



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

DESARROLLO Y DISEÑO DE UN ENTORNO DE TRABAJO PARA LA DETECCIÓN INTELIGENTE DE LESIONES EN LA EPIDERMIS MEDIANTE ANÁLISIS Y CLASIFICACIÓN DE IMÁGENES

Autor: Ramon Villagrasa Benages

Directora: Cristina Puente Águeda

Madrid, Julio 2022

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

**DESARROLLO Y DISEÑO DE UN ENTORNO DE TRABAJO PARA LA DETECCIÓN
INTELIGENTE DE LESIONES EN LA EPIDERMIS MEDIANTE ANÁLISIS Y
CLASIFICACIÓN DE IMÁGENES**

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2021/2022 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.



Fdo.: Ramon Villagrasa Benages

Fecha: 04/07/2022

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Cristina Puente Águeda

Fecha: 04/07/2022



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

DESARROLLO Y DISEÑO DE UN ENTORNO DE TRABAJO PARA LA DETECCIÓN INTELIGENTE DE LESIONES EN LA EPIDERMIS MEDIANTE ANÁLISIS Y CLASIFICACIÓN DE IMÁGENES

Autor: Ramon Villagrasa Benages

Directora: Cristina Puente Águeda

Madrid, Julio 2022

Agradecimientos - REPASAR

Quiero dedicar este trabajo a todas esas personas que me han ayudado a lo largo de esta tarea tan significativa para mí, pues de una forma u otra su aportación ha sido relevante.

Agradecer sin duda los valores y enseñanzas que mis padres me han transmitido, especialmente en esta última etapa del grado en la que se han presentado momentos difíciles, pero no insuperables y en los que la perseverancia que mi abuela me enseñó ha dado sus frutos.

Por último, a mi directora, por guiarme y ayudarme durante todo el desarrollo del proyecto para alcanzar la mejor calidad posible.

DESARROLLO Y DISEÑO DE UN ENTORNO DE TRABAJO PARA LA DETECCIÓN INTELIGENTE DE LESIONES EN LA EPIDERMIS MEDIANTE ANÁLISIS Y CLASIFICACIÓN DE IMÁGENES

Autor: Villagrasa Benages, Ramon.

Directora: Puentes Águeda, Cristina.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Desarrollo de una aplicación no funcional para el análisis de imágenes dermatoscópicas de la piel, clasificando sus afecciones más comunes en siete clases distintas. El algoritmo implementado, basado en técnicas de Aprendizaje Automatizado y haciendo uso de la librería *Keras*, permite obtener un primer diagnóstico para cada persona, según la imagen aportada.

Palabras clave: Aprendizaje Automático, Aprendizaje Profundo, Red Neuronal Residual (ResNet), Red Neuronal Convolutiva (CNN).

1. Introducción

El auge de las tecnologías relacionadas con el Aprendizaje Automático (*Machine Learning* en inglés) destaca por la gran abundancia de datos y la capacidad computacional y de procesamiento de estos. De entre las diferentes técnicas que el campo presenta, el Aprendizaje Profundo (*Deep Learning* en inglés) se impone como una de las herramientas más frecuentes, requiriendo una gran cantidad de datos para alcanzar resultados satisfactorios.

Por otra parte, el órgano más extenso del ser humano, la piel, es una capa de tejido resistente y flexible que cubre y protege el cuerpo. Al tratarse de la primera zona de contacto con el exterior hace que sea una potencial fuente de lesiones, mayoritariamente debido a la sobreexposición solar, pues durante años se ha estado potenciando el atractivo de lucir una piel morena. En la gran mayoría de los casos, este hábito es difícil que tome en cuenta las medidas de protección adecuadas, pudiendo desencadenar consecuencias fatales e irreversibles. Es por ello por lo que se conoce como melanoma el tipo de cáncer de piel más frecuente, estando su aparición directamente relacionada con las exposiciones agudas e intensas a la radiación solar (Hameed, M.Shabut, K.Ghosh, & Hossain, 18 September 2019). A pesar de ello, es habitual tener manchas en la piel desde el momento en que nacemos, pudiéndose también desarrollar con el paso del tiempo. Algunas de las más peligrosas son:

- ✚ Carcinoma de células escamosas: su crecimiento es rápido y aparece en zonas normalmente expuestas al sol, bajo diferentes maneras (bulto o herida con costra).
- ✚ Melanoma maligno: variante del cáncer de piel cuya aparición tiene lugar sobre un lunar. Su forma es asimétrica y destaca por su color rojizo o blanco, con bordes irregulares cuya forma o tamaño puede alterarse.

- ✚ **Carcinoma basocelular:** muy común y de crecimiento lento. Destaca por su aparición en forma de herida incurable de bordes rosados o poco definidos (como una cicatriz).

Actualmente, se sabe que las enfermedades que afectan directamente a la piel son la cuarta causa más frecuente de todas las enfermedades humanas, y afectan a casi un tercio de la población mundial, en torno a 1.900 millones de personas a la vez, de ahí que se genere un interés para su investigación en numerosas disciplinas (Sanofi, 2020).

2. Definición del proyecto

El objetivo de este proyecto de fin de grado es el desarrollo y diseño de un entorno de trabajo que permita detectar de forma inteligente las lesiones más comunes en la dermis gracias al análisis y clasificación de imágenes dermatoscópicas. El enfoque empleado se basa en el Aprendizaje Automático y Profundo, implementando las librería de *Keras*¹. Para ello, como se detallará más adelante, se ha seguido una metodología ágil estableciendo un *Producto Mínimo Viable* (MVP de sus siglas en inglés) sobre el que se ha ido iterando para conseguir una mayor optimización en función de los inconvenientes sobrevenidos durante el ciclo del trabajo.

El desarrollo del MVP consiste en un clasificador de las lesiones más comunes en la piel, basado en el análisis de imágenes dermatoscópicas. En este caso, se distingue entre siete afecciones: Queratosis Actínicas (akiec), Lesiones vasculares (vas), Dermatofibroma (df), Nevos Melanocíticos (nv), Melanoma (mel), Lesiones benignas similares a Queratosis (bkl), Carcinoma de células basales (ccb).

3. Descripción del modelo/sistema/herramienta

Para realizar el proyecto se han contemplado dos enfoques diferentes para la clasificación de imágenes durante el entrenamiento de los algoritmos:

- 1. Enfoque basado en CNN:** clasificación de imágenes de forma independiente empleando un modelo secuencial basado en una red neuronal Convolutiva (Figura 1 - Arquitectura CNN).

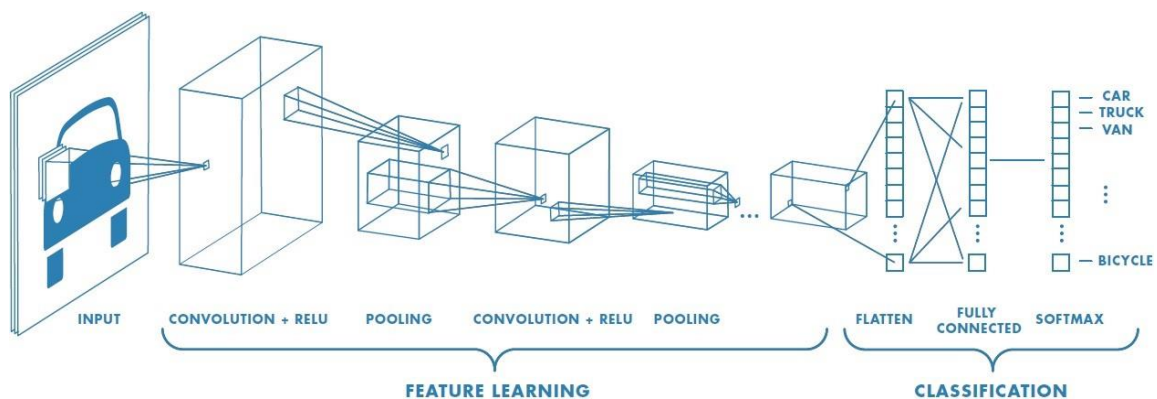


Figura 1 - Arquitectura CNN (Prabhu, 2018)

¹ Biblioteca de Redes Neuronales Artificiales para el entrenamiento de modelos de *Deep Learning*. Su diseño permite la construcción de una red neuronal mediante una arquitectura por bloques.

2. **Enfoque ResNet:** uso de una Red Neuronal Residual (*Figura 2 - Arquitectura ResNet*), previamente entrenada, en la que se ha modificado el último *layer* con las características propias del proyecto (siete niveles correspondientes a las lesiones).

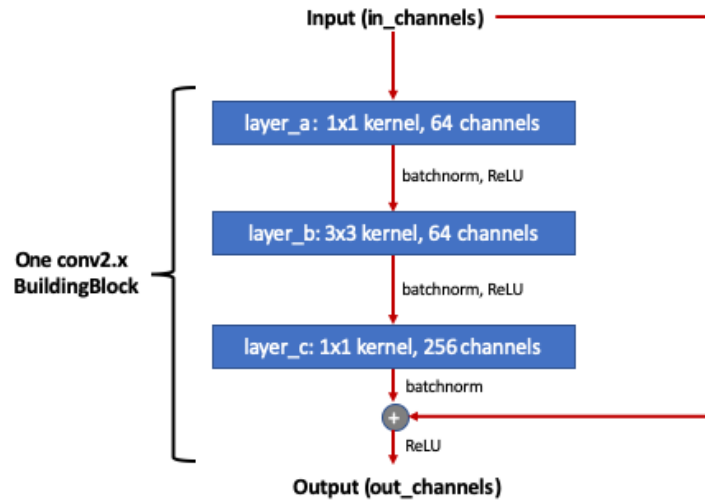


Figura 2 - Arquitectura ResNet (Steve, 2020)

En ambos casos, los modelos han sido alimentados con imágenes pertenecientes al conjunto de datos público *Human Against Machine with 10000 training images*. Este conjunto está compuesto de un total de 10015 imágenes dermatoscópicas pertenecientes a distintos individuos, adquiridas y almacenadas según su modalidad (Tschandl, Rosendahl, & Kittler, 2018).

4. Resultados

Tras la aplicación del modelo secuencial el porcentaje de acierto fijando el número total de imágenes en cada clase (primer enfoque) ha sido de 81,79%, mientras que, si se equilibran las clases empleando la totalidad de las imágenes, el porcentaje desciende ligeramente hasta el 73,12% (*Tabla 1 - Comparación ResNet y modelo Secuencial según el enfoque usado*).

Por una parte, se ha observado que el uso de una red neuronal previamente entrenada como ResNet permite mejorar significativamente el acierto, superando el umbral del 80% (*Tabla 1 - Comparación ResNet y modelo Secuencial según el enfoque usado*).

<i>Modelo</i>	<i>Precisión</i>
CNN (primer enfoque)	81,79%
CNN (segundo enfoque)	73,12%
ResNet50	80,2%

Tabla 1 - Comparación ResNet y modelo Secuencial según el enfoque usado

5. Conclusión

El objetivo del proyecto era el desarrollo y diseño de un entorno de trabajo para la detección inteligente de lesiones en la piel. Se ha comparado la precisión obtenida con un enfoque puramente secuencial de una Red Neuronal Convolutiva (CNN) y una Red Neuronal Residual (ResNet). Por un lado, se ha hecho uso de todas las imágenes disponibles mientras que el otro enfoque ha consistido en fijar el número total de imágenes con las que se ha trabajado, pues no siempre es posible obtener grandes cantidades de datos.

Como era de esperar, la implementación de la Red Neuronal Residual (previamente entrenada) ha supuesto una mejora en la precisión del modelo, visto que se parte de una estructura más robusta (sólo se modifica un nivel de la red) que en el caso del modelo de la Red Neuronal Convolutiva. Sin embargo, es bienvenida la introducción de nuevas imágenes en el conjunto de entrenamiento que sigan la misma estructura del conjunto de datos empleado, con el fin de robustecer todavía más la precisión de las arquitecturas desarrolladas.

6. Referencias

American Cancer Society. (14 de Agosto de 2019). Etapas del cáncer de piel tipo melanoma. *American Cancer Society*, 1.

Aptoide. (s.f.). Obtenido de <https://moleexplore.es.aptoide.com/app>

Arellano, A. R. (23 de Mayo de 2018). *Academia Española de Dermatología y Venerología*. Obtenido de <https://aedv.es/ederma-una-aplicacion-seguimiento-las-lesiones-piel-sospechosas/>

Boesch, G. (2022). *Viso.AI*. Obtenido de Deep Residual Networks (ResNet, ResNet50) – Guide in 2022: <https://viso.ai/deep-learning/resnet-residual-neural-network/>

Calzado, A., & Geleijns, J. (24 de Diciembre de 2010). Tomografía computarizada. Evolución, principios técnicos y aplicaciones. *Didac*, 18.

DiarioFarma. (23 de Mayo de 2018). Obtenido de <https://www.diariofarma.com/2018/05/23/una-aplicacion-util-para-la-autoexploracion-de-piel-contra-el-melanoma>

Flores, D. W. (s.f.). *Cinvestav*. Obtenido de Perceptrón multicapa: https://www.tamps.cinvestav.mx/~wgomez/notas/RP/Notas_Clase12.pdf

- G.Cid. (11 de Febrero de 2022). *El Confidencial*. Obtenido de https://www.elconfidencial.com/tecnologia/ciencia/2020-02-11/apps-cancer-piel-aplicaciones-peligro-inseguras_2448988/
- Géron, A. (2017). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly.
- Gluon. (2021). *Gluon*. Obtenido de Prepare the ImageNet dataset: https://cv.gluon.ai/build/examples_datasets/imagenet.html
- Google. (s.f.). *Colaboratory: Frequently Asked Questions*. Obtenido de Google: <https://research.google.com/colaboratory/faq.html>
- Hameed, N., M.Shabut, A., K.Ghosh, M., & Hossain, M. (18 September 2019). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. 1-18.
- He, K., Zhang, X., Ren, S., & Sun, J. (10 de Diciembre de 2015). *Cornell University*. Obtenido de Deep Residual Learning for Image Recognition: <https://arxiv.org/abs/1512.03385>

DEVELOPMENT AND DESIGN OF A WORKING ENVIRONMENT FOR THE INTELLIGENT DETECTION OF LESIONS IN THE EPIDERMIS THROUGH ANALYSIS AND CLASSIFICATION OF IMAGES

Author: Villagrasa Benages, Ramon.

Supervisor: Puente Águeda, Cristina.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

Creating a working environment for skin imaging analysis of skin copies, classifying their most common conditions. The algorithm implemented, based on automated learning techniques, and making use of the *Keras* library, allows obtaining a first diagnosis for each person, according to the image provided.

Keywords: Machine Learning, Deep Learning, Neural Network, ResNet, CNN.

1. Introduction

The rise of technologies related to Machine Learning stands out because of the great abundance of data and the computational and processing capacity of these. Among the different techniques that the field presents, Deep Learning is imposed as one of the most frequent tools, requiring a great amount of data to achieve satisfactory results.

