas.

**Conjunto de validación:** Este conjunto se emplea para ajustar los hiperparámetros y evaluar el rendimiento del modelo. A diferencia del conjunto de entrenamiento, los datos en el conjunto de validación no se utilizan para entrenar el modelo, sino que se utilizan para ajustar la configuración del modelo y seleccionar los mejores hiperparámetros.

**Conjunto de prueba:** Este conjunto se emplea exclusivamente para evaluar al modelo, es decir el modelo solo lo ve en el último paso de evaluación.

El etiquetado de las imágenes se basa en los diagnósticos realizados por dermatólogos expertos, que representan la variable del mundo real que se modela con el sistema de aprendizaje automático. Una vez que se ha entrenado el modelo, se introduce el subconjunto de prueba, que no incluye las etiquetas de diagnóstico, y los resultados se suelen presentar en una matriz de confusión. Esta matriz relaciona las predicciones realizadas por el modelo con la realidad observada, como se puede apreciar en la figura 15.

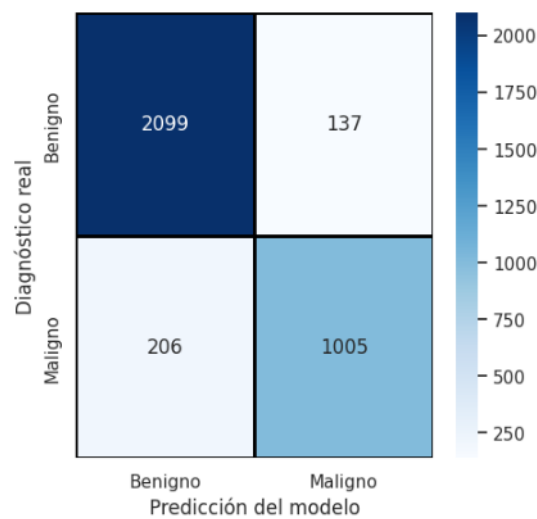
La matriz de confusión estará formada por los siguientes campos:

**Verdadero positivo (TP):** El modelo predice de manera correcta su diagnóstico, siendo este positivo, es decir es maligno. Corresponde al cuadrado de inferior derecho de la figura.

**Verdadero negativo (TN):** El modelo predice de manera correcta su diagnóstico, siendo este negativo, es decir es benigno. Corresponde al cuadrado superior izquierdo de la figura.

**Falso positivo (FP):** El modelo predice de manera errónea su diagnóstico, indicando el modelo que es maligno, pero en realidad es benigno. Corresponde al cuadrado superior derecho de la figura.

**Falso negativo (FN):** El modelo predice de manera errónea su diagnóstico, indicando el modelo que es benigno, pero en realidad es maligno. Corresponde al cuadrado inferior izquierdo de la figura.



**Figura 15:** Matriz de confusión.

**Fuente:** Propia.

Las métricas que han sido empleadas para evaluar los distintos modelos y así poder compararlos han sido las siguientes, todas varían entre 0 y 1, siendo el valor ideal 1:

**Exactitud (accuracy):** Indica el número de casos clasificados de forma correcta en comparación con todos los casos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

**Precisión (precision):** Indica el número de verdaderos positivos que son realmente positivos en comparación con el número total de todos los valores positivos predichos.

$$Precision = \frac{TP}{TP + FP}$$

**Sensibilidad (sensitivity):** Indica la cantidad de verdaderos positivos que el modelo ha clasificado respecto del número total de valores positivos. Representa la probabilidad de una imagen maligna sea identificada de manera adecuada.

$$Sensitivity = \frac{TP}{TP + FN}$$

**Especificidad (specificity):** Indica la cantidad de verdaderos negativos que el modelo ha clasificado respecto del número total de valores positivos. Representa la probabilidad de una imagen benigna sea identificada de manera adecuada.

$$Specifity = \frac{TN}{TN + FP}$$

Aunque el objetivo principal del sistema de diagnóstico es predecir correctamente todos los casos, es preferible que el sistema cometa un error de predicción falso positivo en lugar de un falso negativo, ya que estos pueden contribuir a retrasar la detección de la enfermedad. En este contexto, donde los falsos negativos se penalizan, la métrica de sensibilidad es la más relevante, por lo que tendrá un mayor peso respecto a las demás. Sin embargo, es importante tener en cuenta que la predicción de falsos positivos también conlleva costes aun siendo menos perjudiciales.

## 2.6 Detección fototipo de piel

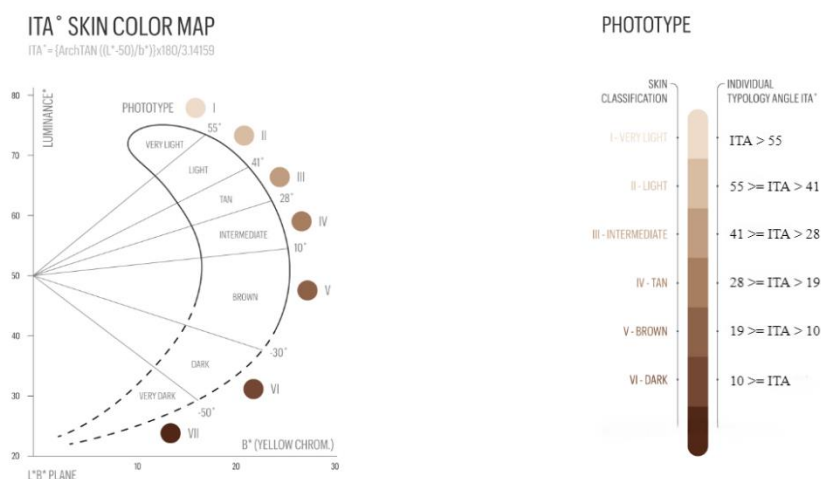
El fototipo es la capacidad de adaptación al sol que tiene cada persona desde que nace, es decir, el conjunto de características que determinan si una piel se broncea o no, y cómo y en qué grado lo hace. Cuanto más baja sea esta capacidad, menos se contrarrestarán los efectos de las radiaciones solares en la piel [20].

En la base de datos de imágenes empleada, cada imagen está asociada con varios metadatos, como el sexo, la edad, la ubicación de la lesión u otra información relevante sobre los historiales clínicos de los pacientes. Pero, uno de los datos que no se suele encontrar en estos metadatos es el fototipo de piel, es decir, el tono de color de la piel. Esta información puede ser crucial para evaluar la equidad de un modelo de diagnóstico.

Una aproximación para detectar el fototipo de piel en una imagen es mediante el cálculo del Ángulo de Tipología Individual (ITA) de la piel sana presente en cada imagen [21]. Se puede apreciar en la figura 16 la equivalencia de la ITA a los tonos de piel. Para calcular el tono de piel aproximado, se utiliza la siguiente fórmula:

$$ITA = \arctan\left(\frac{L - 50}{b}\right) \times \frac{180}{\pi}$$

Siendo L y b la conversión de los valores RGB de cada píxel al espacio de color CIELAB [22].



**Figura 16:** Fototipo y clasificación usando el ITA.

**Fuente:** [21],[23]

Para calcular el ITA, se toman parches de piel no afectada alrededor de los bordes de cada imagen, por ejemplo, 6 muestras de 20x20 píxeles. Se selecciona la muestra con el valor de ITA más alto que corresponde al tono de piel más claro y se toma como tono de piel de la imagen. Antes de tomar los parches, se aplica un preprocesado a la imagen que elimina el pelo. Este método reduce el impacto de las variaciones en la iluminación y es más simple de implementar que la segmentación de la imagen [24].

## 2.7 Aprendizaje automático aplicado al cáncer de piel

En el ámbito de la Inteligencia Artificial y visión artificial, existen varias aplicaciones de aprendizaje automático que pueden ser útiles para analizar imágenes. Este proyecto se enfoca en la tarea de clasificación binaria de imágenes para determinar si son malignas o benignas.

La clasificación binaria es un tipo de tarea de aprendizaje automático donde el modelo debe asignar una de dos clases posibles a cada instancia de entrada, maligna o benigna. En este contexto, la clasificación binaria implica determinar si una imagen de piel contiene o no algún cáncer de piel. Es decir, el modelo debe etiquetar la imagen como positiva si hay presencia y como negativa si no lo hay.

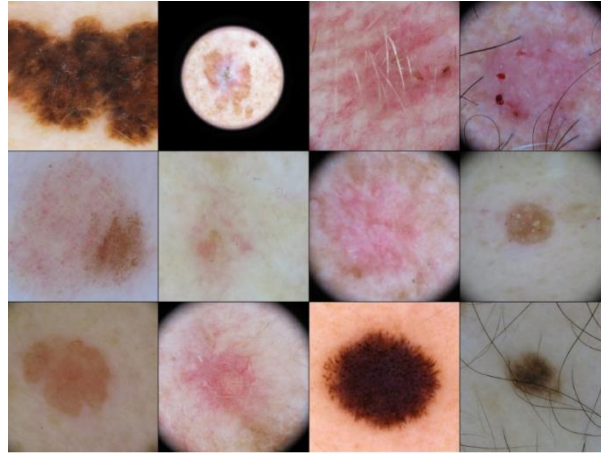
El principal estudio de referencia en este campo de detección de cáncer de piel mediante machine learning es *Dermatologist-level classification of skin cancer with deep neural networks*. Fue realizado en 2017 por la Universidad de Stanford, el modelo fue entrenado con más de 125.000 imágenes y basado en una adaptación de CNN InceptionV3 [25].

Los resultados del estudio fueron muy favorables obteniendo aproximadamente un 94% de tasa de acierto, muy similar al de dermatólogos experimentados [26].

Respecto a las fuentes de datos de imágenes, una iniciativa internacional para crear una base de datos de imágenes dermatoscópicas aptas para el diagnóstico automático es la ISIC [13]. Esta asociación surge de la colaboración entre más de 30 centros académicos y empresas en todo el mundo. El principal objetivo de la ISIC es reducir las muertes



relacionadas con el cáncer de piel y mejora la eficiencia en la detección temprana del cáncer utilizando imágenes digitales de la piel. En la figura 17 se pueden apreciar algunas de sus imágenes. Para lograr esto, la ISIC está desarrollando recursos para las comunidades de Dermatología, Ciencia de Datos e Inteligencia Artificial, incluyendo un gran repositorio de datos público con imágenes de piel, metadatos y diagnósticos precisos [25].



**Figura 17:** Mosaico de imágenes de ISIC.

**Fuente:** Propia

### 3. Análisis de objetivos y metodología

El primer objetivo de este proyecto es desarrollar un modelo de diagnóstico para la detección del cáncer de piel utilizando imágenes dermatoscópicas, con el fin de lograr una alta precisión en la clasificación de lesiones cutáneas como benignas o malignas. Para alcanzar este propósito, se cuenta con recursos disponibles en Kaggle, específicamente la base de datos ISIC-2019, que consta de 25.000 imágenes etiquetadas [13]. Se ha optado por utilizar la versión en formato TFRecord preparada por Chris Deotte [14], lo que permite un manejo más eficiente de las imágenes. Además, se ha empleado Google Colab, en su versión pro, para aprovechar el poder de las Unidades de Procesamiento Tensorial (TPUs) y agilizar el proceso de desarrollo y entrenamiento del modelo hasta obtener resultados satisfactorios.