Moreover, the most extensive organ of the human being, the skin, is a layer of resistant and flexible tissue that covers and protects the body. Being the first contact area with the exterior makes it a potential source of injury, mainly due to the solar overexposure, since for years it has been potentiating the attraction of sporting a brown skin. In many cases, this habit is difficult to consider adequate protective measures, with the result that it can lead to fatal and irreversible consequences. Therefore, it is known as melanoma the most frequent type of skin cancer, being its appearance directly related to acute and intense exposures to solar radiation (Hameed, M.Shabut, K.Ghosh, & Hossain, 18 September 2019). Despite this, it is common to have spots on the skin from the moment we are born, and it is also possible to develop with the passage of time. Some of the most dangerous are:

- ✚ Squamous cell carcinoma: its growth is rapid and appears in areas normally exposed to the sun, in different ways (lump or scab wound).
- ✚ Malignant melanoma: A variant of skin cancer whose appearance occurs on a mole. Its shape is asymmetrical and stands out for its reddish or white colour, with irregular edges whose shape or size can be altered.
- ✚ Basal cell carcinoma: very common and slow growth. It stands out for its appearance in the form of an incurable wound of pink or undefined edges (such as a scar).

Currently, diseases that directly affect the skin are known to be the fourth most common cause of all human diseases, and they affect nearly one-third of the world's population, around 1.9 billion people at a time, hence generating an interest for their research in numerous disciplines (Sanofi, 2020).

2. Project definition

The aim of this project is the development and design of a working environment that can intelligently detect the most common injuries in the dermis thanks to the analysis and classification of dermatoscopic images. The approach employed is based on machine learning and deep implementing the Keras Bookstore. To this end, as detailed below, an agile methodology has been followed by establishing a Minimum Viable Product (MVP) that has been iterating to achieve its highest optimization according to the disadvantages that have occurred during the work cycle.

The development of the MVP consists of a classifier of the most common lesions in the skin, based on the analysis of dermatoscopic images. In this case, we distinguish between seven conditions: Actinic Keratosis (akiec), Vascular lesions (vas), Dermatofibroma (df), Melanocytic Nevos (nv), Melanoma (mel), Benign lesions like Keratosis (bkl), Basal cell carcinoma (bcc).

3. Model description/system/tool

To carry out the project, two different approaches have been contemplated for the classification of images during the training of the algorithms:

1. **CNN-based approach:** independently imaging using a sequential model based on a convolutional neural network (*Figura 1 - Arquitectura CNN Architecture*).
2. **ResNet approach:** use of a residual network (*Figura 2 - Arquitectura ResNet*), previously trained, in which the last layer has been modified with the characteristics of the project (seven levels corresponding to the injuries).

In both cases, the models have been fed images belonging to the public data set *Human Against Machine with 10000 training images*. This set consists of a total of 10015 dermatoscopic images belonging to different individuals, acquired, and stored according to their modality (Tschandl, Rosendahl, & Kittler, 2018).

4. Results

After the application of the sequential model, the percentage of success by setting the total number of images in each class has been 81,79%, whereas if the classes are balanced using the totality of the images, the percentage decreases slightly to 73.12% (Table 1- Comparison ResNet and Sequential model according to the used approach).

On the one hand, it has been observed that the use of a previously trained neural network such as ResNet makes it possible to significantly improve the accuracy, exceeding in both cases the threshold of 80% (*Table 1-ResNet Comparison and Sequential model according to the used approach*).

5. Conclusion

The aim of the project was the development and design of a working environment for the intelligent detection of skin lesions. The accuracy obtained with a Convolutional Neural Network (CNN) approach, and a Residual Neural Network (ResNet) has been compared, making use in both cases of two distinct scenarios. On the one hand, all available images have been used while the other approach has been to fix the total number of images that have been worked with, since it is not always possible to obtain large amounts of data.

As expected, the implementation of a previously trained Neural Network has resulted in an improvement in the accuracy of the model, seen that it is part of a more robust structure (only one layer of the network is modified) than in the case of the sequential model. However, it is welcomed to introduce new images in the training set that follow the same structure of the data set used to further strengthen the accuracy of the developed architectures.

6. References

American Cancer Society. (14 de Agosto de 2019). Etapas del cáncer de piel tipo melanoma.

American Cancer Society, 1.

Aptoide. (s.f.). Obtenido de <https://moleexplore.es.aptoide.com/app>

Arellano, A. R. (23 de Mayo de 2018). *Academia Española de Dermatología y Venerología*.

Obtenido de <https://aedv.es/ederma-una-aplicacion-seguimiento-las-lesiones-piel-sospechosas/>

Boesch, G. (2022). *Viso.AI*. Obtenido de Deep Residual Networks (ResNet, ResNet50) –

Guide in 2022: <https://viso.ai/deep-learning/resnet-residual-neural-network/>

Calzado, A., & Geleijns, J. (24 de Diciembre de 2010). Tomografía computarizada.

Evolución, principios técnicos y aplicaciones. *Didac*, 18.

DiarioFarma. (23 de Mayo de 2018). Obtenido de

<https://www.diariofarma.com/2018/05/23/una-aplicacion-util-para-la-autoexploracion-de-piel-contra-el-melanoma>

Flores, D. W. (s.f.). *Cinvestav*. Obtenido de Perceptrón multicapa:

https://www.tamps.cinvestav.mx/~wgomez/notas/RP/Notas_Clase12.pdf

- G.Cid. (11 de Febrero de 2022). *El Confidencial*. Obtenido de https://www.elconfidencial.com/tecnologia/ciencia/2020-02-11/apps-cancer-piel-aplicaciones-peligro-inseguras_2448988/
- Géron, A. (2017). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly.
- Gluon. (2021). *Gluon*. Obtenido de Prepare the ImageNet dataset: https://cv.gluon.ai/build/examples_datasets/imagenet.html
- Google. (s.f.). *Colaboratory: Frequently Asked Questions*. Obtenido de Google: <https://research.google.com/colaboratory/faq.html>
- Hameed, N., M.Shabut, A., K.Ghosh, M., & Hossain, M. (18 September 2019). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. 1-18.
- He, K., Zhang, X., Ren, S., & Sun, J. (10 de Diciembre de 2015). *Cornell University*. Obtenido de Deep Residual Learning for Image Recognition: <https://arxiv.org/abs/1512.03385>



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

DESARROLLO Y DISEÑO DE UN ENTORNO DE TRABAJO PARA LA DETECCIÓN INTELIGENTE DE LESIONES EN LA EPIDERMIS MEDIANTE ANÁLISIS Y CLASIFICACIÓN DE IMÁGENES

Autor: Ramon Villagrasa Benages

Directora: Cristina Puente Águeda

Madrid, Julio 2022

Índice de la memoria

<i>Índice de figuras</i>	III
<i>Índice de tablas</i>	VII
Capítulo 1. Introducción	1
1.1 Motivación del proyecto.....	3
1.2 Definición del proyecto	6
1.3 Objetivos	7
Capítulo 2. Descripción de las Tecnologías	8
2.1 Keras.....	8
2.2 Google Colaboratory	8
2.3 HAM10000 Dataset.....	9
Capítulo 3. Estado de la Cuestión	12
3.1 SkinVision.....	12
3.2 Molexplore	14
3.3 eDerma	15
3.4 Deep Learning	16
3.4.1 Inspiración biológica.....	16
3.4.2 Perceptrón.....	18
3.4.3 Redes Neuronales Profundas.....	19
3.4.4 Sobre aprendizaje.....	22
3.4.5 Hiper parámetros	25
3.5 Redes Neuronales Convolucionales	26
3.5.1 Arquitectura.....	27
3.6 Redes Neuronales previamente entrenadas	28
3.6.1 ResNet50.....	31
3.6.2 ImageNet	33
3.7 Conclusión.....	34
Capítulo 4. Definición del Trabajo	36
4.1 Justificación.....	36

4.2 Metodología.....	36
4.3 Planificación y Estimación de Costes.....	38
Capítulo 5. Sistema Desarrollado	42
5.1 Análisis del Dataset y Enfoques Planteados.....	42
5.2 Diseño de las Redes Neuronales.....	46
5.2.1 Modelo basado en la Red Neuronal Convolucional.....	46
5.2.2 Modelo basado en la Red Neuronal Residual	53
5.3 Diseño del Interfaz Gráfico	58
Capítulo 6. Análisis de Resultados.....	60
Capítulo 7. Conclusiones y Trabajos Futuros.....	64
7.1 Modelos de Aprendizaje Profundo Desarrollados.....	64
7.2 Mejoras Generales	65
7.3 Consecución de los Objetivos	66
Capítulo 8. Bibliografía.....	68
ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS	73
ANEXO II	74

Índice de figuras

Figura 1 - Arquitectura CNN (Prabhu, 2018).....	II
Figura 2 - Arquitectura ResNet (Steve, 2020).....	III
Figura 3 - Etapas del melanoma maligno (American Cancer Society, 2019)	3
Figura 4 - Ejemplo de escala de colores y TAC (Calzado & Geleijns, 2010).....	5
Figura 5 - Flujo esquemático de los métodos de elaboración del conjunto final de datos (Tschandl, Rosendahl, & Kittler, 2018)	11
Figura 6 - Interfaz gráfico y funcionamiento de SkinVision (MHealth Belgium, s.f.)	13
Figura 7 - Interfaz gráfico de la aplicación Molexplore (Aptoide, s.f.)	14
Figura 8 - Interfaz gráfico de eDerma (Arellano, 2018).....	15
Figura 9 - Diagrama del cuerpo celular de una neurona (I.E.S. Suel, s.f.).....	17
Figura 10 - Cálculo de diferentes funciones lógicas a partir del modelado de una neurona (Piccinini, 2004)	17
Figura 11 - Representación de la Unidad Lógica de Umbral (Sah, 2018)	18
Figura 12 - Funciones lógicas más comunes	19
Figura 13 - Arquitectura Perceptrón Multicapa (Flores).....	20
Figura 14 - Representación del algoritmo de Descenso de Gradiente (Keshav, s.f.)	20
Figura 15 - Funciones de activación (Thinkage, 2017).....	21
Figura 16 - Comparación entre modelos con falta de aprendizaje, adecuado y con sobre aprendizaje (Rahaman, 2020).....	24
Figura 17 - Etapas en una Red Neuronal Convolutiva	26
Figura 18 - Aplicación de <i>max pooling</i> y <i>average pooling</i> (Rahaman, 2020)	27
Figura 19 - Arquitectura de una Red Neuronal Convolutiva (Tensorflow, 2017).....	28
Figura 20 - Transferencia de Aprendizaje (ViewNext, 2020).....	29
Figura 21 - Criterios para la aplicación de transferencia de aprendizaje (Rahaman, 2020)	30
Figura 22 - Arquitectura de ResNet50 (Sachan, 2021)	32
Figura 23 - Conjunto de datos de ImageNet (Gluon, 2021)	33

Figura 24 - Distribución de las imágenes en el conjunto de datos <i>HAM10000</i>	43
Figura 25 -Distribución de los métodos de validación del conjunto de datos <i>HAM10000</i> . 45	
Figura 26 - Resumen del modelo de la CNN	49
Figura 27 - Resultados del modelo basado en CNN con 30 épocas	50
Figura 28 - Resultados del modelo basado en CNN con 100 épocas	51
Figura 29 - Resultados del modelo basado en CNN con 100 épocas y todas las imágenes disponibles en el conjunto de datos	52
Figura 30 - Resumen del modelo de ResNet50	55
Figura 31 - Árbol de decisión en el entrenamiento de una red neuronal.....	57
Figura 32 - Propuesta del interfaz gráfico para la aplicación de clasificación de lesiones cutáneas <i>Skin App</i>	59
Figura 33 - Error cometido por el modelo CNN desarrollado en el primer enfoque	61
Figura 34 - Error cometido por el modelo CNN desarrollado en el segundo enfoque.....	62
Figura 35 - Error cometido por el modelo ResNet desarrollado	63
Figura 36 - Visión general de las lesiones en el conjunto de datos <i>HAM10000</i>	66

Índice de tablas

Tabla 1 - Comparación ResNet y modelo Secuencial según el enfoque usado.....	III
Tabla 2 - Diagrama de GANT del cronograma del proyecto	38
Tabla 3 - Especificaciones técnicas del ordenador portátil	40
Tabla 4 - Horas dedicadas al desarrollo del proyecto.....	40
Tabla 5 - Recursos humanos empleados en el proyecto	41
Tabla 6 - Resultados obtenidos con los modelos desarrollados	65

Capítulo 1. INTRODUCCIÓN

Conocido como el órgano más extenso del ser humano, la piel es una capa de tejido resistente y flexible que cubre y protege el cuerpo. Su principal función permite al individuo conectar y adaptarse con el medio ambiente, siendo además el órgano más pesado, pudiendo llegar a alcanzar hasta los 4 kg de masa total. Difiere de una región a otra: tiene zonas más gruesas (plantas de los pies y palmas de las manos) y zonas más finas (que constituyen los párpados, los pliegues o superficies de flexión y extensión).

Al tratarse del órgano con mayor dimensión del cuerpo, puede verse comprometido a causa de distintos factores como la exposición a la radiación ultravioleta, el bronceado voluntario, el tabaco, el consumo de alcohol, las actividades físicas (especialmente al aire libre), los virus y el entorno laboral de cada persona (Hameed, M.Shabut, K.Ghosh, & Hossain, 18 September 2019). En definitiva, estos factores dependen mucho del estilo de vida de cada persona, pero son los principales contribuyentes para comprometer su integridad y tienen un impacto significativo y devastador en su bienestar. A pesar de ello, es habitual tener manchas en ésta desde el momento en que nacemos, pudiéndose también desarrollar con el paso del tiempo. Sin embargo, debemos ser conscientes de que algunas de estas manchas pueden ser peligrosas y es por eso por lo que deben ser vigiladas con frecuencia. Entre las más dañinas se encuentran:

- ✚ Carcinoma de células escamosas: su crecimiento es rápido y aparece en zonas normalmente expuestas al sol, bajo diferentes maneras (bulto o herida con costra).
- ✚ Melanoma maligno: variante del cáncer de piel cuya aparición tiene lugar sobre un lunar. Su forma es asimétrica y destaca por su color rojizo o blanco, con bordes irregulares cuya forma o tamaño puede alterarse.

- ✚ Carcinoma basocelular: muy común y de crecimiento lento. Destaca por su aparición en forma de herida que no se cura y con bordes rosados o poco definidos (como una cicatriz).

En ciertos casos, es evidente que la evolución de estas manchas puede desencadenar en peligrosas afecciones si su seguimiento no es el adecuado. Actualmente, se sabe que las enfermedades que afectan directamente a la piel son la cuarta causa más frecuente de todas las enfermedades humanas, involucrando a casi un tercio de la población mundial, en torno a 1.900 millones de personas a la vez, de ahí que se genere un interés para su investigación en numerosas disciplinas. En particular, en nuestro país se estima que fallecen cada año cerca de 700 personas por *melanoma* (el tipo de cáncer de piel más común entre la población) y en estas últimas décadas, ha habido un repunte de hasta un 38% en estos tumores. Según el informe que emitió la Sociedad Española de Oncología Médica en 2020, el número de casos de melanoma en España fue de 6.179 (Sanofi, 2020). Este incremento se debe a un cambio en nuestros hábitos respecto a la sobreexposición al sol, sin tomar en cuenta las debidas medidas de protección. Los cánones de belleza han influido en este aspecto, ya que durante años se ha venido potenciando como “bello” lucir una piel morena. Es por ello por lo que, el melanoma, es el tipo de cáncer de piel más frecuente en tanto que su aparición está directamente relacionada con las exposiciones agudas e intensas a la radiación solar.