Después de obtener un modelo de diagnóstico efectivo, el siguiente objetivo es crear una aplicación accesible para todos que pueda proporcionar diagnósticos precisos a partir de imágenes de piel basándose en el modelo de diagnóstico previamente creado. Para este propósito se empleará la herramienta Qt Designer para diseñar la interfaz de usuario. El resto de la lógica será implementada en Python 3, aprovechando la compatibilidad con el modelo existente el cual fue desarrollado también en Python 3.

Finalmente, tras alcanzar los objetivos mencionados anteriormente, el objetivo final es desarrollar una herramienta que pueda ser utilizada tanto por profesionales médicos como por personas no expertas. Esta herramienta consistirá en un sistema informático basado en IA capaz de proporcionar un diagnóstico preliminar a partir de una fotografía de la piel con alguna anomalía. Su propósito será servir como un detector de riesgo para los cánceres de piel, incluyendo melanoma, carcinoma basocelular y carcinoma espinocelular. Es crucial destacar que esta herramienta no pretende reemplazar un diagnóstico profesional realizado por un médico, sino que debe considerarse como una opinión preliminar, ya que sus diagnósticos pueden ser erróneos en ciertos casos. La herramienta estará compuesta de dos componentes principalmente:

**Modelo de diagnóstico:** Utilizando métodos de Inteligencia Artificial. Este modelo realiza de forma automática la clasificación binaria entre maligno y benigno. Se preparan los datos e imágenes provenientes de conjuntos de datos y anonimizados, los cuales incluyen diagnósticos realizados por dermatólogos expertos. A través de técnicas de aprendizaje automático supervisado, se genera un modelo de diagnóstico preciso.

**Aplicación:** La interfaz de la aplicación consta de tres secciones principales. En primer lugar, la sección de *Información* ofrece al usuario un contexto sobre el propósito de la aplicación, además de proporcionar un enlace a imágenes de ejemplo debidamente etiquetadas para poner a prueba la funcionalidad. La sección de *Análisis* permite a los usuarios cargar imágenes dermatoscópicas que serán evaluadas rápidamente por el modelo, mostrando el diagnóstico correspondiente. Por último, la sección de *Historial* muestra las últimas imágenes analizadas junto con sus diagnósticos.

En conclusión, el objetivo final es desarrollar una herramienta totalmente accesible por cualquiera y que sea capaz de ayudar a detectar en fases tempranas el cáncer de piel. De esta manera se podrán ahorrar recursos como ya se ha comentado y reducir la mortalidad de este tipo de enfermedades que se encuentran en auge.

### 3.1 Metodología

La metodología en cascada se ha elegido por ser el enfoque más adecuado para este proyecto debido a su naturaleza con un inicio y un final claramente definidos, así como a un resultado final establecido en colaboración entre el estudiante y los tutores del Trabajo de Fin de Grado. La metodología empleada en este proyecto sigue una aproximación en cascada, que consta de varias fases:

**Análisis de objetivos y metodología:** Se identifican los objetivos del proyecto, que incluyen el desarrollo de un modelo de diagnóstico de cáncer de piel y la creación de una aplicación accesible para proporcionar diagnósticos precisos. Se define la metodología para lograr estos objetivos, que implica el análisis de la base de datos ISIC-2019 y el desarrollo de un modelo de diagnóstico efectivo.

**Desarrollo del modelo de diagnóstico:** En esta fase, se preparan los datos e imágenes de la base de datos ISIC-2019, se emplean técnicas de aprendizaje automático supervisado para entrenar un modelo de clasificación binaria entre lesiones de piel malignas y benignas.

**Validación y evaluación:** Se realiza una validación del modelo de diagnóstico y de la aplicación para garantizar su precisión y fiabilidad. Se utilizan métricas de evaluación del rendimiento para medir la eficacia del modelo de diagnóstico y se realizan pruebas exhaustivas para garantizar la robustez y la exactitud de la aplicación.

**Desarrollo de la aplicación:** Se crea una aplicación en Python 3 que permite a los usuarios cargar imágenes dermatoscópicas. Esta aplicación utiliza el modelo de diagnóstico previamente desarrollado para proporcionar un diagnóstico preliminar de la lesión de piel cargada, indicando si es maligna o benigna.

**Despliegue y uso:** Una vez validado, el modelo de diagnóstico y la aplicación se ponen a disposición del público.

Concisamente, la metodología en cascada utilizada en este proyecto abarca desde el análisis inicial de los objetivos hasta el despliegue y uso de la herramienta desarrollada. Se prioriza la precisión del modelo de diagnóstico y la accesibilidad de la aplicación para ayudar a detectar el cáncer de piel en fases tempranas, con el objetivo de reducir la mortalidad asociada a esta enfermedad.

## 3.2 Materiales y recursos

Una vez explicada la finalidad del trabajo, el siguiente paso es detallar los recursos empleados para su desarrollo.

En primer lugar, el modelo de diagnóstico se ha creado, entrenado y evaluado utilizando el lenguaje de programación Python en su versión 3 empleando Google Colab, concretamente su versión Pro. Para el conjunto de datos, se ha empleado la base de datos de Kaggle "*Skin Lesion Images for Melanoma Classification*" [27], que consta de aproximadamente 25.000 imágenes diagnosticadas por médicos. Específicamente, se ha optado por utilizar la versión del año 2019 debido a su contenido relevante en términos de proporción de casos malignos y benignos, como se aprecia en la tabla 1. Además, se ha utilizado una versión en formato TFRecords, creada por Chris Deotte [28] que ayudará a optimizar el proceso. Dentro de estas versiones de la base de datos, se han seleccionado resoluciones de 256x256, 384x384, 512x512 y 768x768 debido a limitaciones de recursos.

Class	2017	2018	2019	2020
Melanoma	374	1113	4522	584
Atypical melanocytic proliferation	0	0	0	1
Cafe-au-lait macule	0	0	0	1
Lentigo NOS	0	0	0	44
Lichenoid keratosis	0	0	0	37
Nevus	0	0	0	5193
Seborrheic keratosis	254	0	0	135
Solar lentigo	0	0	0	7
Melanocytic nevus	0	6705	12875	0
Basal cell carcinoma	0	514	3323	0
Actinic keratosis	0	327	867	0
Benign keratosis	0	1099	2624	0
Dermatofibroma	0	115	239	0
Vascular lesion	0	142	253	0
Squamous cell carcinoma	0	0	628	0
Other / Unknown	1372	0	0	27124
Total	2000	10015	25331	33126

**Tabla 1:** Distribución de diagnósticos de enfermedades de la piel en la base de datos ISIC.

**Fuente:** [29]

Una decisión importante ha sido la modificación de los metadatos de la base de datos, esta decisión ha sido tomada con la ayuda de experta del médico Javier Andreu. Originalmente, solo se consideraban malignos los melanomas, lo cual no reflejaba completamente la realidad. Por lo tanto, se ha tomado la decisión de modificar los metadatos para considerar también como malignos a los carcinomas basocelulares y a los carcinomas espinocelulares. Se ha llegado a esta conclusión debido a que se considera más relevante tener una herramienta que informe si una anomalía en la piel es maligna o benigna en general, y no únicamente si se trata de un melanoma, dado que existen otros tipos de diagnósticos malignos además del melanoma.

### 3.2.1 Entorno de desarrollo Google Colab

La plataforma Google Colab [30] ha sido utilizada para el desarrollo de la experimentación en la búsqueda de un buen modelo de predicción, ya que proporciona potencia computacional necesaria, además permite que usuario pueda desarrollar y ejecutar código en Python directamente desde un navegador web, el funcionamiento de Google Colab se basa en la infraestructura de Google Cloud Platform (GCP) para proporcionar un entorno de notebook Jupyter basado en la nube además ofrece accesos a recursos informáticos [31].

La principal ventaja de Google Colab a parte de las mencionadas anteriormente es la posibilidad de usar TPUs en su versión Pro. Las TPUs son circuitos integrados diseñados específicamente por Google para acelerar de forma considerable las tareas relacionadas con el entrenamiento de modelos de aprendizaje automático. Estos chips, del tipo Application Specific Integrated Circuit (ASIC), se han creado con el propósito de mejorar el rendimiento y la eficiencia en el procesamiento de grandes volúmenes de datos.

La API de TensorFlow ha sido especialmente optimizada para aprovechar al máximo las capacidades de las TPUs, permitiendo así un rendimiento óptimo en la ejecución de modelos de aprendizaje automático. Sin embargo, es importante tener en cuenta que el flujo de datos hacia las TPUs está restringido a fuentes almacenadas en Google Cloud. Esto significa que no se permite la ingestión de datos desde repositorios locales o fuera de la infraestructura de Google Cloud. Por lo tanto, el dataset con el que se quiera trabajar debe ser subido a Google Cloud Storage (GCS).

### 3.2.2 Análisis de la base de datos ISIC

*The International Skin Imaging Collaboration* o también conocido como ISIC es una asociación entre la academia y la industria diseñada para utilizar la imagen digital de la piel para ayudar a reducir la mortalidad por cáncer de piel. Además, ISIC trabaja para alcanzar sus objetivos mediante el desarrollo y la promoción de estándares para la imagen digital de la piel, y mediante la participación de las comunidades de dermatología y visión por computadora hacia una mejora en el diagnóstico patrocinando retos anuales desde 2016. Los retos han crecido en complejidad y participación, empleando datos de entrenamiento de gran calidad supervisados por expertos y todo ha sido respaldado con conjuntos de datos con miles de imágenes y metadatos [32].

De entre todos los años en los que se han realizado los retos, ha sido seleccionada la base de datos del año 2019 ya que contiene el mayor número de imágenes etiquetadas, en total 25.331. Además, contiene el mejor ratio de benignos y malignos respecto a las demás, teniendo un total de 4522 melanomas, 3323 carcinomas basocelulares y 628 carcinomas espinocelulares, teniendo un ratio de 33,4% malignos/benignos, como se puede apreciar en la tabla 1.

Por lo tanto, con todos estos datos se decidió basarse en la base de datos de ISIC-2019 ya que es la que mejor se adecúa a este caso de estudio y facilitará el desarrollo del modelo de diagnóstico.

## 4. Diseño y resolución del trabajo realizado

### 4.1 Creación del modelo de diagnóstico

En este apartado se explicarán las distintas fases que fueron necesarias para el desarrollo del modelo de diagnóstico final y las técnicas empleadas.

#### 4.1.1 Enriquecimiento y preparación ISIC-2019

Una vez se ha decidido emplear la base de datos de ISIC-2019, se han de realizar una serie de cambios sobre esta base de datos para adecuarla más a este proyecto. Inicialmente los metadatos que nos proporciona la base de datos son los siguientes, en la tabla 2 se puede apreciar un ejemplo de estos campos:

**patient\_id:** Este campo indica el identificador del paciente; al ser un dato privado será siempre -1 por protección de datos.

**sex:** Este campo indica el sexo del paciente; podrá ser female, male o unknown.

**age\_approx:** Este campo indica la edad aproximada del paciente; tendrá siempre un valor numérico entre 0 y 85

**anatom\_site\_general\_challenge:** Este campo indica en qué parte del cuerpo se encuentra la lesión; puede tener los siguientes valores: anterior torso, upper extremity, posterior torso, lower extremity, nan, lateral torso, head/neck, palms/soles, oral/genital.