A pesar de estar constituida por tres principales bloques (epidermis, dermis e hipodermis, siendo respectivamente cada uno más profundo que el anterior), es habitual encontrar lesiones en la capa más superficial y delgada de todas, cuya presencia celular es significativa. Al ser su exposición mayor que en el resto de los niveles, este contacto directo con el exterior puede provocar la aparición de manchas de color que muchas veces terminen siendo dañinas para el organismo, aunque cierto es que también pueden resultar ser benignas e inofensivas.

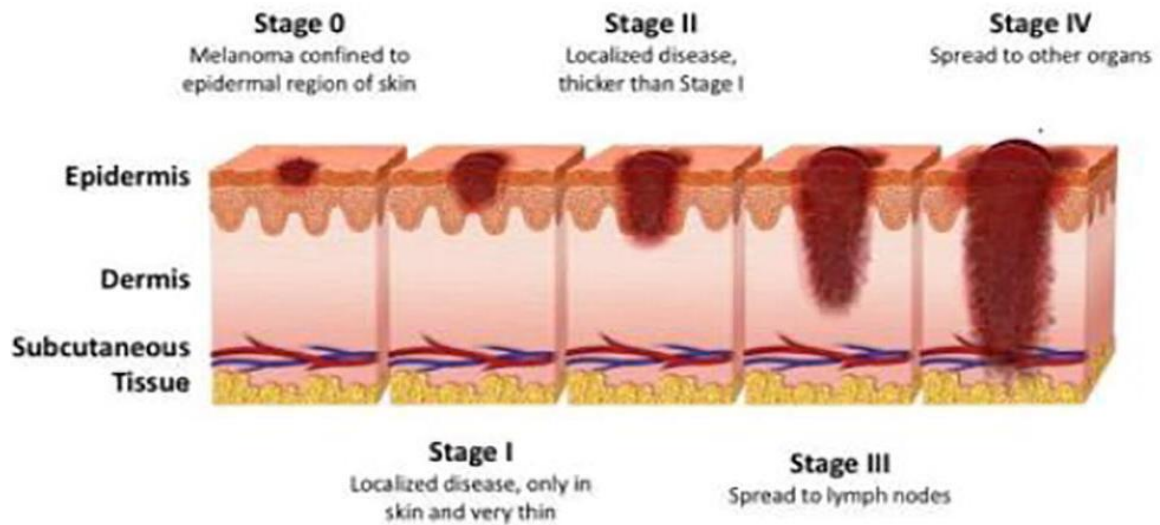


Figura 3 - Etapas del melanoma maligno (American Cancer Society, 2019)

Los tumores de piel se categorizan en función de si son de tipo melanoma o no melanoma, pues este término es un tipo de malignidad cutánea de riesgo (en caso de ser maligno). Como se ha mencionado, puede surgir a partir de lunares existentes, aunque suele ser más habitual observar su presencia como una nueva marca en la piel. A pesar de que su temprana aparición tenga lugar en la capa más superficial, en caso de ser peligroso el melanoma puede atacar los niveles inferiores, pudiéndose extender hacia otras regiones del cuerpo (*Figura 3 - Etapas del melanoma maligno*). Por otro lado, la buena noticia es que la temprana y correcta detección de un melanoma maligno permite su tratamiento, para así ser extirpado gracias a un método conocido como escisión quirúrgica amplia, en el que el cirujano extirpa toda la región del tumor junto con una parte de la piel sana que lo rodea, evitando a toda costa su reproducción.

1.1 MOTIVACIÓN DEL PROYECTO

El enfoque de este proyecto es el desarrollo y entrenamiento de un algoritmo de clasificación enfocado en el Aprendizaje Automático a partir del análisis de imágenes de distintos tipos de afecciones en la piel. Con ello, se quiere enfatizar la posibilidad que

presentan las máquinas para detectar cambios milimétricos que escapan al ojo humano, destacando por su importante papel a la hora de detectar cualquier anomalía que pueda resultar preocupante. Cada año se diagnostican en España unos 4.000 casos de melanoma, cifra que no deja de aumentar pues cada vez la esperanza de vida y la exposición al sol por el bronceado de la piel incrementan notablemente. Se trata de uno de los cánceres que representa solamente el 10% de las neoplasias cutáneas, pero es el responsable directo de más del 90% de las muertes por cáncer de piel, además de ser uno de los más costosos (considerándose su diagnóstico, tratamiento y seguimiento (DiarioFarma, 2018)).

Desde el punto de vista científico, siempre se trata de incorporar nuevos tratamientos medicinales preventivos o preferiblemente mejorar aquellos de los que se dispone hoy en día, cuyos efectos han sido demostrados en forma de varios tipos de ungüentos, como los conocidos protectores solares. Sin embargo, la atención de los grupos científicos recae en su responsabilidad para optimizar y desarrollar estos tratamientos, acudiendo con frecuencia a medios tecnológicos que resulten de gran ayuda para disponer de capacidad suficiente para abordar la temática desde otras perspectivas, estando éstas enfocadas en la puesta en marcha de nuevas tecnologías o entornos de trabajo.

En la sociedad en la que nos encontramos es habitual disponer de capacidades a nivel de computación y cálculo prácticamente ilimitadas que permitan hacer frente a problemas de gran envergadura. Por ello surge un amplio abanico para el sector de la investigación y sus integrantes que permita una colaboración estrecha con el campo medicinal mediante el tratamiento en detalle de imágenes. Este nexo se viene presentando con frecuencia en los últimos años (*Figura 4 - Ejemplo de escala de colores y TAC*), en los que se han obtenido destacables avances (entre ellos uno de los más conocidos es la Tomografía Axial Computarizada (TAC) y su procesamiento en escala de colores).

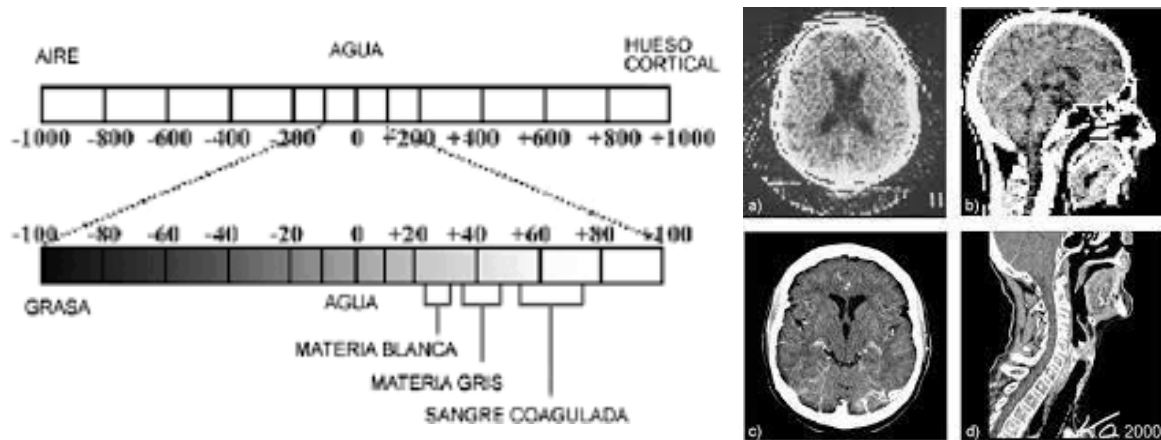


Figura 4 - Ejemplo de escala de colores y TAC (Calzado & Geleijns, 2010)

El presente Trabajo Fin de Grado aprovecha la relación de estos dos importantes campos científicos para, mediante el conocimiento en cada uno de ellos, afrontar un sistema que permita detectar afecciones relacionadas con el cáncer de piel, o por el contrario ser capaz de obtener una respuesta óptima en caso de que se trate de una mancha inofensiva o benigna, pudiendo así clasificar el estado del paciente según el resultado.

De forma general, la mayoría de la población no conoce el tipo y estado de una enfermedad en la piel, pues algunas de ellas muestran los síntomas varios meses después, lo que hace que la enfermedad se desarrolle y crezca más, debido a la falta de conocimientos médicos del público. A veces, un dermatólogo también puede tener dificultades para llevar a cabo un diagnóstico de la afección cutánea, requiriendo costosas pruebas de laboratorio para su correcta identificación. Además, un diagnóstico precoz puede ser esencial para salvar una vida y en el caso de cáncer de piel, una semana de retraso puede hacer que disminuya la tasa de supervivencia en un 30%.

El avance de la tecnología médica basada en el láser ha permitido diagnosticar las enfermedades de la piel de forma mucho más rápida y precisa, siendo por otra parte su coste limitado y muy caro. Por lo tanto, se propone un enfoque basado en el procesamiento de imágenes para diagnosticar ciertas enfermedades de la piel. Si bien es cierto que estas son numerosas, en el presente estudio se enfatizan las más comunes en pieles pigmentadas. Como se ha comentado, se toman las imágenes (dermatoscópicas) del área de la piel afectada

permitiendo entrenar el algoritmo basado en Aprendizaje Profundo para identificar el tipo de enfermedad por clasificación. El enfoque propuesto pretende ser rápido, y no requiere un equipo costoso más allá de un dispositivo móvil con una cámara y un ordenador.

Es sumamente importante recalcar que tanto la Inteligencia Artificial como la información disponible en Internet no sustituye el conocimiento humano. A pesar de que esta herramienta pueda ser útil para identificar posibles afecciones, se debe consultar la opinión y el diagnóstico de un especialista dermatólogo.

1.2 DEFINICIÓN DEL PROYECTO

Previamente se ha hecho mención del coste que pueden alcanzar ciertos diagnósticos médicos profesionales relacionados con el campo de la dermatología en el que nos encontramos. Si bien es cierto que buena parte de la población dispone de los recursos necesarios para acceder a éste, son numerosas las persona que por desgracia no tienen los medios suficientes para lograrlo. Como se ha mencionado, la importancia del diagnóstico temprano en afecciones relacionadas con el cáncer de piel es vital si se quiere intentar remediarlo, por lo que afrontar los costes que puede representar y su seguimiento no será fácil para aquellas personas que no cuenten con capital o suficiente poder adquisitivo para afrontarlo. Es por esto por lo que este proyecto se enfoca en dar accesibilidad y en la obtención de un primer diagnóstico que pueda orientar al paciente en cuestión, dotándole de una herramienta que pueda servirle de guía para conocer más en detalle su estado dermatológico sin necesidad de pago por ello.

De esta forma, el objetivo principal del proyecto es permitir prevenir el avance de lesiones dermatológicas que supongan un peligro para los pacientes. Para lograrlo se requieren dos pasos:

- ✚ Desarrollo de un entorno de trabajo.
- ✚ Aplicación en formato de programa informático.

En nuestro caso, se pretende dar solución al primer punto, concluyendo el alcance del trabajo con el diseño del interfaz para que la aplicación funcional sea desarrollada como trabajo futuro. En este sentido, se tratará de combinar el trabajo realizado con todas las funcionalidades que se quieran añadir en el desarrollo de la aplicación, además de las definidas en el diseño contemplado en el presente estudio.

Para el desarrollo de la totalidad del trabajo, como se detallará en una sección más adelante dedicada a este aspecto, se ha empleado una metodología ágil para ser capaces de poder mejorar el *Producto Mínimo Viable* (MVP de sus siglas en inglés) establecido inicialmente. En este caso, se trata de un clasificador entrenado para distinguir entre los tipos de lesiones más comunes en la piel que, como se verá posteriormente, están clasificadas en 7 clases diferentes.

1.3 OBJETIVOS

La finalidad de este proyecto, como se ha recalcado previamente, es facilitar el acceso a un primer diagnóstico que permita orientar al paciente sobre su estado dermatológico de su piel. Para lograrlo, se han establecido diferentes objetivos, pudiéndose hacer una distinción entre los que debe cumplir el proyecto y la aportación que hace el mismo a nuestra sociedad.

Por una parte, los correspondientes al desarrollo del entorno de trabajo son los siguientes:

- ✚ Creación del entorno de trabajo y del clasificador de imágenes dermatoscópicas.
- ✚ Comparación de un enfoque secuencial a partir de una CNN con una ResNet.
- ✚ Diseño gráfico del interfaz para el entorno de trabajo.

En cuanto a la aportación social, se quiere cumplir con:

- ✚ Facilitar accesibilidad a un diagnóstico temprano que en muchas ocasiones supone un coste importante para los pacientes y es difícil de asumir según cada caso y entorno personal.
- ✚ Optar a una detección temprana del cáncer de piel pues en caso contrario resulta incurable.

Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

2.1 KERAS

El desarrollo llevado a cabo en este proyecto se hecho empleando el lenguaje de programación *Python*, haciendo uso principalmente de la librería *Keras*. Se trata de un lenguaje altamente utilizado en el sector del Aprendizaje Automatizado, al igual que la librería citada, perteneciente a *TensorFlow* y que cuenta con un desarrollo muy extenso, permitiendo de esta forma automatizar numerosos procesos. Otro aspecto por considerar es que, gracias a su popularidad, ambas herramientas cuentan con extensa documentación disponible al público, facilitando cualquier búsqueda y en ocasiones, permitiendo resolver problemas de forma adecuada y accesible.

Para entender correctamente *Keras*, es importante conocer sus orígenes, es decir, *TensorFlow*. Al igual que la primera, se trata de una librería *open source*, en este caso desarrollada por *Google*, capaz de entrenar algoritmos de Aprendizaje Automático Profundo, pudiendo ser testeados para así medir su rendimiento (*TensorFlow*, s.f.). Por otro lado, *Keras* contiene un amplia variedad de Redes Neuronales, accesibles a alto nivel y ejecutables sobre *TensorFlow*, de ahí su estrecha relación. Con todo esto presente, el uso de estas herramientas queda justificado, pues en ambos casos permiten procesar imágenes para ser clasificadas según las clases definidas en el proyecto, y son de gran utilidad para el entrenamiento de los modelos de Aprendizaje Profundo (*Wikimedia*, 2018).

2.2 GOOGLE COLABORATORY

Esta herramienta consiste en un interfaz muy similar a los *Jupyter Notebooks* que *Conda* proporciona, permitiendo al usuario desarrollar código *Python* en el navegador y ejecutarlo a la vez. De nuevo, la compañía *Google* ha sido la encargada de llevar a cabo este

producto, ofreciendo recursos adicionales de GPU (Unidad de Procesado Gráfico) y TPU (Unidad de Procesado de Tensores) óptimos para trabajos relacionados con el Aprendizaje Automatizado, cuyos requisitos implican procesar grandes cantidades de datos a la vez. La versión empleada en este proyecto ofrece el modelo K80 de GPU y un tiempo de ejecución de hasta 12 horas por cuaderno (Google, s.f.).