**diagnosis:** Este campo indica el diagnóstico correspondiente a la imagen; puede tener los siguientes valores: NV (Nevus), MEL (Melanoma), BKL (Benign keratosis), DF (Dermatofibroma), SCC (Squamous cell carcinoma), BCC (Basal cell carcinoma), VASC (Vascular lesion), AK (Actinic keratosis).

**benign\_malignant:** Este campo indica el valor del diagnóstico de la imagen, es decir si dado el diagnóstico de **diagnosis** es maligno o benigno. La base de datos solo considera maligno a los melanomas, puede tener dos valores: benign (el diagnóstico no es un melanoma) o malignant (el diagnóstico es un melanoma)

**target:** Este campo indica en binario el valor del campo **benign\_malignant**, pudiendo tener dos valores 0 en caso de que sea benigno y 1 en caso de ser maligno.

**tfrecord:** Este campo indica si la resolución de la imagen era 1024x1024 antes de cambiarla de tamaño; puede tener dos valores: impares tenían un tamaño original de 1024x1024 y pares los cuales no tenían ese tamaño.

**width:** Este campo indica el ancho de la imagen original antes de cambiar el tamaño del recorte del cuadrado central, tiene un valor numérico.

**height:** Este campo indica el alto de la imagen original antes de cambiar el tamaño del recorte del cuadrado central, tiene un valor numérico.

	image_name	patient_id	sex	age_approx	anatom_site_general_challenge	diagnosis	benign_malignant	target	tfrecord	width	height
0	ISIC_0000000	-1	female	55.0	anterior torso	NV	benign	0	4	1022	767
1	ISIC_0000001	-1	female	30.0	anterior torso	NV	benign	0	18	1022	767
2	ISIC_0000002	-1	female	60.0	upper extremity	MEL	malignant	1	0	1022	767
3	ISIC_0000003	-1	male	30.0	upper extremity	NV	benign	0	24	1022	767
4	ISIC_0000004	-1	male	80.0	posterior torso	MEL	malignant	1	14	1022	767

**Tabla 2:** Metadatos originales ISIC-2019.

**Fuente:** Propia.

Como se puede apreciar, muchos de estos metadatos no son relevantes ni para el desarrollo del modelo de predicción ni para comprender la información. Por eso mismo se pasará de tener los 11 campos de metadatos a solo los relevantes: **image\_name**, **sex**, **age\_approx**, **anatom\_site\_general\_challenge** y **diagnosis**. Además de esos campos, se considera también interesante poder incluir dos campos más y modificar el campo:

**age\_group:** Este campo indica la edad, pero agrupándola en secciones de 15 años, pudiendo tener valores (0-15) hasta (75-90).

**benign\_malignant:** Este campo ha sido modificado para no solo contemplar el **diagnosis** de melanoma maligno, ya que de entre todos los diagnósticos el SCC (*Squamous cell carcinoma*), BCC (*Basal cell carcinoma*) han sido considerados malignos ya que son otros tipos de cánceres de piel, por lo tanto indicará un valor de maligno al tener un diagnóstico de MEL, SCC o BCC y en caso contrario indicará benign.

**ITA:** Este campo indica el tono de la piel, pudiendo tener valores entre 1 y 6, correspondiendo valores mayores a pieles más oscuras [21].

Con todos estos cambios, los metadatos de la base de datos ISIC-2019 quedan con 8 campos: **image\_name**, **sex**, **age\_approx**, **anatom\_site\_general\_challenge**, **diagnosis**, **age\_group**, **benign\_malignant** y **ITA** como se puede apreciar en la tabla 3.



	image_name	sex	age_approx	anatom_site_general_challenge	diagnosis	age_group	benign_malignant	ITA
0	ISIC_0000000	female	55.0	anterior torso	NV	(45, 60]	benign	6
1	ISIC_0000001	female	30.0	anterior torso	NV	(15, 30]	benign	6
2	ISIC_0000002	female	60.0	upper extremity	MEL	(45, 60]	malignant	6
3	ISIC_0000003	male	30.0	upper extremity	NV	(15, 30]	benign	1
4	ISIC_0000004	male	80.0	posterior torso	MEL	(75, 90]	malignant	1

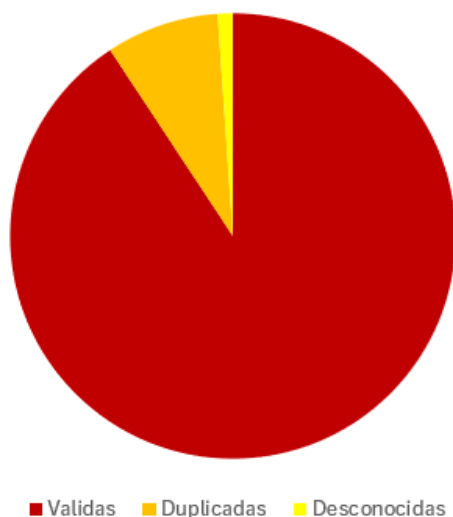
**Tabla 3:** Metadatos modificados ISIC-2019.

**Fuente:** Propia.

Por último, se han descartado algunos datos del total. En primer lugar, se han descartado toda la información relacionada con las filas que contienen en **image\_name** el prefijo de *\_downsampled* ya que son imágenes ya contenidas, modificando su posición/ángulo solamente. Estas imágenes son un total de 2074, lo que representa un total aproximado del 8% de las imágenes, quedando un total de 23.257.

Además de ese conjunto de datos, se han descartado todas las filas que contuvieran en alguna de sus columnas un campo *unknown*, ya que esa información no es precisa ni fiable, siendo un total de 277 filas y representando un total aproximado del 1% del total restante.

Finalmente, la base de datos con la que se trabajará tiene un total de 22.980 imágenes correctamente etiquetadas sin información duplicada ni datos desconocidos. En la figura 18 se muestra gráficamente el porcentaje de los datos disponibles, duplicados y descartados.



**Figura 18:** Gráfica de metadatos ISIC-2019.

**Fuente:** Propia.

### 4.1.2 Análisis ISIC-2019

Como ya se ha comentado en el punto anterior, finalmente tenemos un total de 22.980

imágenes con sus metadatos, distribuidos como se puede ver en la tabla 4.

Tipo	Cantidad	Porcentaje	benigno_maligno
NV	11355	49.41%	benigno
MEL	4114	17.90%	maligno
BCC	3317	14.43%	maligno
BKL	2214	9.63%	benigno
AK	864	3.76%	benigno
SCC	628	2.73%	maligno
VASC	249	1.08%	benigno
DF	239	1.04%	benigno

**Tabla 4:** Distribución de los tipos de diagnósticos en el dataset ISIC-2019 modificado.

**Fuente:** Propia.

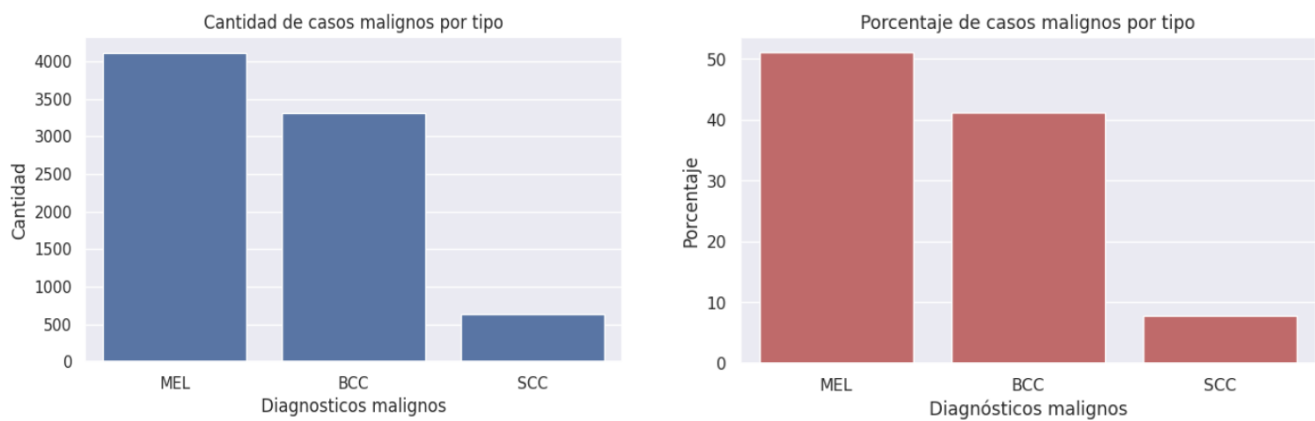
Al mismo tiempo, se puede apreciar un desbalance entre la cantidad de benignos y malignos como se aprecia en la tabla 5, por lo que tuvo que ser tratado en el desarrollo del modelo de diagnóstico.

Tipo	Cantidad	Porcentaje
Benigno	14921	64.93%
Maligno	8059	35.07%

**Tabla 5:** Distribución de los diagnósticos maligno/benigno en el dataset ISIC-2019 modificado.

**Fuente:** Propia.

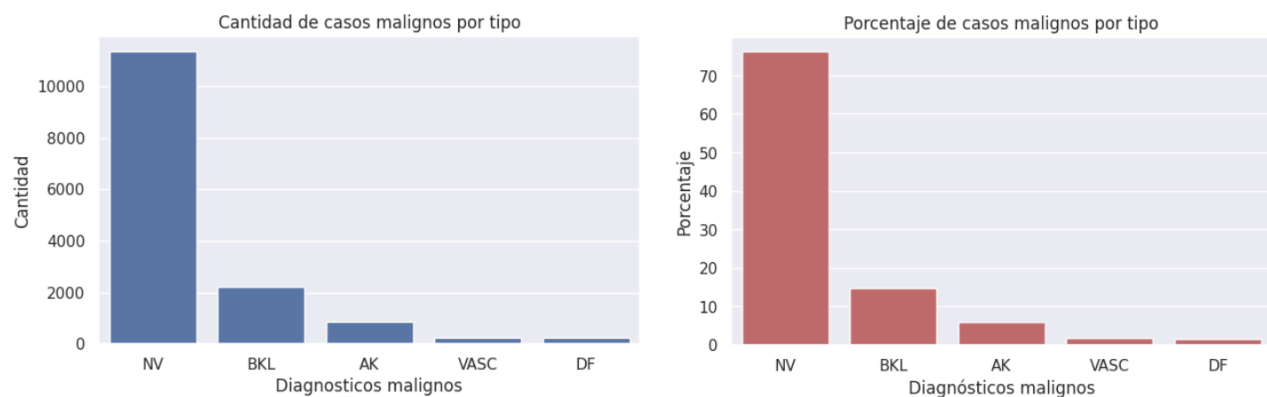
Dentro de los malignos podemos apreciar también un cierto desbalance entre los distintos tipos, sobre todo en el tipo SCC, como se puede ver en la figura 19.



**Figura 19:** Cantidad/porcentaje de casos malignos en el dataset ISIC-2019.

**Fuente:** Propia.

Además, dentro de los benignos también podemos apreciar un gran desbalanceo con el diagnóstico NV respecto a todos los demás, como se puede ver en la figura 20.

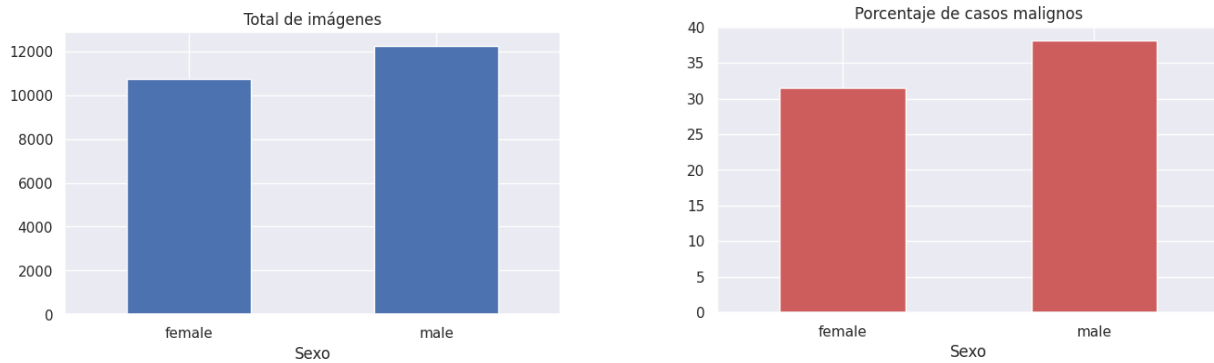


**Figura 20:** Cantidad/porcentaje de casos benignos en el dataset ISIC-2019.

**Fuente:** Propia.

Una vez analizados todos los distintos casos dentro de ambos diagnósticos, se analizarán las distribuciones de diagnóstico maligno y benigno dentro de las características más relevantes, **sex**, **age\_approx**, **age\_group**. Los resultados se pueden ver en las figuras 21-23.

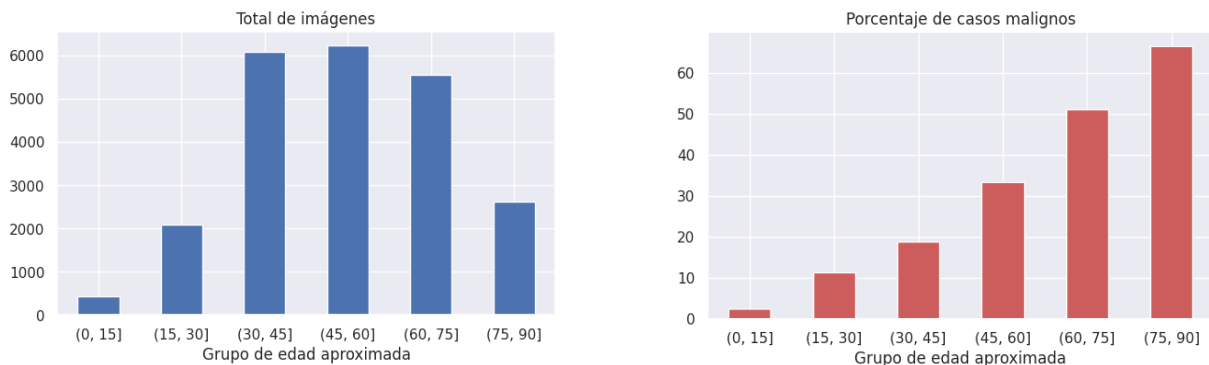
Como apreciamos en la figura 21, se puede ver que la cantidad de imágenes es similar para ambos sexos. También se puede apreciar que la cantidad de casos benignos y malignos es similar para ambos sexos. Con lo que se puede concluir que no hay una mayor incidencia de casos malignos en función del sexo y por lo tanto el sexo no es un parámetro que afecte ante la detección de enfermedades cutáneas.



**Figura 21:** Cantidad/porcentaje de casos respecto al sexo en el dataset ISIC-2019.

**Fuente:** Propia.

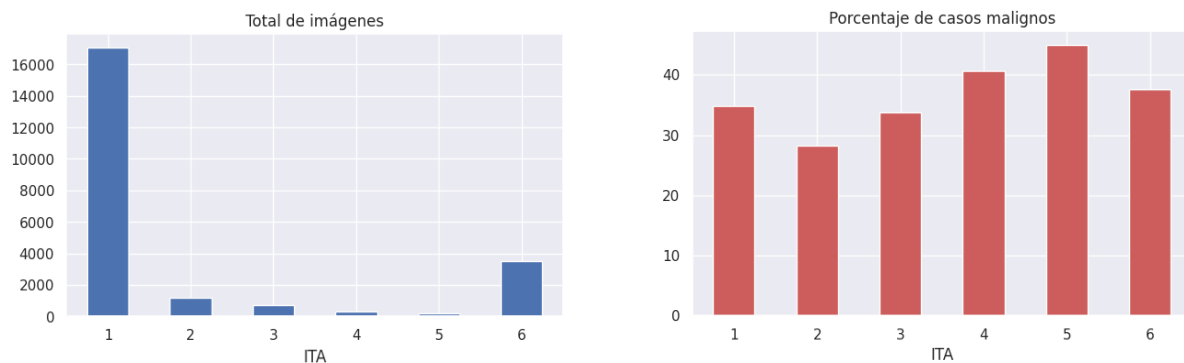
Como apreciamos en la figura 22, se observa que la mayor cantidad de imágenes se encuentra entre los rangos de edad de 30-90 años. En este caso, se puede ver una relación entre la edad y la cantidad de diagnósticos malignos. Con todo esto se puede concluir que la edad sí es una característica importante ante la detección de enfermedades cutáneas malignas. En edades tempranas menores a 30 años es más rara la presencia de afecciones malignas en la piel. Por el contrario, se puede apreciar una correlación entre el aumento de la edad y el porcentaje de casos malignos.



**Figura 22:** Cantidad/porcentaje de casos respecto a los grupos de edad en el dataset ISIC-2019.

**Fuente:** Propia.

En cuanto a la ITA, en primer lugar, se debe tener en cuenta que la ITA no ha sido proporcionada por la propia base de datos. Por lo tanto, Es muy probable que los valores de la ITA no sean correctos o estén distorsionados, es bastante posible que el tipo de ITA 6 esté aumentado debido a la presencia de bordes negros en algunas imágenes producidos por el dermatoscopio, como sucede en el estudio [21]. Una vez hechas estas aclaraciones, se puede apreciar un gran desbalanceo entre los tipos de piel, siendo mucho más común el tipo ITA 1, es decir, la piel clara. Por lo tanto, se puede concluir que la ITA posiblemente sea una característica que afecte ante la detección de enfermedades cutáneas, siendo más comunes en pieles claras.



**Figura 23:** Cantidad/Porcentaje casos respecto ITA ISIC-2019.

**Fuente:** Propia.

### 4.1.3 Preparación y carga de datos

Como se mencionó anteriormente, al utilizar las Unidades de Procesamiento Tensorial (TPU) en Google Colab, es imprescindible que la base de datos esté alojada en Google Cloud Storage, dado que el flujo de datos hacia las TPUs está restringido. Además, dada la gran cantidad de información (más de 25.000 imágenes y sus metadatos), se optó por utilizar el formato TFRecords [33][34].

Este formato consiste en datos binarios simples orientados a registros, ampliamente utilizado en aplicaciones de TensorFlow para el entrenamiento de datos debido a su mayor eficiencia en comparación con los archivos de imágenes originales. Otro gran beneficio del uso de TFRecords es que el tamaño de las imágenes es homogéneo en todo el conjunto de datos, siendo cuadradas y tomando la parte central si es necesario recortar. Esta homogeneidad en el tamaño de las imágenes favorece el rendimiento en el entrenamiento de las redes neuronales convolucionales. Concretamente, se utilizaron los conjuntos de TFRecords creados por Chris Deotte a partir del conjunto de datos ISIC-2019 [28].

Chris Deotte generó varias versiones de los TFRecords de ISIC-2019, organizándolos por resolución, desde grupos de 128x128 píxeles hasta 768x768 píxeles. Para este proyecto, se utilizaron las versiones de 256x256, 384x384, 512x512 y 768x768 píxeles. Cada conjunto de TFRecords contenía un archivo en formato .csv con los metadatos de todas las imágenes y treinta archivos de formato .tfrec que contenían las imágenes.

El primer paso para desarrollar un modelo de diagnóstico fue la carga de los TFRecords. Se cargaron un total de cuatro, que fueron las versiones comentadas anteriormente. De esos cuatro se extrajeron de los TFRecords los campos necesarios: **image\_name** el nombre de la imagen, **image** la propia imagen y **diagnosis** el diagnóstico asociado a la imagen. Una vez realizada la carga de los TFRecords y habiendo extraído la información de interés, se realizaron algunas comprobaciones para asegurar que la información había sido cargada de manera correcta. Se hizo un conteo del total de imágenes cargadas y se mostraron algunas imágenes con su nombre y su diagnóstico, como se ve en la figura 24.



**Figura 24:** Imágenes cargadas TFRecords.

**Fuente:** Propia.

Una vez comprobado que los datos han sido cargados correctamente fue necesario realizar un ajuste sobre el campo **diagnosis** como se realizó se explicó en el punto 4.2.2. De esta manera se reconoció como un diagnóstico maligno a los casos de MEL, SCC y BCC al mismo tiempo que se eliminaron los casos con diagnósticos *unknown* y los casos con *\_downsampled* en el campo **image\_name**.

Tras realizar esos ajustes, se dividió toda la información en tres conjuntos que serán usados para entrenar, validar y evaluar al modelo de diagnóstico como se aprecia en la tabla 6.

Conjunto	Total	Benigno	Maligno
Entrenamiento	16774 (72.25%)	10930 (65.16%)	5844 (34.84%)
Validación	2961 (12.75%)	1929 (65.15%)	1032 (34.85%)
Prueba	3483 (15%)	2269 (65.14%)	1214 (34.86%)

**Tabla 6:** Distribución de los conjuntos para el entrenamiento y prueba de modelos.

**Fuente:** Propia.

#### 4.1.4 Experimentación

Una vez realizados todos los cambios anteriores para preparar los datos con los que se va a desarrollar el modelo de diagnóstico, se inició la fase de experimentación. Para la experimentación se han tenido en cuenta los siguientes hiperparámetros:

**Resolución:** Este término se refiere a la selección del conjunto de datos. Como se mencionó previamente, se cargaron cuatro conjuntos de datos diferentes, cada uno con imágenes de diferentes resoluciones.

**Versión del modelo:** Controla la versión específica del modelo EfficientNet que se utilizará durante el entrenamiento o la inferencia del modelo.

**Épocas (Epochs):** Una época se refiere a una pasada completa a través de todo el conjunto de datos de entrenamiento durante el proceso de entrenamiento de la red neuronal. Durante una época, la red neuronal ve todas las muestras de entrenamiento una vez y actualiza los pesos en función de los errores calculados en la retropropagación.

**Tasa de aprendizaje (Learning Rate):** Controla la magnitud de los ajustes que se realizan en los pesos de la red durante el entrenamiento. Determina qué tan rápido o lento se actualizan los pesos en respuesta a los errores calculados durante la retropropagación. Una tasa de aprendizaje alta puede llevar a oscilaciones o divergencia, mientras que una tasa de aprendizaje baja puede llevar a un entrenamiento lento.

**Sesgo inicial (Initial Bias):** El sesgo (bias) es un término adicional en las neuronas que permite a la red aprender funciones más complejas. El sesgo inicial se refiere a los valores iniciales asignados a los sesgos de las neuronas en la red al comienzo del entrenamiento.