2.3 HAM10000 DATASET

El conjunto de datos utilizado en este Trabajo Fin de Grado está compuesto de 10015 imágenes dermatoscópicas y que se recogieron durante un periodo de 20 años de dos sitios diferentes. Por un lado, el Departamento de Dermatología de la Universidad Médica de Viena (Austria) intervino en el proceso y, por otra parte, la consulta de cáncer de piel de Cliff Rosendahl en Queensland (Australia) también hizo su respectiva aportación. El centro australiano almacenó imágenes y metadatos en archivos PowerPoint y bases de datos Excel mientras en Austria recopilaron imágenes antes de la era de las cámaras digitales y almacenaron todos los conjuntos. Posterior a esta recopilación, diferentes profesionales se encargaron de realizar algunos retoques de calidad y filtrado para estandarizar las imágenes (*Figura 5 - Flujo esquemático de los métodos de elaboración del conjunto final de datos* (Tschandl, Rosendahl, & Kittler, 2018)). Es importante recalcar que el conjunto de datos se ha generado también teniendo en cuenta las siguientes fases:

- ✚ Diseño: creación de la base de datos, integración de los datos y conversión del formato de imagen.
- ✚ Mediciones: a partir de lesiones en la dermis (muestras corporales humanas).
- ✚ Tecnología empleada: conservación digital de imágenes.
- ✚ Factores: diagnóstico y respectivo procedimiento, edad, sexo, región corporal de la afección.

Finalmente, el *dataset* definitivo sirve como conjunto de entrenamiento para fines académicos de Aprendizaje Automático. En él, se cuenta con una colección representativa

de todas las categorías de diagnóstico importantes y más relevantes del ámbito de las lesiones dermatológicas en pieles pigmentadas:

- ✚ Queratosis actínicas y carcinoma intraepitelial o enfermedad de Bowen (akiec)
- ✚ Carcinoma de células basales (bcc)
- ✚ Lesiones benignas similares a la queratosis (bkl)
- ✚ Dermatofibroma (df)
- ✚ Melanoma (mel)
- ✚ Nevos melanocíticos (nv)
- ✚ Lesiones vasculares (vasc)

Más de la mitad de las lesiones aportadas han sido confirmadas mediante histopatología (histo). El porcentaje restante es validado por exámenes de seguimiento (follow_up), consensos de expertos (consensus) o la confirmación por microscopía confocal in vivo (confocal).

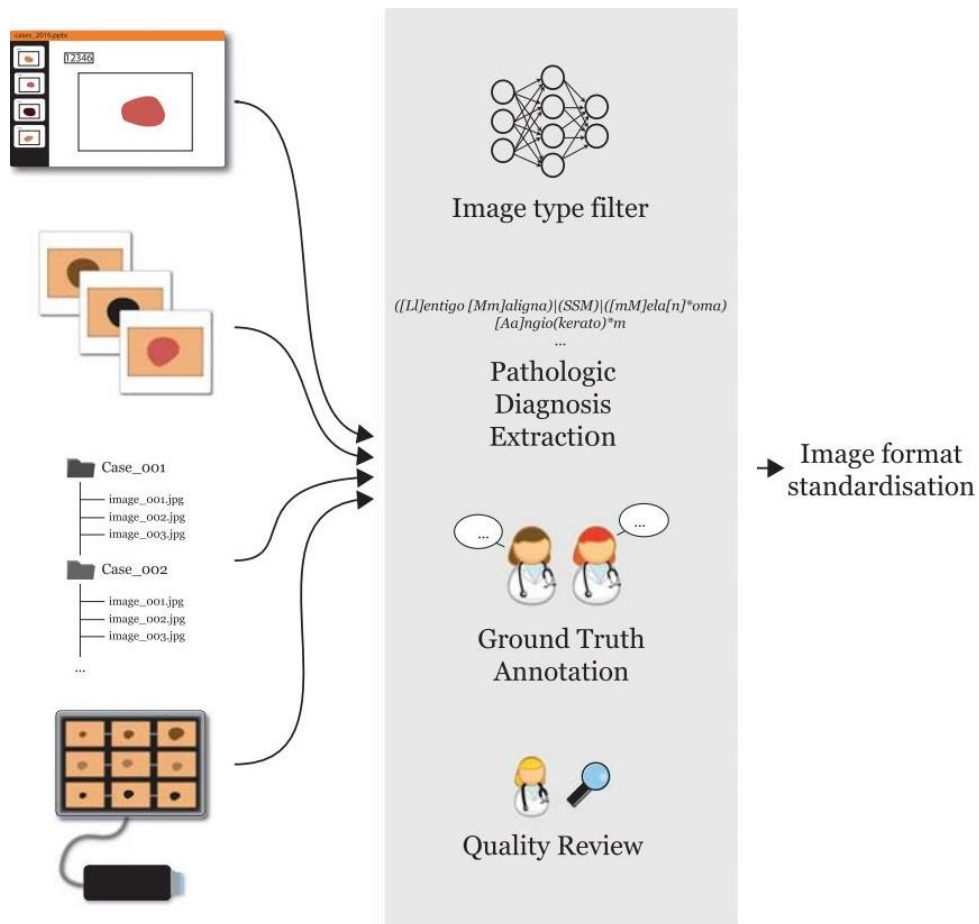


Figura 5 - Flujo esquemático de los métodos de elaboración del conjunto final de datos (Tschandl, Rosendahl, & Kittler, 2018)

Capítulo 3. ESTADO DE LA CUESTIÓN

Es evidente que, los avances y desarrollos que se han obtenido hasta la fecha y los que se están llevando a cabo en el sector de la medicina relacionados con la detección de afecciones en la dermis requieren una aportación relevante del sector tecnológico. En el caso que nos atañe, juega un papel importante el campo de la Inteligencia Artificial y más precisamente el área del Aprendizaje Automatizado, pues sus algoritmos permiten observar patrones que a simple vista podrían pasar desapercibidos o variables que pueden haber sido consideradas como no relevantes.

Por otra parte, una de las medidas más repetidas por los especialistas médicos es la supervisión de nuestra piel y el control de nuestros lunares para, ante cualquier cambio inesperado, poder acudir al médico y descartar melanomas o cualquier otro posible problema derivado de manchas en la dermis. Distintas son las organizaciones que, mediante el desarrollo de aplicaciones, han querido hacer su contribución para ayudar a saber cuan peligrosa es una mancha presente en nuestro cuerpo. La combinación de estos desarrollos con el uso de la metodología de Redes Neuronales en los últimos años ha destacado por su imparable ascenso y popularidad en los entornos de programación, en detalle su técnica de Aprendizaje Profundo.

3.1 SKINVISION

SkinVision es una aplicación disponible para versiones móviles que implementa un algoritmo capaz de analizar manchas y lunares presentes en la dermis registrándolos en una fotografía, pudiéndose hacer directamente con el propio dispositivo móvil (aplicando el sistema de iluminación proporcionado por la linterna del dispositivo para obtener una mayor calidad). Si se detecta algún tipo de anomalía, es capaz de generar recomendaciones para el usuario, incluyendo una evaluación médica en un centro especializado compartiendo los

datos con los profesionales sanitarios si así lo requiere la situación. Para obtener un seguimiento adecuado de la afección, cuenta con la opción de almacenar diferentes capturas de una misma región para poder archivarlas y obtener un evolutivo para evaluar posibles cambios anómalos (*Figura 6 - Interfaz gráfico y funcionamiento de SkinVision*).

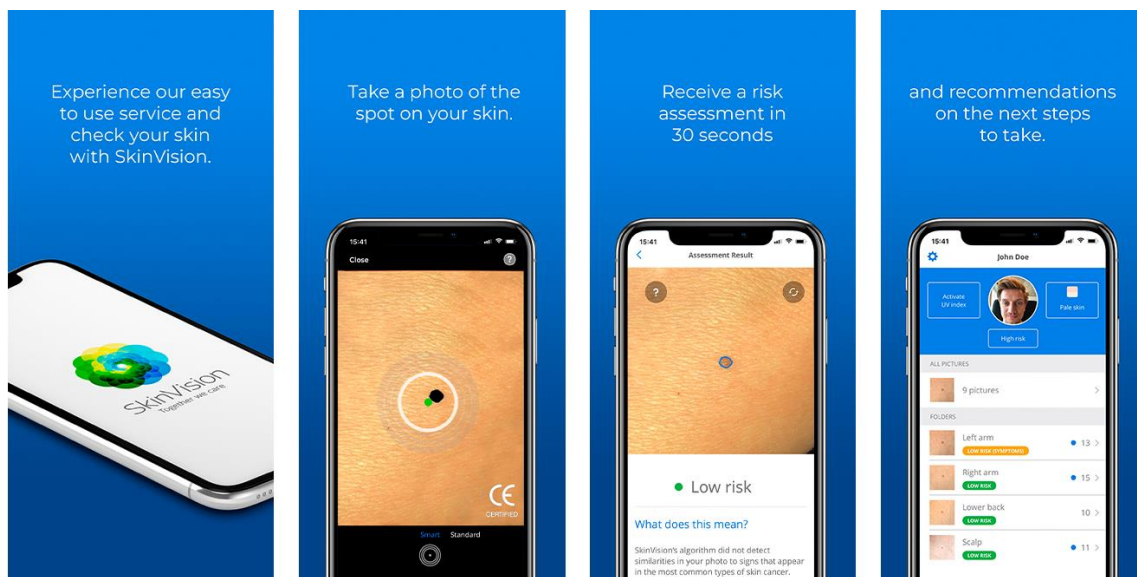


Figura 6 - Interfaz gráfico y funcionamiento de SkinVision (MHealth Belgium, s.f.)

El algoritmo detrás de este interfaz implementó el Aprendizaje Automatizado a principios de 2018 y desde entonces se ha ido entrenado con grandes cantidades de imágenes evaluadas previamente por un equipo de dermatólogos, permitiéndole aprender qué lesiones son peligrosas y cuáles no (SkinVision, 2022). A día de hoy, el servicio ofrecido por esta aplicación es únicamente de pago y está dividido en diferentes tarifas, donde destaca la más cara (49,99 euros anuales) ofreciendo varias funcionalidades de seguimiento no disponibles en categorías inferiores (haciendo hincapié en el enfoque personal).

3.2 MOLEXPLORE

De forma parecida al caso visto anteriormente, Molexplore sigue un funcionamiento similar a SkinVision, añadiendo un paso más y permitiendo también un seguimiento en paralelo de otros usuarios, como pueden ser padres o hijos. Mediante copias de seguridad para almacenar toda la información, se tiene un historial de cada persona que también, al igual que en SkinVision, facilita la visión de la evolución de cada individuo (*Figura 7 - Interfaz gráfico de la aplicación Molexplore*). Además, su importante valor añadido reside en un dispositivo médico portátil llamado *Matiz*, usado para tomar fotografías multiespectrales (revelando información clave, invisible al ojo humano), encargadas de alimentar el sistema de Inteligencia Artificial conocido como Big Derma. Una ventaja evidente frente al sistema mencionado antes es que en este caso se puede tener acceso al servicio ofrecido por la aplicación de manera gratuita y sin ningún coste de descarga (Molexplore, 2022). Sin embargo, el dispositivo dermatoscópico empleado para la toma de imágenes supone un gasto no asumible para todo el público en caso de querer realizar un trabajo o estudio independiente, sin presupuesto alguno, como resulta en nuestro caso.

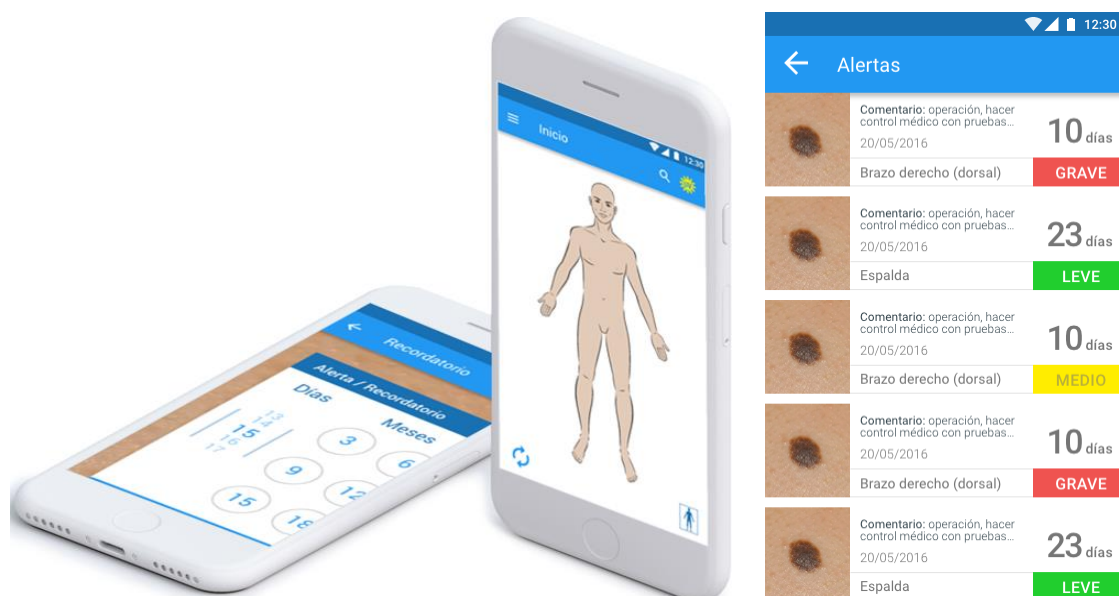


Figura 7 - Interfaz gráfico de la aplicación Molexplore (Aptoide, s.f.)

3.3 *EDERMA*

Avalada por la Academia Española de Dermatología y Venereología (AEDV), este proyecto ayuda a realizar el seguimiento de las lesiones de piel, haciendo hincapié en la autoexploración de la dermis. De esta forma, supone una ayuda adicional para que cualquier persona realice un correcto seguimiento de sus lesiones y lunares (Arellano, 2018), permitiendo, al igual que en el resto de los casos, hacer fotos específicas de cualquier región del cuerpo, pudiéndose clasificar y almacenar para valorar, con el paso del tiempo, si hay cambios significativos para comunicárselos a un especialista. Esta vez, la aplicación hace énfasis en lo importante que resulta tener una estrategia útil como lo es la memoria objetiva con el registro de las fotografías de la lesión, más allá de la foto fija que el ojo ve en un momento determinado y de nuestros recuerdos, que pueden resultar ser variables o poco claros con el paso del tiempo.