**Dropout:** Es una técnica de regularización utilizada durante el entrenamiento de redes neuronales para reducir el sobreajuste. Durante el entrenamiento, se apaga aleatoriamente un porcentaje de las neuronas en cada capa oculta, lo que hace que la red sea más robusta y menos propensa al sobreajuste al obligarla a aprender características más generalizadas. El valor de dropout generalmente se establece entre 0 y 1, donde 0 significa ninguna neurona se apaga y 1 significa que todas las neuronas se apagan.

Todos esos hiperparámetros se irán ajustando mediante experimentos concretos para ver qué valor a cada hiperparámetro consigue mejores resultados ante el conjunto de validación. Para determinar qué modelo tiene mejores resultados se emplean las siguientes métricas:

**Área bajo la curva ROC (AUC-ROC):** Es una medida de la capacidad de un modelo para distinguir entre clases positivas y negativas. La curva ROC representa la tasa de verdaderos positivos frente a la tasa de falsos positivos en varios umbrales de clasificación. El AUC-ROC es el área bajo esta curva. Cuanto más alto sea el valor del AUC-ROC (más cercano a 1), mejor será el modelo en términos de clasificación.

**Área bajo la curva de precisión y recuperación (AUC-PR):** Similar al anterior, pero utiliza la precisión y la recuperación en lugar de la tasa de verdaderos positivos y la tasa de falsos positivos. Es especialmente útil cuando hay un desequilibrio significativo entre las clases positivas y negativas.

**Precisión (Precision):** Es la proporción de verdaderos positivos (instancias correctamente clasificadas como positivas) respecto al total de instancias clasificadas como positivas (verdaderos positivos y falsos positivos). Cuanto más alta sea la precisión, menor será el número de falsos positivos.

**Recuperación o Sensibilidad (Recall):** Es la proporción de verdaderos positivos respecto al total de instancias que realmente son positivas (verdaderos positivos más falsos negativos). Cuanto más alta sea la recuperación, menor será el número de falsos negativos.

Tras detallar las métricas que serán utilizadas para evaluar los resultados de los experimentos, es importante resaltar que no todas tendrán igual importancia. Se le otorgará un peso superior a la métrica de **Recall**, dado que uno de los objetivos fundamentales de este modelo de diagnóstico es minimizar los falsos negativos. Estos casos representan individuos con un diagnóstico maligno erróneamente clasificados como benignos, lo cual constituye un grave error en este contexto médico.

El primer experimento realizado fue sobre el **Learning Rate**. Los resultados obtenidos con los distintos valores de **Learning Rate** sobre el conjunto de validación se pueden apreciar en la tabla 7.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
0	256x256	B0	15	0.00500	0	0.0	0.837	0.720	0.653	0.733
1	256x256	B0	15	0.00100	0	0.0	0.887	0.850	0.872	0.659
2	256x256	B0	15	0.00050	0	0.0	0.916	0.863	0.812	0.806
3	256x256	B0	15	0.00010	0	0.0	0.918	0.878	0.839	0.781
4	256x256	B0	15	0.00005	0	0.0	0.920	0.873	0.808	0.777

**Tabla 7:** Experimentos Learning Rate.

**Fuente:** Propia.

Como se puede apreciar en este experimento **Learning Rate** toma estos valores [0.00500, 0.00100, 0.00050, 0.00010, 0.00005], una vez obtenidas las distintas métricas, se procede a seleccionar los tres mejores resultados. Para este caso son la configuración 3, la 2 y la 4.

Para el siguiente experimento **Learning Rate** toma los valores de los mejores resultados del experimento anterior, además de un cuarto valor ya que se observó que los tres mejores eran los tres **Learning Rate** menores [0.0005, 0.0001, 0.00005, 0.000001]. También para este experimento se empleó el **Sesgo inicial**. Los resultados del experimento se puede apreciar en la tabla 8.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
5	256x256	B0	15	0.000500	1	0.0	0.908	0.859	0.857	0.726
6	256x256	B0	15	0.000100	1	0.0	0.921	0.873	0.834	0.783
7	256x256	B0	15	0.000050	1	0.0	0.923	0.878	0.834	0.769
8	256x256	B0	15	0.000001	1	0.0	0.819	0.676	0.723	0.410

**Tabla 8:** Experimentos Learning Rate y Sesgo inicial.

**Fuente:** Propia.

Una vez obtenidas las distintas métricas, se procede a comparar los resultados, tanto estos últimos como con los anterior, pudiendo apreciar que el **Sesgo inicial** activo no mejora los



resultados y además que el mejor **Learning Rate** sigue siendo 0.000100. Por lo tanto, la configuración con mejores resultados sigue siendo la 3.

El siguiente experimento está dedicado al **Dropout**. En este experimento tomará los valores de [0.2, 0.3, 0.4, 0.5], al mismo tiempo se mantendrá el **Learning Rate** a 0.000100 y no se empleará **Sesgo inicial** al no haber mostrado ninguna mejora. Los resultados del experimento se pueden apreciar en la tabla 9.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
9	256x256	B0	15	0.000100	0	0.2	0.922	0.876	0.829	0.786
10	256x256	B0	15	0.000100	0	0.3	0.919	0.875	0.815	0.800
11	256x256	B0	15	0.000100	0	0.4	0.924	0.886	0.825	0.801
12	256x256	B0	15	0.000100	0	0.5	0.927	0.879	0.774	0.859

**Tabla 9:** Experimentos Dropout.

**Fuente:** Propia.

Una vez obtenidas las distintas métricas, se procede a comparar los resultados tanto estos últimos como con los anterior, pudiendo apreciar que el **Dropout** mejora los resultados, siendo el valor de 0.4 el que obtiene un mejor resultado junto con los parámetros anterior. Por lo tanto, la configuración con mejores resultados es la 11.

El siguiente experimento está dedicado a la **Versión del modelo**. En este experimento se tomarán las distintas versiones de EfficientNet [1, 2, 3, 4, 5, 6, 7], al mismo tiempo que se mantendrán los parámetros anteriores de la configuración 11. Los resultados del experimento se pueden apreciar en la tabla 10.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
13	256x256	B1	15	0.000100	0	0.4	0.932	0.882	0.772	0.855
14	256x256	B2	15	0.000100	0	0.4	0.930	0.887	0.832	0.818
15	256x256	B3	15	0.000100	0	0.4	0.932	0.890	0.873	0.781
16	256x256	B4	15	0.000100	0	0.4	0.936	0.894	0.820	0.858
17	256x256	B5	15	0.000100	0	0.4	0.932	0.903	0.866	0.797
18	256x256	B6	15	0.000100	0	0.4	0.930	0.903	0.863	0.778

**Tabla 10:** Experimentos Versiones del modelo.

**Fuente:** Propia.

Una vez obtenidas las distintas métricas, se procede a comparar los resultados tanto estos últimos como con los anteriores. Se pudo observar que no fue posible realizar el experimento con la versión 7 de EfficientNet por fallos en el entorno producidos por falta de memoria. Se pudo apreciar que la mejor **Versión del modelo** es B5 mejorando los resultados. Por lo tanto, la configuración con mejores resultados es la 17.

El siguiente experimento está dedicado a las **Épocas**. En este experimento se tomarán todas las características de la mejor configuración anterior, es decir la 17 y se le modificaron el número de épocas durante su entrenamiento tomando valores [10,13,16,19,22,25]. Los resultados del experimento se pueden apreciar en la tabla 11.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
19	256x256	B5	10	0.000100	0	0.4	0.936	0.901	0.871	0.783
20	256x256	B5	13	0.000100	0	0.4	0.928	0.894	0.840	0.810
21	256x256	B5	16	0.000100	0	0.4	0.930	0.899	0.885	0.799
22	256x256	B5	19	0.000100	0	0.4	0.936	0.896	0.828	0.856
23	256x256	B5	22	0.000100	0	0.4	0.935	0.898	0.857	0.824
24	256x256	B5	25	0.000100	0	0.4	0.931	0.892	0.865	0.823

**Tabla 11:** Experimentos Épocas.

**Fuente:** Propia.

Una vez obtenidas las distintas métricas, se procede a comparar los resultados tanto estos últimos como con los anteriores, pudiendo observar una ligera mejora al tener 16 **Épocas**. Por lo tanto, la configuración con mejores resultados es la 21.

Como último experimento está dedicado a la **Resolución**. En este experimento se tomarán todas las características de la mejor configuración anterior, siendo la 21 y se modificarán las resoluciones de las imágenes tomando valores de [384x384, 512x512, 768x768]. Los resultados del experimento se pueden apreciar en la tabla 12.

	resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
25	384x384	B5	16	0.000100	0	0.4	0.943	0.908	0.852	0.837
26	512x512	B5	16	0.000100	0	0.4	0.936	0.911	0.892	0.804

**Tabla 12:** Experimentos Resolución.

**Fuente:** Propia.

Una vez obtenidas las distintas métricas, se procede a comparar los resultados tanto estos últimos como con los anteriores, se pudo observar que no fue posible realizar el experimento con la resolución de 768x768 debido a la falta de memoria del entorno. Sin embargo, se pudo apreciar una mejora al emplear la **Resolución** de 512x512. Por lo tanto, la configuración con mejores resultados es la 26.

Tras la realización de todos los experimentos se puede ver que la configuración 26 es la que obtiene los mejores resultados, por lo tanto, se empleará esa configuración para crear el modelo de diagnóstico óptimo.

### 4.1.5 Entrenamiento del modelo óptimo

Una vez terminado todo el proceso de experimentación donde se han ajustado todos los hiperparámetros del modelo, se entrenará con la configuración óptima al modelo:

- **Resolución:** 512x512 píxeles
- **Versión del modelo:** EfficieNet-B5
- **Épocas (Epochs):** 16
- **Tasa de aprendizaje (Learning Rate):** 0.0001

- **Sesgo inicial (Initial Bias):** 0
- **Dropout:** 0.4

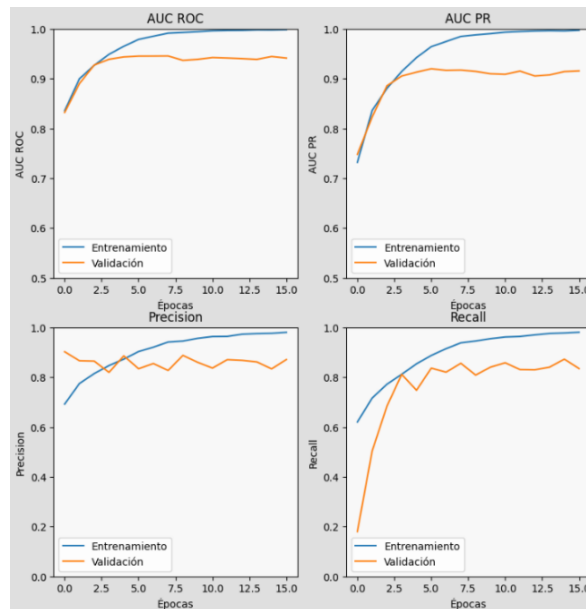
Tras realizar el entrenamiento del modelo con esos parámetros, se obtuvieron los siguientes resultados que se muestran en la tabla 13.

resolucion	version_modelo	epocas	lr	bias	dropout	val_auc_roc	val_auc_pr	val_precision	val_recall
512x512	B5	16	0.0001	0	0.4	0.941	0.916	0.872	0.835

**Tabla 3:** Hiperparámetros y métricas finales modelo óptimo sobre el conjunto de validación.

**Fuente:** Propia.

Finalmente se muestra en la figura 25 cómo han ido evolucionando a lo largo del tiempo las cuatro métricas sobre los conjuntos de entrenamiento y validación, donde se aprecia que las métricas en ambos conjuntos fueron estables y se también se aprecia que la mayor parte de la mejora sucedió en las primeras épocas.

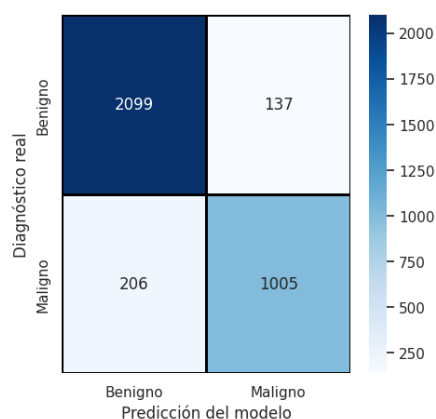


**Figura 25:** Métricas de rendimiento sobre los conjuntos de entrenamiento y validación.

**Fuente:** Propia.

#### 4.1.6 Evaluación del modelo óptimo

Tras obtener el modelo de predicción óptimo en base a los experimentos, el siguiente paso fue evaluarlo frente al conjunto de prueba, el cual contiene imágenes que nunca había visto durante su entrenamiento. En la figura 26 se puede ver la matriz de dispersión correspondiente al modelo.



**Figura 26:** Matriz de confusión obtenida con el modelo óptimo.

**Fuente:** Propia.

Cuando se evaluaron las métricas sobre este modelo óptimo con el conjunto de prueba, se puede observar que se obtuvieron unos valores similares a los obtenidos con el conjunto de validación , como se aprecia en la tabla 14.

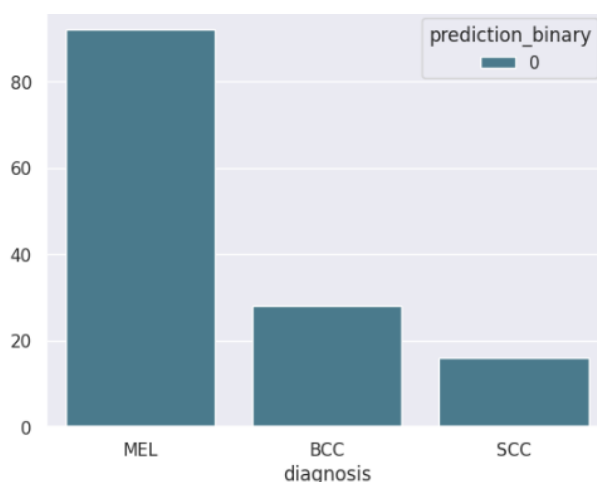
Métrica	AUC ROC	AUC PR	Precisión	Sensibilidad
Valor	0.96	0.931	0.846	0.888

**Tabla 14:** Métricas finales modelo óptimo sobre conjunto de prueba.

**Fuente:** Propia.

Una vez vista la matriz de dispersión y las métricas, se analizarán en profundidad los casos de errores tanto los falsos negativos como los falsos positivos.

Como se ha visto anteriormente, hay un total de 206 falsos negativos. Su distribución dentro de las tres clases malignas se aprecia en la figura 27.



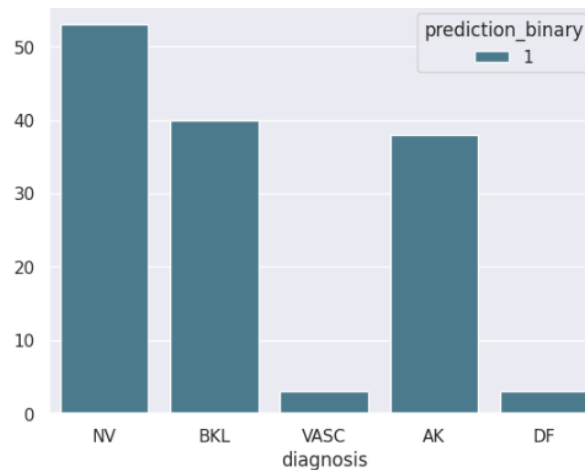
**Figura 27:** Distribución del diagnóstico real en falsos negativos.

**Fuente:** Propia.

Se puede apreciar claramente que la mayoría de los casos de falsos negativos corresponden a diagnósticos de melanomas y en mucha menor medida a carcinomas de células basales y carcinomas espinocelulares. Al ser contrastado con los datos que se vieron

en el punto 4.2.3, podemos darnos cuenta de que la mayor parte de diagnósticos malignos son melanomas. Por lo tanto, es normal que la mayor parte de falsos negativos correspondan a esta clase. Lo que es destacable es la similitud entre falsos negativos de carcinomas ya que la cantidad de carcinomas basocelulares es muy superior a la de carcinomas espinocelulares, por lo que se puede intuir que al modelo le cuesta más distinguir a los carcinomas espinocelulares, seguramente producido por la falta de casos de ejemplo.

Respecto a falsos positivos, hay un total de 137 casos. Su distribución dentro de las cinco clases benignas se aprecia en la figura 28.

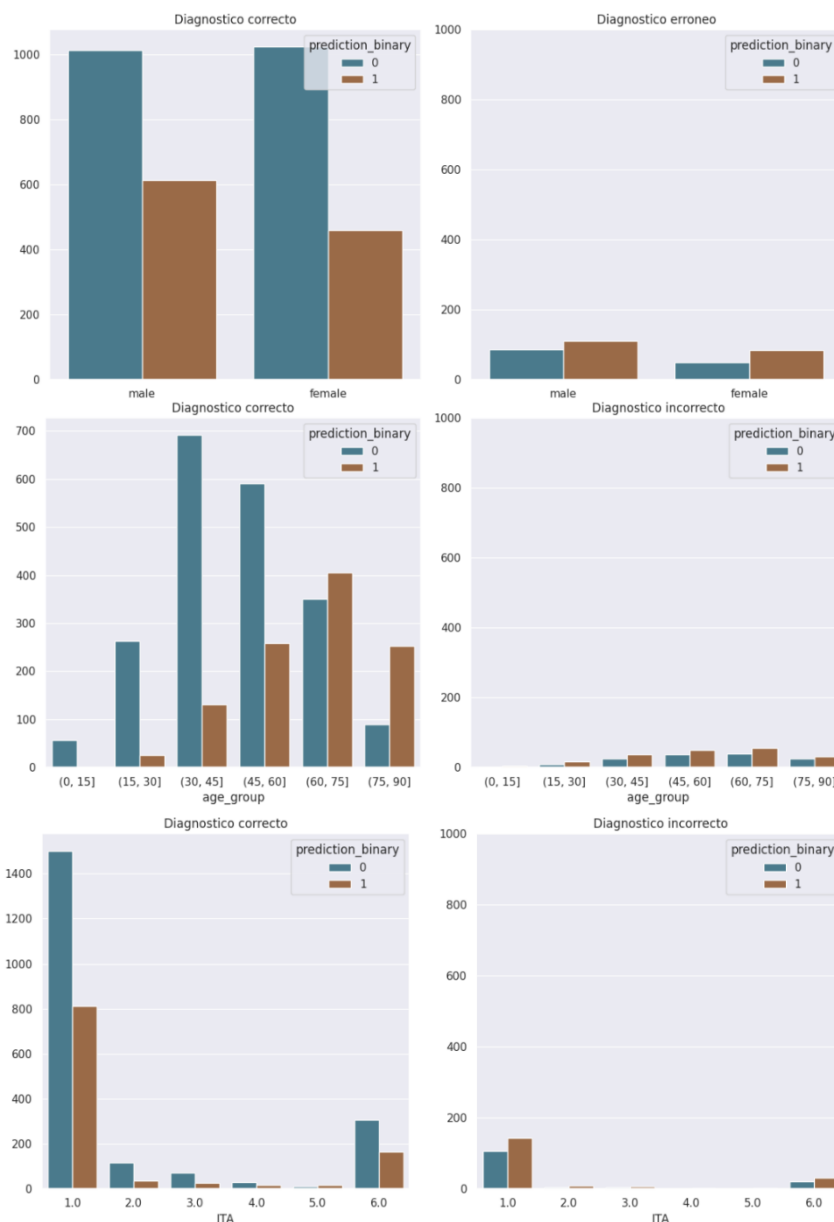


**Figura 28:** Distribución del diagnóstico real en falsos positivos.

**Fuente:** Propia.

Se puede apreciar claramente que la mayoría de los casos de falsos positivos corresponden a diagnósticos de nevus, queratosis benigna y queratosis actínica. Al contrastarlo con los datos que se vieron en el punto 4.2.3, podemos darnos cuenta de que la mayor parte de los diagnósticos benignos son nevus, seguidamente de queratosis benigna y queratosis actínica. Por lo tanto, podemos asumir la cantidad de errores del tipo nevus como razonables. Además, se puede intuir también que al modelo le cuesta más predecir la queratosis benigna y queratosis actínica, seguramente producido por la falta de casos de ejemplo.

A continuación, se estudiará la distribución de los diagnósticos correctos frente a los diagnósticos incorrectos respecto a algunas características como: **sex**, **agre\_group** y **ITA**, como se puede ver en la figura 29.



**Figura 29:** Distribución de los diagnósticos correctos frente a incorrectos en función de características.

**Fuente:** Propia.

Después de la visualización de las gráficas se puede concluir que no hay ninguna característica que aumente la cantidad de diagnósticos incorrectos. En todos los casos la distribución de diagnósticos incorrectos es igual y proporcional al total de casos de diagnósticos correctos.

## 4.2 Aplicación de diagnóstico

Finalmente, el último paso de este proyecto fue diseñar una aplicación capaz de cargar el modelo de diagnóstico y poder realizar predicciones de imágenes dermatoscópicas. Se desarrolló una aplicación en Python 3 apoyándose en Qt Designer para la creación de la interfaz.

La aplicación consta de tres secciones:

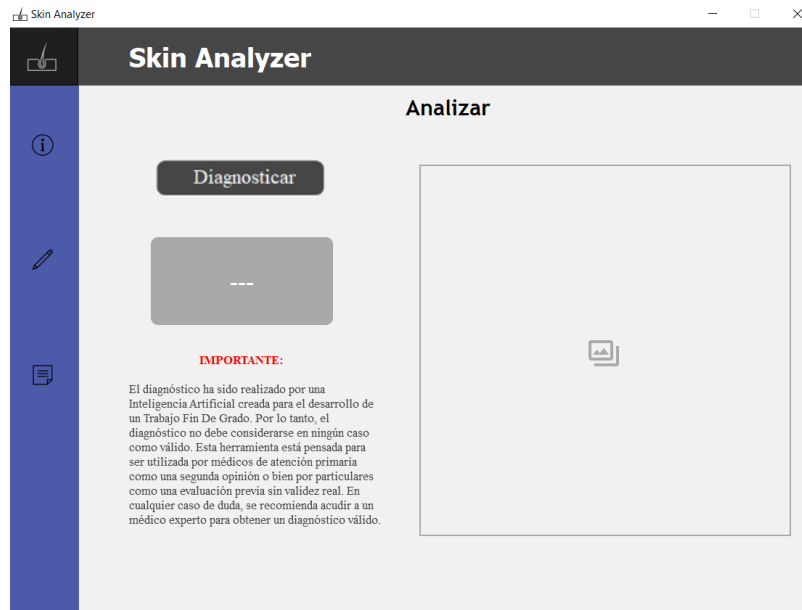
**Información:** Esta es la sección de inicio de la aplicación, se da un contexto general de la aplicación, su uso, los motivos por los que ha sido realizada y sus limitaciones, al mismo tiempo que se expone un archivo descargable en “ejemlo\_50\_imagenes” con cincuenta imágenes etiquetadas con el diagnóstico correcto y listas para probar en la aplicación en la sección de analizar. Esta sección se puede ver en la figura 30.