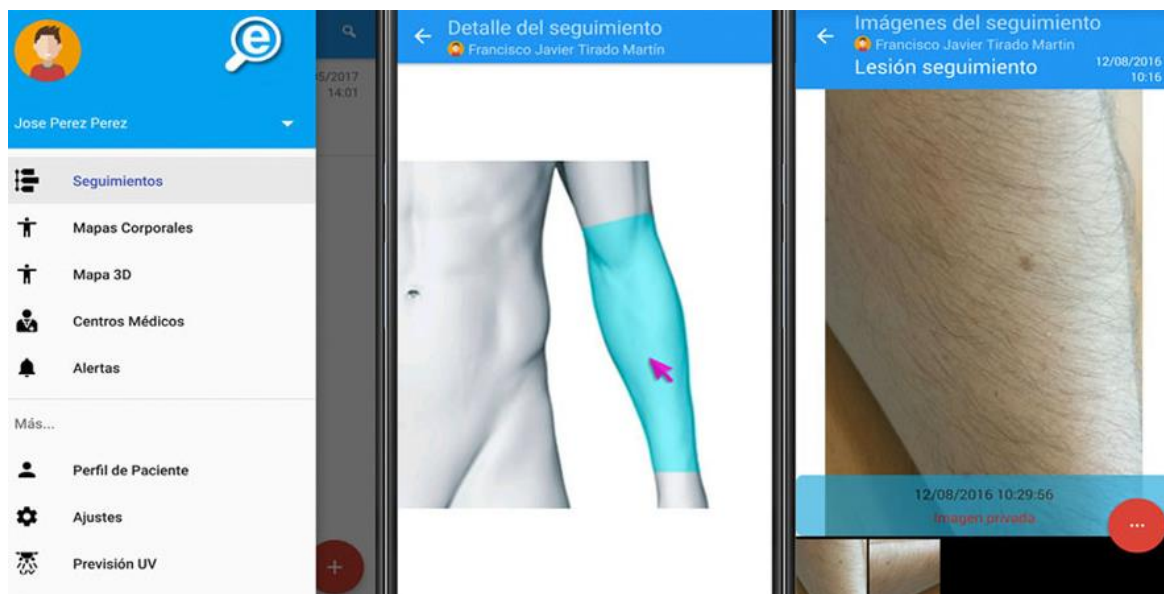


Figura 8 - Interfaz gráfico de eDerma (Arellano, 2018)

3.4 DEEP LEARNING

El campo del Aprendizaje Profundo está basado en las conocidas Redes Neuronales artificiales (ANNs de sus siglas en inglés), que resultan ser modelos de Aprendizaje Automatizado inspirados en las neuronas naturales (biológicas) del cerebro humano. De forma más detallada, se trata de una red de nodos, neuronas, enlazadas entre sí mismas según sus respectivos pesos. La red es la encargada de transformar los valores de entrada para obtener los valores de salida (*inputs* y *outputs* respectivamente) que permiten representar conceptos tales que la probabilidad de pertenecer a una cierta categoría (clasificar). Por lo tanto, las ANNs son la base necesaria del Aprendizaje Profundo para poder afrontar problemáticas complejas de reconocimiento textual o clasificación de imágenes, pues resultan ser versátiles, potentes y escalables (Géron, 2017).

3.4.1 INSPIRACIÓN BIOLÓGICA

Las neuronas son células localizadas en el cerebro capaces de comunicarse entre ellas a través de las sinapsis, conexiones especializadas que se encuentran entre los terminales de los axones de una neurona y las dendritas de su vecina (*Figura 9 - Diagrama del cuerpo celular de una neurona*). Es en esta conexión celular donde tiene lugar la onda de descarga eléctrica (potencial de acción) que establece la liberación de un neurotransmisor por parte del cuerpo neuronal. A su vez, este neurotransmisor puede ser capaz de desencadenar el envío del potencial de acción por parte de otra neurona en caso de disponer de la cantidad necesaria en escasos milisegundos para ello (Uq.edu.au, 2016). En 1949, Donald Hebb publicó el libro *The Organisation of Behavior*, donde sugirió el hecho de que una conexión entre dos neuronas utilizada de forma frecuente y recurrente concluye en un refuerzo, lo que en tal caso habilita el aprendizaje. A este proceso se suele hacer referencia con el nombre de teoría *Hebbiana*. Gracias a esta constatación, se sabe que las neuronas, a nivel individual, poseen un comportamiento sencillo, pero, sin embargo, cuando éstas se organizan en redes (implicando millones de neuronas), su capacidad de cálculo aumenta hasta el punto de poder realizar operaciones altamente complejas (Géron, 2017).

Neuron

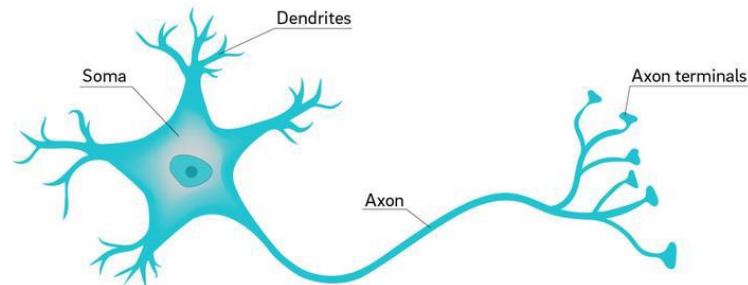


Figura 9 - Diagrama del cuerpo celular de una neurona (I.E.S. Suel, s.f.)

Previa a la constatación de Hebb, en el año 1943 se mencionan en el artículo *A logical calculus of ideas inmanent in nervous activity* a las Redes Neuronales Artificiales de la mano de McCulloch y Pitts, que proponen un modelo abstracto de la neurona biológica. En éste, cada cuerpo celular cuenta con uno o varios *inputs* y un *output* binarios. La salida es activada si más de cierto umbral de sus entradas han sido activadas. De esta manera y a partir de la representación de la *Figura 10 - Cálculo de diferentes funciones lógicas a partir del modelado de una neurona* se obtienen las siguientes funciones lógicas:

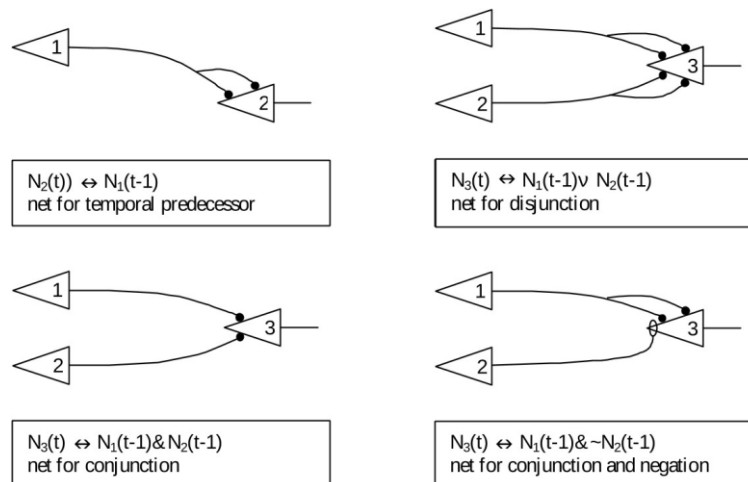


Figura 10 - Cálculo de diferentes funciones lógicas a partir del modelado de una neurona (Piccinini, 2004)

3.4.2 PERCEPTRÓN

Se conoce como perceptrón a la arquitectura básica y más simple de las ANNs, cuya composición neuronal incluye la Unidad Lógica de Umbral (TLU de sus siglas en inglés). Principalmente, la diferencia con la inspiración biológica tratada en el apartado anterior reside en la interpretación de sus *inputs* y *outputs*, pues estos ya no son binarios. La salida se obtiene mediante la suma ponderada de los valores de entrada a los que se les aplica una función de activación, permitiendo así realizar una toma de decisión para el siguiente paso (Figura 11 - Representación de la Unidad Lógica de Umbral).

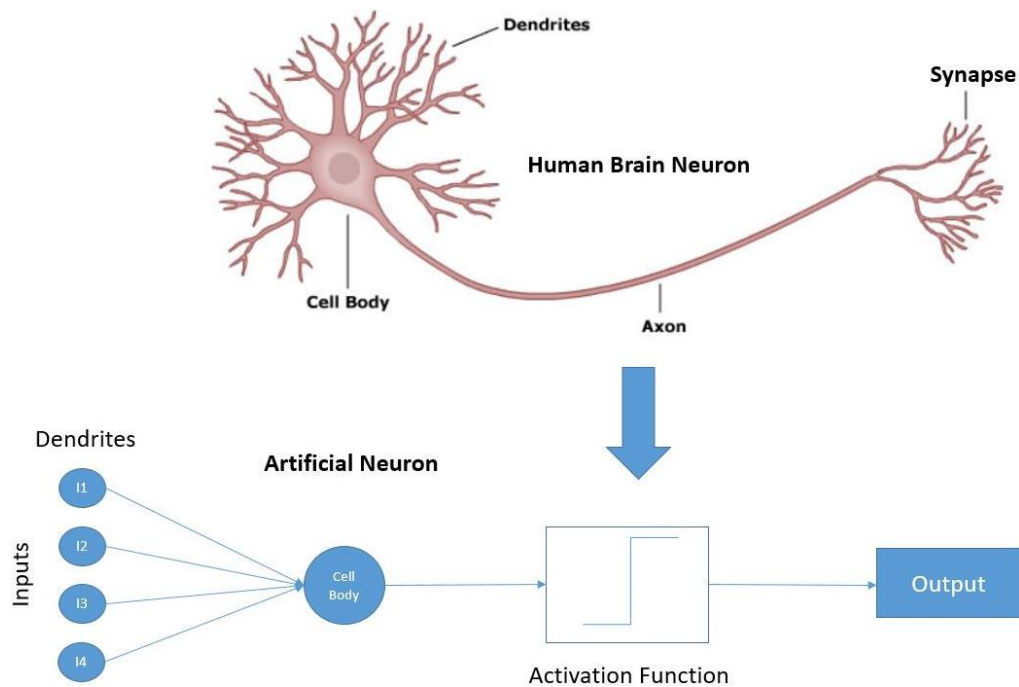


Figura 11 - Representación de la Unidad Lógica de Umbral (Sah, 2018)

Numerosas son las funciones de activación que se pueden emplear según el propósito de la arquitectura, pero las más habituales resultan ser la conocida función signo y la función escalón *Heavyside* (Figura 12 - Funciones lógicas más comunes). Gracias a estas operaciones matemáticas, la Unidad Lógica de Umbral puede realizar una clasificación (binaria) donde cada valor de salida represente una clase distinta. Por otra parte, un solo perceptrón se compone de una capa de estas TLUs en la que cada TLU individualmente se

conecta a todos los valores de entrada; en algunos casos, puede ser habitual añadir una constante *bias*.

$$\text{sign}(x) = \begin{cases} 1, & x > 0; \\ 0, & x = 0; \\ -1, & x < 0. \end{cases} \quad H(x) = \begin{cases} 1, & x > 0; \\ \frac{1}{2}, & x = 0; \\ 0, & x < 0. \end{cases}$$

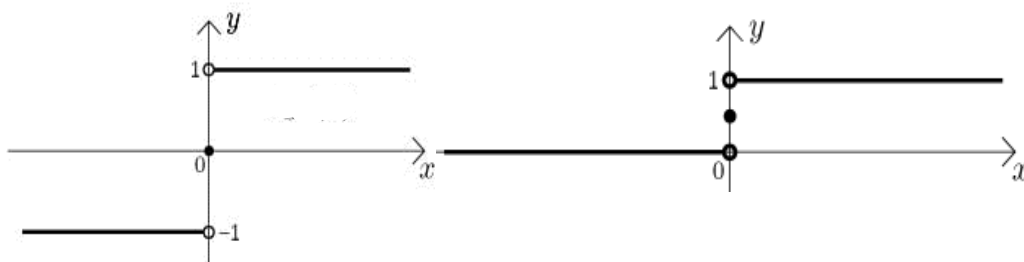


Figura 12 - Funciones lógicas más comunes

El entrenamiento de un perceptrón resulta muy parecido al de la teoría desarrollada por *Donald Hebb* descrita en el apartado anterior. En este caso, se considera el error cometido durante una predicción, permitiendo así reforzar las conexiones que ayudan a reducirlo de forma significativa. En otras palabras, el conjunto se alimenta con una entrada, pero si ésta concluye en una predicción equivocada, los pesos que hubiesen contribuido a realizar la predicción correcta son reforzados (Géron, 2017).

3.4.3 REDES NEURONALES PROFUNDAS

Desafortunadamente, las limitaciones de los perceptrones son significativas. Por una parte, no son capaces de calcular la probabilidad de pertenecer a cada clase y, además, no pueden afrontar problemáticas de escasa dificultad como las puertas lógicas *XOR*. A pesar de ello, si se apilan varias capas de perceptrones, las debilidades intrínsecas del componente pueden ser erradicadas. Este concepto es conocido como Perceptrón Multicapa (MLP de sus siglas en inglés), cuya composición se refleja en *Figura 13 - Arquitectura Perceptrón Multicapa*. En primer lugar, esta estructura se compone por una capa con los valores de entrada, seguida de una o varias capas intermedias (conocidas como capas ocultas) conectadas a la última capa (de salida). El entrenamiento de esta estructura se basa en el

algoritmo del Descenso de Gradiente² (*Error! No se encuentra el origen de la referencia.*), también llamado algoritmo de propagación hacia atrás (Wikipedia Contributors, 2021).

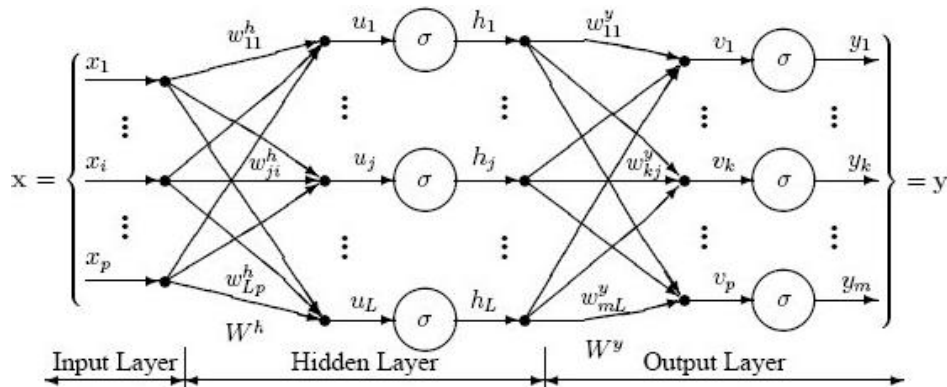


Figura 13 - Arquitectura Perceptrón Multicapa (Flores)

Como bien indica su nombre, su funcionalidad permite obtener el gradiente de la red únicamente realizando dos pasos por ella (primero hacia delante y después hacia atrás). Con el gradiente ya calculado, se lleva a cabo su descenso y se repite el proceso hasta lograr la convergencia de la red en una solución.

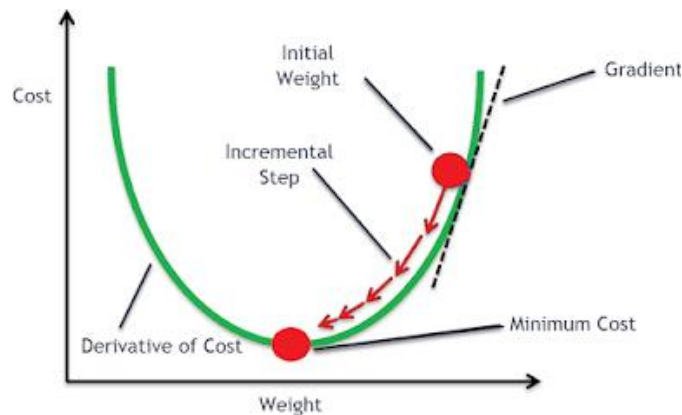


Figura 14 - Representación del algoritmo de Descenso de Gradiente (Keshav, s.f.)

² Conocido como *Gradient Descent*, este algoritmo se optimiza de forma iterativa para encontrar un mínimo de una función diferenciable, dando pasos en la dirección opuesta al gradiente en el punto actual (dirección de descenso más pronunciada).