**Figura 30:** Sección Información aplicación.

**Fuente:** Propia.

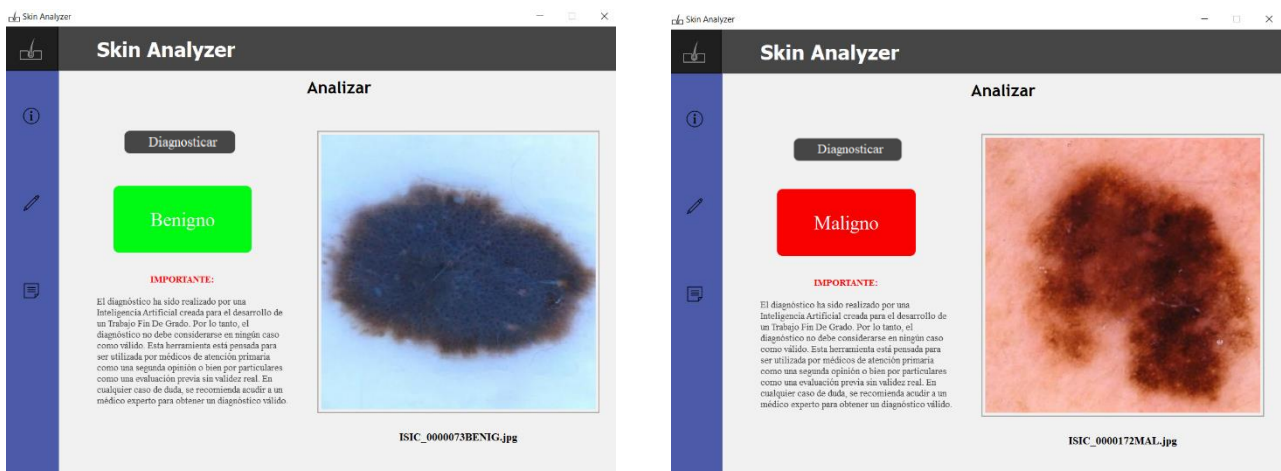
**Analizar:** En esta sección es donde la aplicación recibe imágenes del dispositivo local donde se esté ejecutando para analizar mediante el botón *Diagnosticar*, también se hace una importante aclaración sobre el uso de la aplicación. Inicialmente esta sección está vacía esperando una imagen a analizar, como se muestra en la figura 31.



**Figura 31:** Sección Analizar aplicación.

**Fuente:** Propia.

Cuando esta sección recibe una imagen mostrará su diagnóstico en el rectángulo gris debajo del botón *Diagnosticar*. En caso de ser un diagnóstico benigno cambiará el color de ese rectángulo a verde y mostrará el texto “Benigno”. En caso de ser maligno el rectángulo cambiará a un color rojo mostrando el texto “Maligno”. Al mismo tiempo en el cuadrado derecho se mostrará la imagen introducida a analizar, estas dos interacciones se pueden ver en la figura 32.

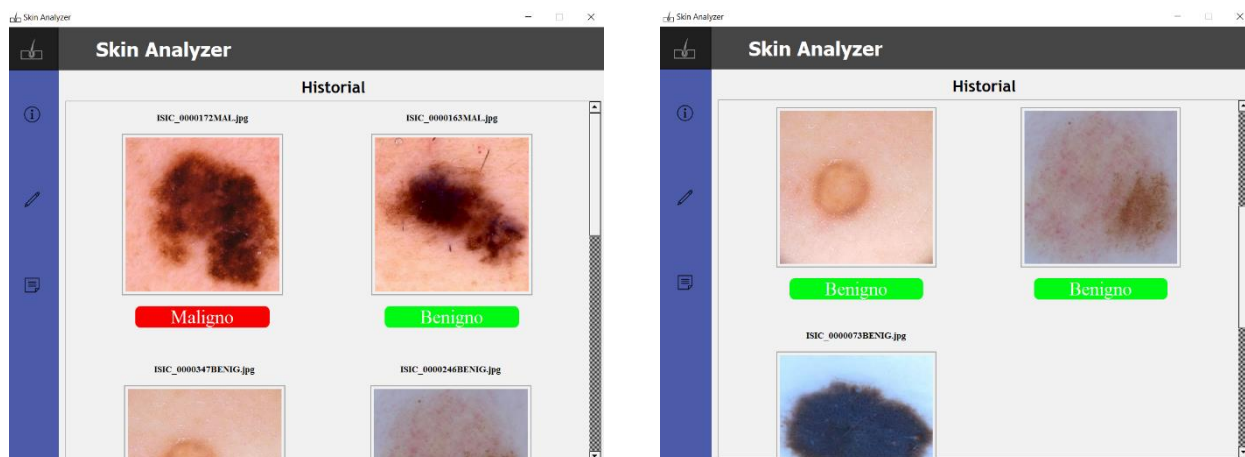


**Figura 32:** Sección Analizar funcionamiento.

**Fuente:** Propia.

**Historial:** En esta sección se muestran los resultados de las últimas ocho imágenes analizadas, su nombre y su diagnóstico. La última imagen analizada aparecerá arriba a la izquierda, si una imagen que ya ha sido analizada se encuentra dentro de esas ocho últimas se moverá al inicio de las imágenes analizadas, pero no saldrá repetida. El funcionamiento de esta sección se puede ver en la figura 33.





**Figura 33:** Sección Historial funcionamiento.

**Fuente:** Propia.

La aplicación se encuentra disponible en [umubox](https://umubox.com), el comprimido contiene la carpeta App que contiene dos carpetas, dist y build, para ejecutar la aplicación hay que ir a la carpeta dist, esta contiene la carpeta SkinAnalyzer donde se encuentra el ejecutable SkinAnalyzer.exe. Para ejecutarlo es tan sencillo como hacerle doble click o darle a “Abrir”.

## 5. Conclusiones y vías futuras

Una vez finalizado este proyecto, se pueden considerar los resultados obtenidos como satisfactorios, teniendo en cuenta los recursos disponibles comentados en el punto 1.5. El producto obtenido por este Trabajo Fin de Grado es una herramienta de diagnóstico automático de casos de riesgo de cáncer de piel basado en imágenes dermatoscópicas. Sus componentes y características principales son las siguientes:

Se ha creado un modelo de diagnóstico optimizado para imágenes dermatoscópicas de 512x512 píxeles con 3 canales de color RGB. Este modelo se basa en una red neuronal convolucional del tipo EfficientNet-B5, entrenada en Google Colab con más de 20.000 imágenes de la base de datos ISIC-2019. El rendimiento del modelo se evaluó utilizando el conjunto prueba, el cual contiene 3.500 imágenes de prueba que no fueron vistas por el modelo durante el entrenamiento. Las métricas de precisión y sensibilidad obtenidas superan 0,85.

Se ha creado una aplicación en Python 3 que emplea un modelo de diagnóstico para evaluar las imágenes cargadas en la sección de *Análisis*. Esta aplicación muestra un diagnóstico, indicando si la imagen es benigna o maligna. Además de la sección de *Análisis*, la aplicación cuenta con dos secciones adicionales: *Información* e *Historial*. En la sección de *Información*, se proporciona contexto sobre el propósito y el desarrollo de la aplicación, junto con un archivo de imágenes de ejemplo para el usuario. Por otro lado, en la sección de *Historial*, se muestra un registro de los diagnósticos más recientes, que incluye el nombre del archivo, su imagen y el diagnóstico obtenido.

El campo de la salud es uno de los que más está aprovechando la Inteligencia Artificial en la actualidad, y se espera que su uso aumente exponencialmente en el futuro, dada la capacidad de tratamiento y análisis de grandes volúmenes de datos. Este trabajo ha presentado un ejemplo de cómo la Inteligencia Artificial se aplica en el ámbito de la salud, creando un modelo de diagnóstico e importándolo a una aplicación funcional para detectar el riesgo de cáncer de piel para apoyar a los profesionales de la salud. Además de abordar la selección y preparación de datos, la elección de modelos mediante experimentos y su implementación en un sistema en producción, el trabajo también destaca métodos para evaluar el rendimiento.

Respecto a las vías futuras, sería muy interesante poder desarrollar un proyecto con las mismas intenciones a este “proyecto piloto” pero con una mayor cantidad de recursos, con la búsqueda de desarrollar varios modelos de predicción con mejores resultados. Con todo eso sería muy interesante desarrollar otra aplicación disponible para todo el público, especialmente para el ámbito médico y que pueda ser usada como una herramienta por profesionales de la salud no expertos en ese ámbito, para tener más de una opinión y decidir con más recursos el tratamiento o derivarlo a un experto.

Algunos de los aspectos que se podrían aplicar en futuras líneas de trabajo para mejorar el modelo de diagnóstico son:

**Utilización de metadatos:** Los metadatos expuestos anteriormente simplemente han sido usados con función informativa, para contextualizar, entender la información de mejor manera y documentar cómo han influido en los resultados, pero estos metadatos no han sido todos empleados en el desarrollo del modelo de predicción, solo se han empleado **image** y **diagnosis**.

**Especificidad dentro de los diagnósticos malignos:** El modelo de diagnóstico solo es capaz de realizar una clasificación binaria entre malignos y benignos, por lo que sería muy interesante reevaluar los resultados malignos por otro modelo de diagnóstico capaz de distinguir entre los tres tipos MEL, BCC y SCC, especialmente distinguir los MEL de los demás debido a su peligrosidad.

**Empleo de varios modelos:** En la aplicación desarrollada solamente se usa un modelo, por lo que sería interesante tener la posibilidad de emplear varios modelos y ver sus distintos diagnósticos con la finalidad de comprobar si coinciden en sus diagnósticos, de esa manera tener más de una predicción.

**Mejora de imágenes:** Muchas de las imágenes empleadas, aunque están centradas y recortadas gracias al empleo de TFRecords [28], contienen secciones negras producidas por el uso de dermatoscopios, por lo que genera algunos errores como al calcular el ITA. Por lo tanto, sería interesante mejorar las imágenes, al mismo tiempo que desarrollar un modelo capaz de conseguir resultados similares con imágenes no dermatoscópicas.

**Empleo de preprocesado:** Las imágenes empleadas contienen algunas imperfecciones como pelos que pueden llegar a confundir al modelo de diagnóstico. Por lo tanto, sería interesante aplicar algún filtro de preprocesado para eliminarlos, pero ese mismo filtro se debería aplicar también a las imágenes introducidas a la aplicación.