Sin embargo, a este algoritmo no es posible aplicarle ningún tipo de función escalón a modo de función de activación, pues únicamente contiene secciones planas y no hay gradientes con los que trabajar, motivo por el cual la función sigmoide ($\sigma(x) = \frac{1}{1+e^{-x}}$) es empleada en este caso. A pesar de ello, en la actualidad es más habitual hacer uso de la función $ReLU(x) = \max(0, x)$ porque su velocidad de cálculo es mayor y permite esquivar problemas de desvanecimiento de gradientes que provocan la reducción de este a valores tan sumamente pequeños que resulta de extrema dificultad modificar su valor. A continuación, en *Figura 15 - Funciones de activación* se pueden apreciar las tres funciones de activación citadas previamente, donde todas ellas anulan posibles valores negativos.

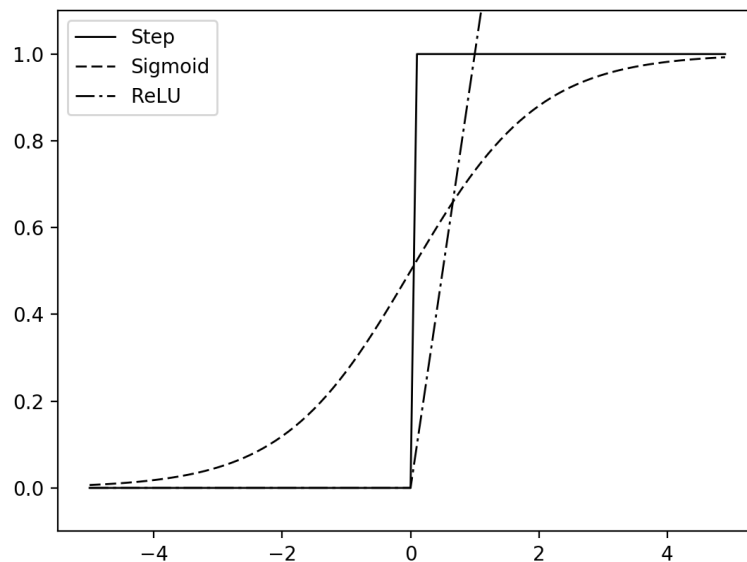


Figura 15 - Funciones de activación (Thinkage, 2017)

Pasamos ahora a detallar los pasos que se llevan a cabo durante la etapa de entrenamiento del famoso algoritmo de Descenso de Gradiente. En primer lugar, se realiza una predicción sobre un subconjunto de entrenamiento (*training set*) permitiendo actualizar los pesos del modelo sin necesidad de emplear todas sus instancias³ (para determinar el tamaño del subconjunto es habitual variar entre potencias de dos tales como 16, 32, 64, etc.);

³ Esta técnica permite reducir el tiempo de convergencia en conjuntos de datos de gran envergadura.

cada paso por el subconjunto de entrenamiento se conoce como época (*epoch*). Tras esta predicción, se mide el error cometido y se procede a pasar de nuevo por las diferentes capas del modelo, pero en sentido inverso (*backpropagation*), midiendo así la participación de cada peso en el error obtenido previamente. Por último, se modifican los pesos asignados para reducir el error total, acción conocida como *tunear* o *tunning*.

Con esta arquitectura, podemos realizar múltiples tareas de clasificación. En el caso de la clasificación binaria, la salida puede ser interpretada como la probabilidad de pertenecer a una clase definida, pues se trata de un valor entre 0 y 1. Si existieran varias clases no mutuamente excluyentes (con uno o más resultados en común), cada valor de salida representaría la probabilidad de pertenencia a cada una de estas. Si por lo contrario fuesen mutuamente excluyentes (sin resultados compartidos), tal caso requeriría una función de activación *softmax* en la salida, permitiendo así que las probabilidades sumaran 1.

3.4.4 SOBRE APRENDIZAJE

Un buen modelo de Aprendizaje Automático debe ser capaz no solo de predecir de forma correcta las muestras con las que ha sido entrenado, sino que también debe poder generalizarse de forma adecuada para lograr de esta forma mantener la precisión obtenida con nuevos datos. Este objetivo se alcanza dividiendo el conjunto de datos de trabajo en los siguientes subconjuntos:

- ✚ Entrenamiento (*training*).
- ✚ Testeo (*test*).
- ✚ Validación (*validation*).

Como es evidente, el primer subconjunto sirve para alimentar (entrenar) el modelo con el que se está trabajando. En cuanto al segundo y tercer subconjunto, lo más habitual resulta ser combinar ambas secciones dentro de un mismo subgrupo de *validation*, permitiendo así asegurar que se trabaja de forma correcta con nuevas muestras. Con los resultados obtenidos para los dos subconjuntos definidos, se hace uso de diferentes hiper-

parámetros⁴ que permiten seleccionar el modelo que mejor encaje para cada casuística. El proceso termina midiendo el rendimiento del modelo establecido con el conjunto de *validation* definido. De esta forma, podremos validar que no ha habido errores en la elección del modelo más óptimo, pues se selecciona el que mejor respuesta obtiene tanto en los conjuntos de entrenamiento y testeo como en el de validación (el modelo debe ser capaz de generalizar de forma adecuada).

Cuando un modelo no puede retener de manera correcta la relación entre los valores de entrada y los de salida, se obtiene un error considerable en el apartado de entrenamiento y en el de testeo, presentando así *underfitting*. De forma más generalizada, esto ocurre cuando el modelo es demasiado simple (no presenta suficientes variables de entrada o su tiempo de entrenamiento es muy escaso). En estas situaciones, el error que presentan *training* y *test* a lo largo de toda la curva del error medido durante la etapa de entrenamiento del modelo resulta alto, por lo que se concluye que carece de aprendizaje. Sin embargo, esta situación es remediable; por un lado, se puede probar a aumentar el número de iteraciones, aunque si el problema persiste, el modelo tendrá que ser sustituido. Por otra parte, este efecto puede ser observable también mediante la curva de precisión del modelo: si ésta resulta ser relativamente leve a lo largo de todo el periodo de aprendizaje y además presenta signos evidentes de estancamiento en ambos subconjuntos estamos frente a una clara situación de escaso aprendizaje. Entonces, como aumentar en número de épocas no garantiza obtener una mayor precisión, es habitual tener que realizar una modificación en el modelo, aunque es muy importante recalcar que no siempre es posible establecer el error como alto o bajo mediante la función de pérdidas o acierto. Cada caso requiere su propio análisis según sus propios datos, pudiendo así ser necesario definir métricas de interés distintas y evaluar de nuevo el modelo.

En el lado opuesto, el sobre aprendizaje u *overfitting* tiene lugar si el modelo se sobre adapta a los datos de entrada en el conjunto de *training*, concluyendo en no poder

⁴ Parámetros que permiten controlar y calibrar el proceso de aprendizaje del modelo; el resto de los valores de los otros parámetros quedan definidos por el propio entrenamiento.

generalizarse con nuevos datos. Normalmente, estos casos se suelen dar cuando el conjunto de datos de trabajo no es lo suficientemente grande o su interpretación por parte del modelo resulta demasiado compleja. No obstante, también se debe considerar la posibilidad de que el entrenamiento haya sido demasiado extenso, teniéndose así que reducir el número de iteraciones. Si a partir de cierto número de épocas vemos que nuestro modelo se atasca en la etapa de testeo, pero mejora su acierto en el entrenamiento, las alternativas más habituales a este comportamiento pasan por parar antes el entrenamiento si los resultados cumplen con los criterios establecidos inicialmente, incrementar el conjunto de datos en la medida de lo posible o incluso modificar la complejidad del propio modelo.

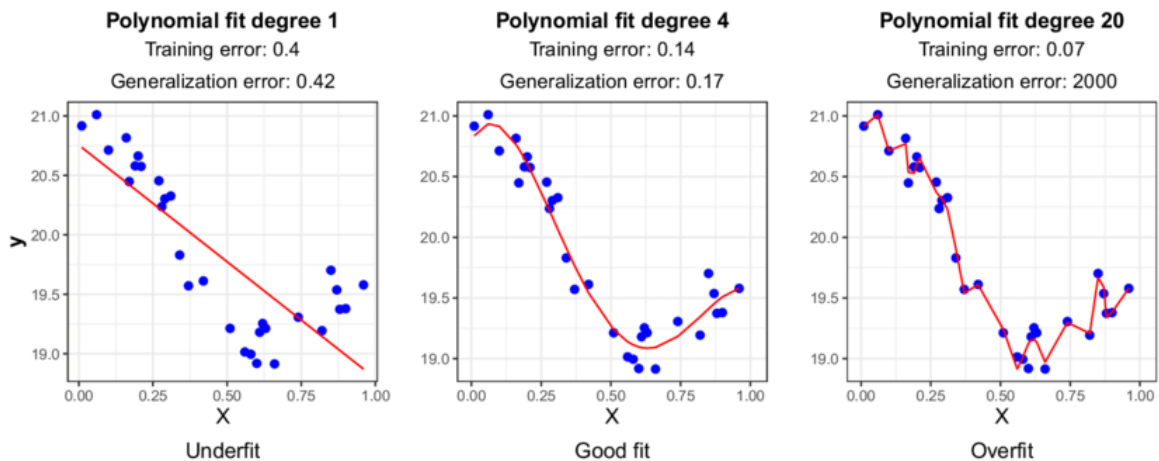


Figura 16 - Comparación entre modelos con falta de aprendizaje, adecuado y con sobre aprendizaje (Rahaman, 2020)

De forma más genérica, en la *Figura 16 - Comparación entre modelos con falta de aprendizaje, adecuado y con sobre aprendizaje* se resumen los conceptos mencionados, aportando una clara visión sobre cada uno de ellos. Mediante tres aproximaciones a un polinomio de distinto grado, se puede observar cómo en cada caso, la diferencia entre el error cometido durante el entrenamiento y una vez se generaliza el modelo (se prueba con otro conjunto de datos), corresponde a los escenarios tratados.

3.4.5 HIPER PARÁMETROS

La selección de hiper parámetros en una red neuronal es considerada una de sus desventajas, pues existen un gran número y su definición no siempre es clara. Para escogerlos correctamente se deben definir⁵:

- ✚ La cantidad de capas ocultas: es recomendable aumentar su número hasta alcanzar sobre aprendizaje.
- ✚ El número de neuronas de cada capa oculta: tanto en la capa de entrada como de salida, el número de neuronas debe coincidir con las respectivas dimensiones de los parámetros de entrada y las clases de salida (se suelen utilizar centenas de neuronas, pero se recomienda emplear más de las que sean necesarias⁶). Normalmente apilar capas resulta más productivo que aumentar el número de neuronas de éstas.
- ✚ El ratio de aprendizaje: por definición es el parámetro que más condiciona el modelo y debe ser ajustado ($[10^{-6};1]$) si se realizan cambios en otros hiper parámetros. Indica la cantidad de pesos que deben ser modificados (durante *backpropagation*) para que nuestro modelo obtenga mejores predicciones.
- ✚ El optimizador: de forma general *Adam optimizer* permite resultados óptimos, pues se basa en el mencionado Descenso de Gradiente, pudiendo así tratar con gradientes dispersos en problemas de elevada dificultad.
- ✚ El tamaño de las muestras: potencias de 2, habitualmente a partir de 32.
- ✚ La función de activación: *ReLU* es una buena opción para emplear en capas ocultas. Sin embargo, en la capa de salida dependerá del objetivo perseguido por el modelo⁷.
- ✚ El número de iteraciones: superior al que se espera implementar.
- ✚ El abandono o *dropout*: permite esquivar el sobre aprendizaje descartando neuronas arbitrariamente ($[0,5;0,8]$).

⁵ Hay algoritmos capaces de optimizar hiper parámetros, pero implican un tiempo de ejecución largo además de un coste computacional alto (normalmente su uso se recomienda cuando se obtiene un buen modelo y se quiere perfeccionar).

⁶ Más vale que sobre que no que falte.

⁷ *Linear* para regresión, *sigmoide* para clasificación no excluyente y *softmax* para clasificación excluyente

3.5 REDES NEURONALES CONVOLUCIONALES

De sus siglas en inglés, las CNNs son de los tipos de Redes Neuronales más utilizados para la clasificación de imágenes. Resulta curioso saber además que estos modelos toman la inspiración del mecanismo utilizado en el córtex visual del cerebro humano (Raschka & Mirjalili, 2017): cada neurona responde a los estímulos que se encuentran en una cierta región del campo visible por nuestro ojo (campo receptivo) y su superposición completa el área visual. Con esta analogía, cada elemento singular del mapa de características de la CNN se asocia con un conjunto específico de píxeles y no con la imagen completa; en cada sección de píxeles se emplean los mismos parámetros (pudiéndose reducir su número).

Este tipo de redes permite un escaso preprocesado de aquellas imágenes con las que se trabaja si lo comparamos con otros algoritmos de clasificación, pues una CNN correctamente entrenada permite aprender los filtros que se tendrían que aplicar manualmente antes de utilizar un modelo más clásico, suponiendo así una clara ventaja. Sus capas se pueden dividir en distintos bloques según la profundidad de cada uno: las capas más superficiales de la red se encargan de extraer aquellas características más genéricas de las imágenes (a bajo nivel), como bordes y manchas, mientras que, a más alto nivel, las capas superiores emplean la información de las capas inferiores para lograr encontrar detalles más precisos (objetos concretos). A continuación, se muestra una representación de las características descritas previamente:

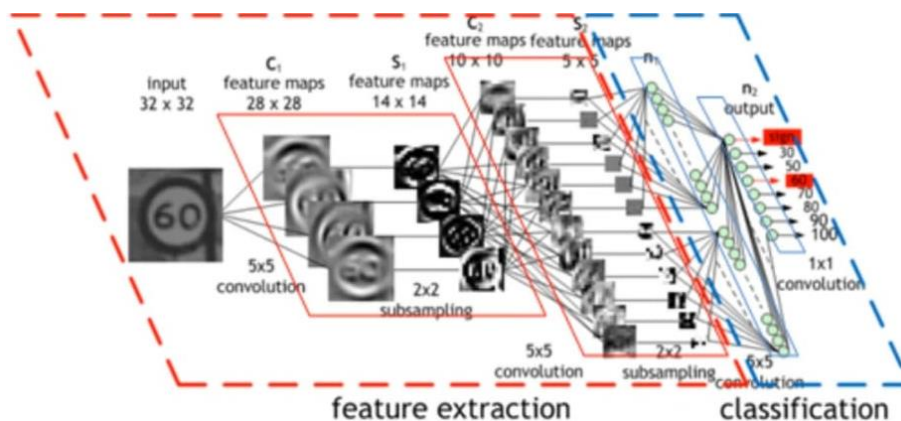


Figura 17 - Etapas en una Red Neuronal Convolutiva

3.5.1 ARQUITECTURA

En primer lugar, la entrada de esta estructura son siempre los propios píxeles de la imagen, todas de tres dimensiones: altura, anchura y profundidad (o canales). En el caso de imágenes de 380 píxeles por 44 píxeles en formato RGB (habitual), la forma de la capa es 380x44x3. Por otra parte, principalmente las Redes Neuronales Convolucionales se componen de las cuatro capas⁸ siguientes (Tensorflow, 2021):

- ✚ Capas convolucionales: equivalen a diferentes filtros (*kernels*) que permiten calcular un producto escalar sobre un área pequeño de píxeles asociado al campo receptivo de la corteza visual humana. Los valores de entrada corresponden a los píxeles de la imagen y como resultado se obtiene un mapa de características con un nuevo dimensionado de altura, anchura y profundidad (en número de filtros).
- ✚ Capas ReLU: se encargan de erradicar los valores negativos asignándolos a cero mediante la aplicación de la función de activación definida.
- ✚ Capas de Pooling: realizan un submuestreo de las dimensiones espaciales de la imagen (altura y anchura), pudiendo así reducir el coste computacional; su función también incluye la extracción de las características más importantes de la imagen. Existen varios tipos de pooling, aunque los más utilizados son *max*⁹ y *average pooling* (Figura 18 - Aplicación de *max pooling* y *average pooling*). Devuelven el valor máximo y el valor medio de cada área de píxeles estudiada, respectivamente.

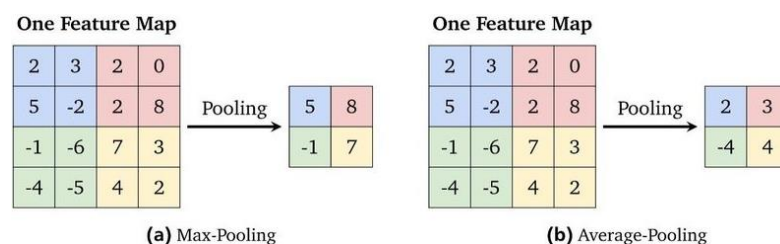


Figura 18 - Aplicación de *max pooling* y *average pooling* (Rahaman, 2020)

⁸ Divididas en dos tipos según la terminología: en primer lugar, la red como conjunto pequeño de capas complejas y en segundo lugar como un conjunto más grande, pero de capas más simples.

⁹ Suele ser más habitual su uso debido a su componente de reducción de ruido.

- ✚ Capas de conexión total: su función permite a la red aprender aquellas combinaciones que no son lineales y pertenecen a las características de más alto nivel.

En la actualidad, varias son las arquitecturas de Redes Neuronales Convolucionales públicas que suelen usarse para afrontar problemas como el de este proyecto. Las más conocidas suelen ser MobileNet, AlexNet, Inception (antes GoogLeNet) o ResNet (todas con sus diferentes versiones).

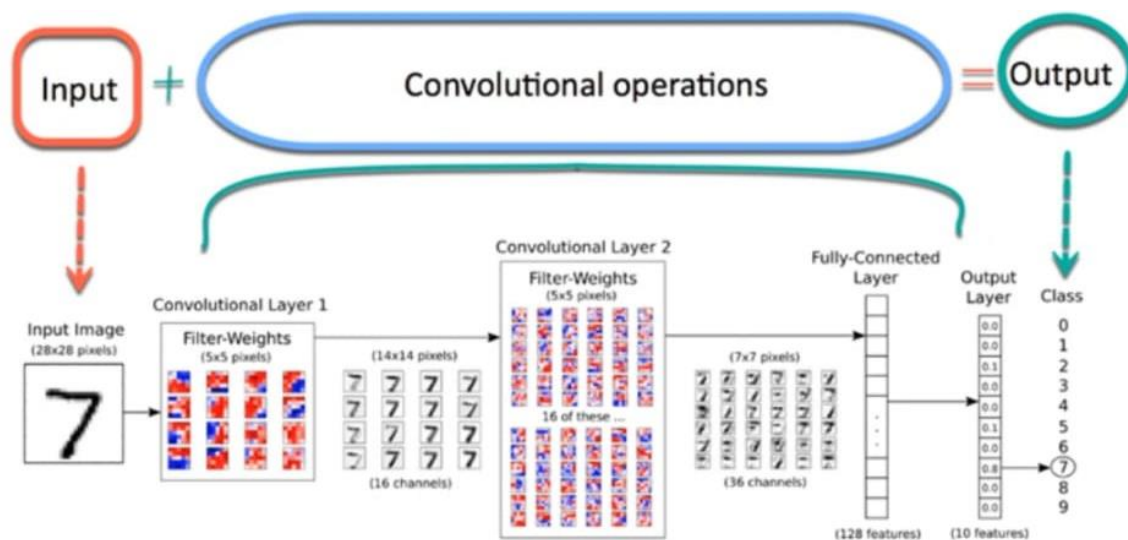


Figura 19 - Arquitectura de una Red Neuronal Convolutiva (Tensorflow, 2017)

3.6 REDES NEURONALES PREVIAMENTE ENTRENADAS

Se conoce como *Transfer Learning* la metodología de aprendizaje supervisado ampliamente utilizada en el campo de la visión por computación, donde la Inteligencia Artificial es capaz de extraer información a partir de imágenes, ofreciendo soluciones a problemas del mundo real. Esta técnica consiste en reutilizar las características (*features*) aprendidas de un problema para resolver otro parecido, pudiendo ser modificadas si es necesario, lo que permite reducir considerablemente el tamaño necesario del conjunto de datos, así como el tiempo de entrenamiento. A pesar de esto, es condición obligatoria que el primer modelo haya sido entrenado con un conjunto de datos bastante amplio (*Figura 20 - Transferencia de Aprendizaje*).

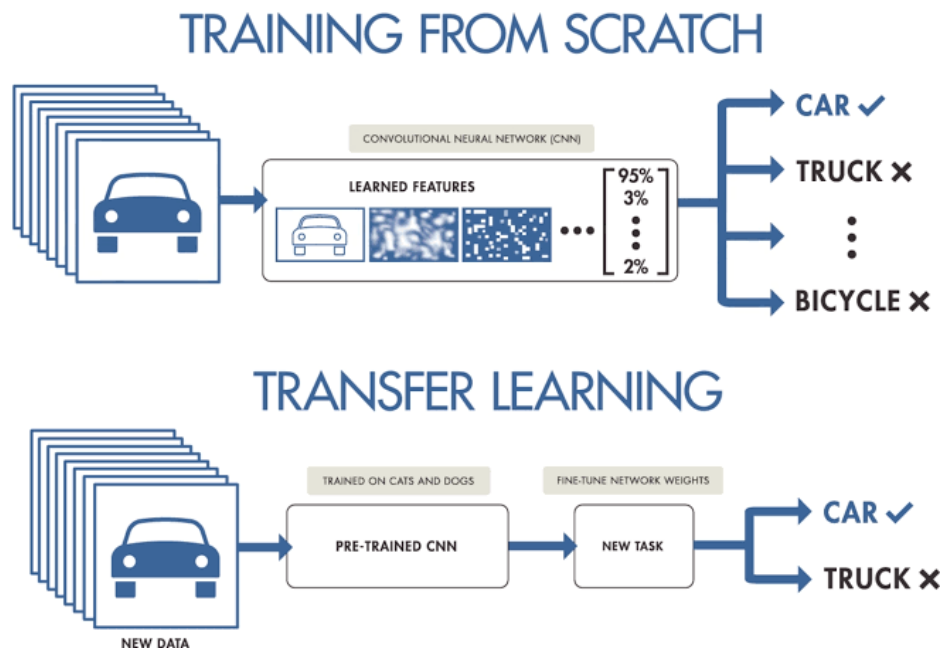


Figura 20 - Transferencia de Aprendizaje (ViewNext, 2020)

El auge en los últimos años de esta metodología se relaciona directamente con el desarrollo de modelos que han demostrado dar resultados por encima de los de los humanos para la clasificación de imágenes. Este hecho ha sido posible por una parte gracias a la cada vez mayor capacidad de hardware y la cantidad de datos disponibles para procesar, pero también gracias al desarrollo de nuevas técnicas que han permitido emplear los recursos disponibles de forma más eficiente. Para su aplicación, la transferencia de aprendizaje requiere de los siguientes pasos:

- ✚ Emplear las capas de un modelo ya entrenado.
- ✚ Establecer ciertas capas como estáticas para evitar que sus pesos no sean alterados durante el nuevo entrenamiento del modelo.
- ✚ Apilar nuevas capas adicionales por encima de las estáticas que puedan aprender a asociar los pesos originales a los nuevos datos.
- ✚ Llevar a cabo el entrenamiento del nuevo modelo con el nuevo conjunto de datos.
- ✚ Adicionalmente, se puede reentrenar de nuevo todo el modelo (incluyendo las capas originales) pero reduciendo considerablemente su ratio de aprendizaje.

La base teórica sobre la que se apoya esta técnica consiste en dejar que las primeras capas de la red neuronal aprenden a reconocer los patrones básicos de la imagen (líneas, por ejemplo). Posteriormente, las capas superiores son las que se encargan de relacionar los distintos patrones para ser capaces de extraer conclusiones a más alto nivel, con lo que una red que haya sido entrenada para clasificar imágenes de forma genérica sirve como primer paso para una red cuyo objetivo sea la clasificación de imágenes más específicas o detalladas. Por otro lado, para decidir el número de capas a descongelar de una red previamente entrenada se suele tener en cuenta el tamaño del conjunto de datos empleado, así como su similitud con los datos para realizar el preentrenamiento del modelo y los nuevos datos con los que se va a entrenar. A mayor similitud entre los conjuntos de datos, menor número de capas será necesario descongelar. Sin embargo, si se dispone de un extenso nuevo conjunto de datos, mayor será el número de capas que se podrán descongelar para que el modelo aprenda características más específicas de sus datos.

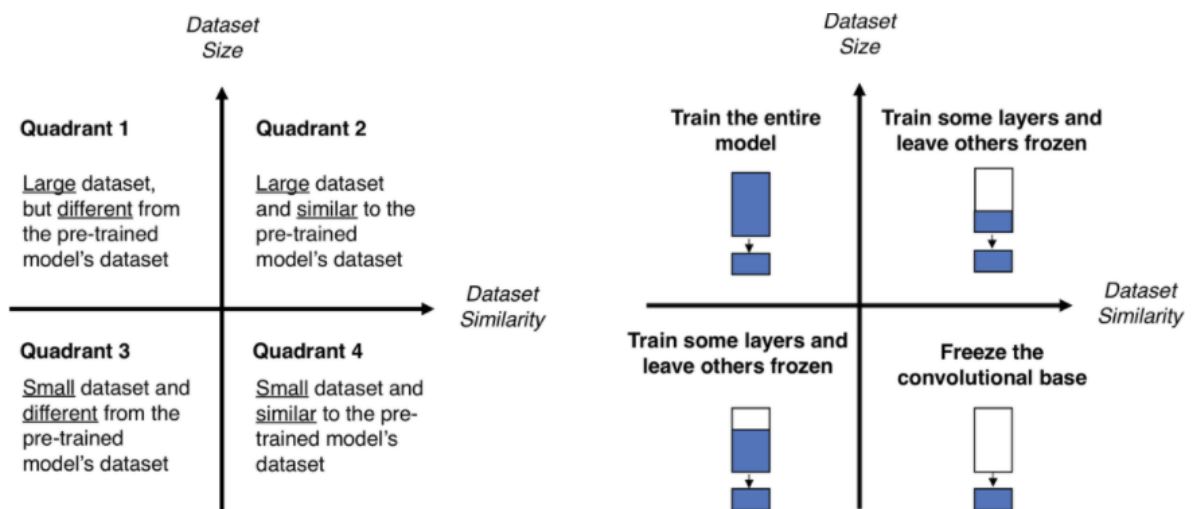


Figura 21 - Criterios para la aplicación de transferencia de aprendizaje (Rahaman, 2020)

En este proyecto, se hace uso de ResNet (en particular ResNet50), una de las Redes Neuronales Profundas más potentes que en su momento consiguió unos resultados de

rendimiento sorprendentes en el reto de clasificación de ILSVRC 2015¹⁰. Esta red ha logrado un excelente rendimiento de generalización en otras tareas de reconocimiento, obteniendo el primer puesto en detección de ImageNet, localización de ImageNet, detección de COCO y segmentación de COCO en las competiciones ILSVRC y COCO 2015 (Russakovsky, Deng, & Su, 2014).

3.6.1 RESNET50

Existen numerosas variantes de la arquitectura ResNet que, aunque se trate del mismo concepto, cada una tiene un número diferente de capas. Desde ResNet-18 hasta ResNet-1202 (el rango más habitual), encontramos una amplia variedad de este tipo de redes pues su nombre seguido de un número de dos o más dígitos simplemente implica la misma arquitectura base ResNet con un determinado número de capas en la red neuronal.

La arquitectura de ResNet50 se compone de cuatro etapas (*Figura 22 - Arquitectura de ResNet50*). La red toma la imagen de entrada con sus parámetros de altura, anchura (múltiplos de 32) y canales (3 para el caso de RGB). A modo de ejemplo, consideraremos el tamaño de entrada de las imágenes de *HAM10000*, fijado en este proyecto a 64x64x3. En primer lugar, se realiza la convolución inicial seguida del *max pooling*, con unos tamaños de núcleo de 7x7 y 3x3 respectivamente. Después, comienza la primera etapa de la red, que tiene 3 bloques residuales con 3 capas cada uno. El tamaño de los núcleos utilizados para realizar la operación de convolución en las 3 capas del bloque de la primera etapa son 64, 64 y 128 respectivamente. Las flechas curvas indican la conexión de identidad, mientras que la flecha discontinua conectada representa que la operación de convolución en el bloque residual se realiza con la segunda zancada (el tamaño de la entrada se reduce a la mitad en términos de altura y anchura, duplicando la anchura del canal). Este proceso se repite a medida que se avanza en las etapas (Boesch, 2022).

¹⁰ ImageNet Large Scale Visual Recognition Challenge o el reto de reconocimiento visual a gran escala de ImageNet evalúa algoritmos de detección de objetos y clasificación de imágenes a gran escala. Un importante incentivo que presenta el proyecto es permitir a los investigadores comparar los avances en la detección de una variedad de objetos más amplia.