**Enriquecimiento progresivo de la base de datos:** Una idea muy interesante y ambiciosa sería crear una gran base de datos a nivel nacional pública, la cual iría recibiendo conjuntos de imágenes correctamente etiquetadas regularmente. Cada vez que se obtuviera un conjunto de imágenes significativamente grande, se tendría que crear y reentrenar de nuevo el modelo de diagnóstico.

**Empleo de una aplicación web:** Una idea algo más sencilla sería crear una aplicación web similar a la aplicación proporcionada y que pudiera estar activa siempre en un servidor web. De esta manera la herramienta sería mucho más accesible para todo el mundo y con cualquier dispositivo con conexión web.

En conclusión, este modelo de diagnóstico de casos de riesgo de cáncer de piel basado en imágenes dermatoscópicas puede ser una propuesta muy interesante en el futuro, ya que podría ser utilizado en el ámbito de la medicina si se realizasen las mejoras en la calidad del diagnóstico y se empleara de manera adecuada. De esta manera se conseguiría reducir la mortalidad de estas enfermedades y mejorar la calidad de vida de personas que las padezcan.

# Bibliografía

- [1] “Cáncer Datos y cifras,” Organización Mundial de la Salud. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/cancer>
- [2] “Epidemiología del cáncer de piel,” Roche Farma. [Online]. Available: <https://rochepacientes.es/cancer/piel/epidemiologia-melanoma.html>
- [3] “Cáncer de piel,” [3] IVO – Instituto Valenciano de Oncología. [Online]. Available: [https://www.ivo.es/actualidad/cancer-de-piel-melanoma/el-cancer-de-piel-representa-cerca-del-50-del-total-de-tumores-que-se-diagnostican-cada-ano-en-el-mundo/#:~:text=de%20piel%20\(melanoma\)-,El%20c%20c%20A%20nc%20er%20de%20piel%20representa%20cerca%20del%2050%25%20del%20total,cada%20a%20C%20B%20lo%20en%20el%20mundo](https://www.ivo.es/actualidad/cancer-de-piel-melanoma/el-cancer-de-piel-representa-cerca-del-50-del-total-de-tumores-que-se-diagnostican-cada-ano-en-el-mundo/#:~:text=de%20piel%20(melanoma)-,El%20c%20c%20A%20nc%20er%20de%20piel%20representa%20cerca%20del%2050%25%20del%20total,cada%20a%20C%20B%20lo%20en%20el%20mundo)
- [4] “Cáncer de piel,” Asociación española contra el cáncer. [Online]. Available: <https://www.contraelcancer.es/es/todo-sobre-cancer/tipos-cancer/cancer-piel/carcinoma-cutaneo>
- [5] J. García Dorado and P. Alonso Fraile, “ANATOMÍA Y FISIOLOGÍA DE LA PIEL.” [Online]. Available: [https://www.pediatriaintegral.es/wp-content/uploads/2021/xxv03/07/n3-156e1-13\\_RB\\_JesusGarcia.pdf](https://www.pediatriaintegral.es/wp-content/uploads/2021/xxv03/07/n3-156e1-13_RB_JesusGarcia.pdf)
- [6] Naeblys, “TYPE OF SKIN CANCER.” [Online]. Available: <https://www.istockphoto.com/br/foto/tumores-de-pele-verrugas-e-manchas-se%20C%20A%20o-3d-da-camada-da-pele-gm1210765902-350886594>
- [7] S. Esponsorizada Por Galderma, S. A. Localizador, P. Zaballo, C. Carrera, S. Puig, and J. Malvey, “Educación Médica Continuada Criterios dermatoscópicos para el diagnóstico del melanoma Dermoscopic Criteria in the Diagnosis of Melanoma,” 2004. [Online]. Available: <https://www.medigraphic.com/pdfs/cutanea/mc-2004/mc041b.pdf>
- [8] Palacios-Martínez, D., & Díaz-Alonso, R. A. (2017). Dermatoscopia para principiantes (i): características generales [Dermoscopy for beginners (i): General information]. *Semergen*, 43(3), 216–221. <https://doi.org/10.1016/j.semerg.2015.11.009>
- [9] “Dermatoscopio DermLite Handyscope,” DermLite. [Online]. Available: <https://www.dhmaterialmedico.com/material-medico/fotos/dermatoscopio-dermlite-handyscope-para-mviles-y-tabletas-1680512625.jpg>
- [10] “Dermatoscópio DermLite DL4 - 3GEN,” Descrição. [Online]. Available: <https://www.bleymed.com.br/media/product/407/dermatoscopio-dermlite-dl4-5c3.jpg>
- [11] Karimkhani, C., Green, A. C., Nijsten, T., Weinstock, M. A., Dellavalle, R. P., Naghavi, M., & Fitzmaurice, C. (2017). The global burden of melanoma: results from the Global Burden of Disease Study 2015. *The British journal of dermatology*, 177(1), 134–140. <https://doi.org/10.1111/bjd.15510>
- [12] Lopes, J., Rodrigues, C. M. P., Gaspar, M. M., & Reis, C. P. (2022). Melanoma Management: From Epidemiology to Treatment and Latest Advances. *Cancers*, 14(19), 4652. <https://doi.org/10.3390/cancers14194652>
- [13] A. Lashbrook, “AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind,” *The Atlantic*, 2018.
- [14] N. J. Nilsson, “INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK,” 1998.
- [15] T. M. (Tom M. Mitchell, *Machine Learning*. [Online]. Available: <https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20>

- [16] Ringa Tech, “Redes Neuronales Convolucionales - Clasificación avanzada de imágenes con IA / ML (CNN),” Youtube. [Online]. Available: [https://www.youtube.com/watch?v=4sWhhQwHqug&ab\\_channel=RingaTech](https://www.youtube.com/watch?v=4sWhhQwHqug&ab_channel=RingaTech)
- [17] Jesús Martínez, “Redes Neuronales Convolucionales en Profundidad,” Data smart. [Online]. Available: <https://datasmarts.net/es/redes-neuronales-convolucionales-en-profundidad/>
- [18] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” May 2019, [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [19] OpenGenus Tech Review Team, “EfficientNet [model architecture],” OpenGenus IQ. [Online]. Available: <https://iq.opengenus.org/efficientnet/>
- [20] D. Marina and A. del Pozoa, “Fototipos cutáneos. Conceptos generales,” *Offarm*, vol. 24, 2005.
- [21] P. J. Bevan and A. Atapour-Abarghouei, “Detecting Melanoma Fairly: Skin Tone Detection and Debiasing for Skin Lesion Classification,” Feb. 2022, [Online]. Available: <http://arxiv.org/abs/2202.02832>
- [22] M. R. Luo, “CIELAB,” in *Encyclopedia of Color Science and Technology*, Springer Berlin Heidelberg, 2015, pp. 1–7. doi: 10.1007/978-3-642-27851-8\_11-1.
- [23] Dina Hasarmi, “lable Digital Solution for Measurement of the Individual Typology Angle (ITA°) and Changes in the Skin Tone Articles • Dec 14, 2021 Scarletred®Vision: A Scalable Digital Solution for Measurement of the Individual Typology Angle (ITA°) and Changes in the Skin Tone,” Scarletred. [Online]. Available: <https://www.mendeley.com/reference-manager/library/all-references/>
- [24] N. Andrea Gil Coca, E. Hernando Hernández Rincón, and J. Contreras Ruíz, “The impact of primary and secondary prevention in reducing skin cancer O impacto da prevenção primária e secundária na redução do cancro de pele”, [Online]. Available: [https://revistas.ces.edu.co/index.php/ces\\_salud\\_publica/article/view/3559/2688](https://revistas.ces.edu.co/index.php/ces_salud_publica/article/view/3559/2688)
- [25] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.0575>
- [26] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [27] Larxel, “Skin Lesion Images for Melanoma Classification,” Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/andrewmvd/isic-2019/data>
- [28] Chris Deotte, “Triple Stratified KFold with TFRecords,” Kaggle. [Online]. Available: <https://www.kaggle.com/code/cdeotte/triple-stratified-kfold-with-tfrecords/input>
- [29] Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., & Yap, M. H. (2022). Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Medical image analysis*, 75, 102305. <https://doi.org/10.1016/j.media.2021.102305> [30] “Google Colab ,” Google. [Online]. Available: <https://colab.google/>
- [31] “Google Colab,” Google . [Online]. Available: <https://research.google.com/colaboratory/intl/es/faq.html>
- [32] “The International Skin Imaging Collaboration,” ISIC. [Online]. Available: <https://www.isic-archive.com/>
- [33] mssaperla, docs-preview-bot, irinaskaya, and mariesalet, “Preparación de datos para el aprendizaje distribuido,” Microsoft Build. [Online]. Available:

<https://learn.microsoft.com/es-es/azure/databricks/machine-learning/load-data/ddl-data>

- [34] Dimitre Oliveira, “Creating TFRecords,” Keras. [Online]. Available: [https://keras.io/examples/keras\\_recipes/creating\\_tfrecords/](https://keras.io/examples/keras_recipes/creating_tfrecords/)

# Lista de Acrónimos y Abreviaturas

<b>TFG</b>	Trabajo Final de Grado.
<b>ML</b>	Maching learning    Aprendizaje computacional
<b>UI</b>	User Interface    Interfaz de Usuario
<b>OMS</b>	Organización Mundial de la Salud.
<b>IRAC</b>	International Agency for Research on Cancer    Agencia Internacional de Investigación del Cáncer.
<b>ISIC</b>	International Skin Imaging Collaboration    Colaboración Internacional de Imágenes de la Piel.
<b>LED</b>	Light Emitting Diode    Diodo Emisor de Luz.
<b>PRO</b>	Professional    Profesional.
<b>IA</b>	Inteligencia Artificial.
<b>RNA</b>	Red Neuronal Artificial.
<b>CNN</b>	Convolutional Neural Network    Redes Neuronales convolucionales.
<b>RLU</b>	Rectified Linear Unit    Unidad Lineal Rectificada
<b>TP</b>	True Positive    Verdadero Positivo.
<b>TN</b>	True Negative    Verdadero Negativo.
<b>FP</b>	Fake Positive    Falso Positivo.
<b>FN</b>	Fake Negative    Falso Negativo.
<b>ITA</b>	Individual Typology Angle    Ángulo de Tipología Individual.
<b>RGB</b>	Red Green Blue    Rojo Verde Azul.
<b>TPU</b>	Tensor Processing Unit    Unidad de Procesamiento Tensorial.
<b>NV</b>	Nevus    Lunar.
<b>MEL</b>	Melanoma    Melanoma.
<b>BKL</b>	Benign keratosis    Queratosis benigna.
<b>DF</b>	Dermatofibroma    Dermatofibroma.
<b>SCC</b>	Squamous cell carcinoma    Carcinoma de células escamosas.
<b>BCC</b>	Basal cell carcinoma    Carcinoma de células basales.
<b>VASC</b>	Vascular lesion    Lesión vascular.
<b>AK</b>	Actinic keratosis    Queratosis actínica.
<b>LR</b>	Learning Rate    Tasa de aprendizaje.
<b>ASIC</b>	Application Specific Integrated Circuit    Circuito Integrado para Aplicaciones Específicas.
<b>GCS</b>	Google Cloud Storage    Almacenamiento en la Nube de Google.