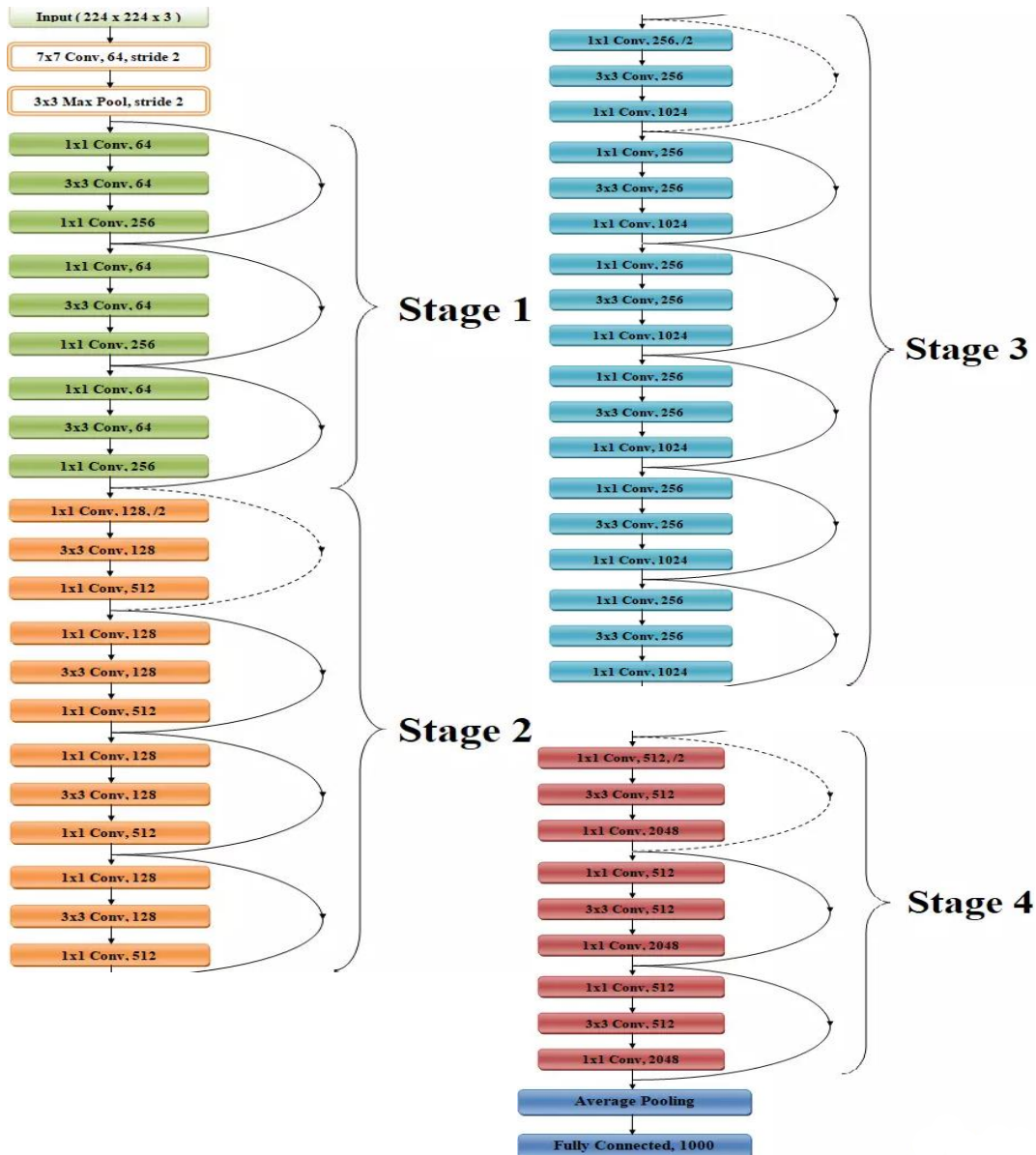


Figura 22 - Arquitectura de ResNet50 (Sachan, 2021)

En las redes más profundas (ResNet101 o superiores) es habitual encontrar un diseño de cuello de botella. A cada función residual F se apilan 3 capas, una sobre la otra. Todas ellas son convoluciones 1×1 , 3×3 y 1×1 : en las capas de convolución 1×1 se reducen primero y luego restauran las dimensiones, y en la capa 3×3 se encuentra el cuello de botella, con dimensiones de entrada y salida inferiores. Por último, la red tiene una capa de agrupación

de medias seguida de una capa totalmente conectada que tiene 1000 neuronas (salida de la clase ImageNet).

3.6.2 IMAGENET

Siendo uno de los datasets más conocidos actualmente, ImageNet es un conjunto de datos que tiene como objetivo participar en el ILSVRC, concurso llevado a cabo a nivel mundial con más de 50 instituciones participantes (desde 2010). En él se presenta el mejor modelo de reconocimiento de objetos, que comprende clasificación (presencia de un objeto en una imagen) y detección (localizar todos los objetos de una imagen). Este evento competitivo se originó con el propósito de poder comparar de forma sencilla distintos modelos para un amplio número de clases, permitiendo así ahorrar a las instituciones participantes el esfuerzo necesario para etiquetar de imágenes.



Figura 23 - Conjunto de datos de ImageNet (Gluon, 2021)

Este dataset (*Figura 23 - Conjunto de datos de ImageNet*) por tanto, es el orquestador en el concurso mencionado y dispone de 1000 clases, con un total de 1.461.406 imágenes.

Debido a su extensión, las imágenes fueron etiquetadas empleando *crowdsourcing*¹¹. Adicionalmente, cada imagen se etiquetó con sinónimos empleando WordNet¹², (permitiendo así evitar problemas de ambigüedades).

3.7 CONCLUSIÓN

La mayoría de los desarrollos citados tienen su fundamento en las Redes de Aprendizaje Profundo para obtener algoritmos de valor útiles para cada funcionalidad. Entre estas destacan las Redes Neuronales Convolucionales (CNN o Convolutional Neural Network en inglés), capaces de tomar una imagen de entrada, asignarle relevancia (según pesos y sesgos) a varios aspectos u objetos de la imagen y poder diferenciar unos de otros. El preprocesamiento requerido en una *ConvNet* es mucho menor en comparación con otros algoritmos de clasificación; mientras que en los métodos primitivos los filtros se diseñan a mano; con suficiente entrenamiento, estas redes están diseñadas para aprender estos filtros y características¹³.

Sin embargo, hay estudios que demuestran la ineficacia de estos conocidos métodos o aplicaciones. Es el caso de un grupo de expertos del Instituto de Investigación Aplicada a la Salud de la Universidad de Birmingham, que, en colaboración con el Centro de Dermatología Basada en la Evidencia de la Universidad de Nottingham, analizaron diferentes estudios realizados durante los últimos años para poner a prueba la exactitud de varias aplicaciones diferentes. Los resultados publicados en el *British Medical Journal* señalan que solo un pequeño número de esos estudios obtienen una conclusión convincente sobre la fiabilidad de los resultados que proporcionan estas aplicaciones para decirte qué tienes en la piel (G.Cid, 2022). Por lo tanto, en este Trabajo Fin de Grado la intención es

¹¹ Práctica por la cual se convoca a un gran número de personas para aportar ideas con el objetivo de resolver una tarea.

¹² Base de datos léxica que agrupa palabras en inglés en conjuntos de sinónimos denominados *synsets*, proporcionando definiciones cortas y generales y almacenando las relaciones semánticas entre los conjuntos de sinónimos (Wikipedia, 2021).

¹³ Curiosamente, la arquitectura de una *ConvNet* es análoga a la del patrón de conectividad que siguen las neuronas en el cerebro humano y se inspira en la organización de la corteza visual que éstas reciben.

asegurar al máximo que las clasificaciones de imágenes son las adecuadas en cada caso, dejando abierta la posibilidad de mantener el algoritmo entrenando de forma recursiva.

Capítulo 4. DEFINICIÓN DEL TRABAJO

4.1 JUSTIFICACIÓN

Al igual que se ha comentado en el primer capítulo de este documento, este proyecto tiene un objetivo claro y necesario que aportar a nuestra sociedad. Por una parte, se sabe que las enfermedades en la piel son una causa importante de la mortalidad en nuestro país y en general en todo el mundo. Actualmente, contamos con desarrollos notables en un mercado extenso como es el de la biomedicina, y que cada vez gana más protagonismo como campo de investigación y desarrollo, pues su impacto en el bienestar de las personas es muy significativo siempre que se obtienen mejoras. Sin embargo, la mayoría de los servicios requieren de una aportación económica necesaria para poder hacer un tratamiento inicial, por básico que sea (realizar una primera exploración de la zona afectada, por ejemplo).

La solución que se propone en este Trabajo Fin de Grado consiste en garantizar el acceso a todas las personas que no puedan disponer de los recursos necesarios para permitirse el primer análisis de cualquier región corporal que presente una mancha que se pueda considerar como sospechosa. Por otra parte también, se facilitará de esta forma cierto trabajo médico a los profesionales especialistas del sector para que así se pueda tomar una decisión con mayor información desde el primer paso del proceso de análisis.

4.2 METODOLOGÍA

De la misma forma que se ha mencionado ligeramente en el primer capítulo de este documento, en este apartado se tratará con más detalle el contenido abarcado en la metodología de trabajo del proyecto.

El camino que se ha usado para conseguir los objetivos consta de una planificación ágil, habiéndose fijado entregas de contenido secuenciales con sus respectivas definiciones previamente. Visto que la estructura de este Trabajo Fin de Grado se compone de la construcción de algoritmos de Inteligencia Artificial por una parte (enfrentando una Red Neuronal Convolutiva *hecha a mano* y una Red Neuronal Residual) y del apartado gráfico del interfaz de la aplicación móvil por otra (donde la combinación de ambas partes se propone como trabajo futuro y se detallará en la sección correspondiente más adelante), se ha tenido que combinar la metodología *Agile* con la *Waterfall*. Este motivo se debe a las entregas de contenido fijadas para los Anexos A, B y la sesión de seguimiento durante el primer cuatrimestre del curso.

En primer lugar y una vez profundizados los conocimientos de Aprendizaje Automático junto con las propuestas disponibles en el mercado, se procedió a estructurar la primera Red Neuronal Convolutiva para tratar los datos¹⁴. Previa a su definición (y en general a cualquier implementación de una red neuronal), fue importante depurar el conjunto de imágenes para conseguir unos resultados adecuados, eliminando imágenes duplicadas y aquella información que no fuera de utilidad para el proyecto como puede ser la edad de los pacientes o la localización de las manchas en el cuerpo (pues no se ha contemplado tener en cuenta su influencia en los resultados finales). Por otra parte, se han añadido columnas necesarias inexistentes en el conjunto inicial, como la asignación binaria para saber a cuál de las clases disponibles pertenece la afección. Acto seguido, se definieron las distintas capas que compondrían la red y sus características (neuronas en cada nivel, funciones de activación en cada capa o el optimizador a emplear en cada caso). Es importante mencionar que, a pesar de contar con una idea y estructura inicial para afrontar el problema, se ha tenido en cuenta en la planificación un periodo para aplicar retoques en aquellos modelos que así lo necesitaran, pudiendo mejorar de esta forma y en la medida de lo posible sus resultados.

¹⁴ Cabe destacar que en ambos enfoques el tratamiento de las imágenes no ha sido idéntico en todos los casos, como se ha mencionado. Este aspecto se desarrollará en profundidad en el capítulo siguiente.

El tiempo de implementación para el segundo enfoque, en el que se ha usado una red neuronal previamente entrenada, ha sido menor a su predecesor puesto que ha requerido invertir menos tiempo modificando parámetros básicos de la red, pues al haber tratado con otros datos y partir de una estructura sólida de base permite empezar con un porcentaje de acierto mayor al de la Red Neuronal Convolutiva *hecha a mano*.

Por último, se reservó un tiempo similar a la clasificación de imágenes para el diseño gráfico de la interfaz propuesta para la aplicación funcional que se propone desarrollar como trabajo futuro en caso de retomar este proyecto, pudiendo también servir de inspiración para otro tipo de implementaciones del mismo campo y quedando sujeta a cambios por parte del desarrollador/a si así lo considera oportuno.

En la siguiente sección se expone la planificación seguida y el tiempo dedicado a cada tarea.

4.3 PLANIFICACIÓN Y ESTIMACIÓN DE COSTES

Para la estimación temporal de las actividades a realizar en este proyecto, se dividió el espacio disponible para las actividades principales que componen el mismo. En primer lugar y aunque se hayan realizado distintas búsquedas durante toda la extensión del trabajo

	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Abril	Mayo	Junio
Estado de la cuestión									
Estudio Aprendizaje Profundo y pruebas									
Construcción de clasificador y evaluación									
Diseño gráfico									
Redacción									

Tabla 2 - Diagrama de GANT del cronograma del proyecto

(en acorde también con la metodología ágil del estudio), se dedicaron los primeros meses a recabar información sobre las varias tecnologías disponibles en el mercado y su impacto a nivel social. La segunda etapa consistió en preparar el terreno para la implementación de los algoritmos basados en Inteligencia Artificial, estudiando con detalle cuáles son sus parámetros más importantes para obtener un correcto funcionamiento y cumplir con el objetivo perseguido en el Trabajo Fin de Grado, permitiendo así clasificar imágenes de forma correcta según el conjunto de datos empleado. Hay que destacar que durante el mes de febrero se tuvo que paralelizar el estudio del Aprendizaje Profundo y las primeras pruebas con las Redes Neuronales Convolucionales con un nuevo estado de la cuestión, pues no se pudo trabajar con el *dataset* que se había planteado inicialmente debido a problemas de confidencialidad con los mismos, adaptando el contenido realizado hasta ~~la~~ entonces (de *dermatitis*) al nuevo área de estudio.

A partir de febrero se ha ido manteniendo en segundo plano la redacción de este documento para poder cumplir con el plazo de entrega, y además mantener un historial de los avances secuenciales del proyecto. Desafortunadamente, la construcción del clasificador de imágenes y su validación ha requerido más tiempo del previsto para llevar a cabo las dos implementaciones planteadas (enfoque con una Red Neuronal Convolutiva y con una Red Neuronal Residual, tratando los datos de dos formas distintas en el primer caso), teniendo así que retrasar ligeramente el desarrollo del diseño del interfaz gráfico sobre el que se propone desplegar la aplicación completamente funcional a modo de trabajo futuro.

En cuanto a la estimación económica del trabajo, el objetivo de la propuesta es ofertar el servicio sin pago alguno, al menos por lo que a los límites de esta propuesta se refiere. Sin embargo, para hacer una estimación del coste de desarrollo se tienen en cuenta los siguientes factores:

- ✚ Precio y uso del ordenador personal.
- ✚ Horas de trabajo y desarrollo.

En cuanto a los recursos materiales usados para el desarrollo de este Trabajo Fin de Grado, se ha empleado un ordenador portátil con las siguientes características (*Tabla 3 - Especificaciones técnicas del ordenador portátil*):

Ordenador portátil HUAWEI MateBook X Pro 2020	
Sistema Operativo	Windows 10 Home
Pantalla	13,9" LTPS (3000x2000) 260 PPI
Procesador	10 th Gen. Intel ® Core ™ i7-10510U Processor
Memoria RAM	LPDDR3 2133 MHz 16 GB
Disco Duro	NVMe PCIe SSD 1 TB
Coste	1.899,00€

Tabla 3 - Especificaciones técnicas del ordenador portátil

Normalmente, se consideran 4 años como vida útil fiscal para un ordenador portátil, lo que equivale a 35040 horas. Sin embargo, se tiene en cuenta para las próximas estimaciones que el ordenador se ha empleado también para otras tareas por lo que su uso no ha sido exclusivamente para este proyecto. En la *Tabla 4 - Horas dedicadas al desarrollo del proyecto* se detalla el tiempo total de uso para el proyecto y a continuación, en las ecuaciones, el coste estimado por horas.

Actividad	Horas
Organización y planificación del proyecto	22
Análisis y estudio de tecnologías	140
Desarrollo y diseño (algoritmos)	380
Evaluación y mejora (algoritmos)	200
Desarrollo y diseño (interfaz)	100
Documentación del contenido	400

Tabla 4 - Horas dedicadas al desarrollo del proyecto

$$\text{Coste por hora} = \frac{1.899,00 \text{ €}}{35040 \text{ horas}} = 0,0542 \text{ €/hora}$$

$$\text{Coste total} = \frac{1242 \text{ horas}}{35040 \text{ horas}} \times 1.899,00\text{€} = 67,31 \text{ €}$$

El conjunto de programas empleados para el trabajo han sido todos de libre distribución por lo que no han supuesto un coste añadido. De esta manera, el coste final de los recursos materiales para este Trabajo Fin de Grado queda por debajo de los 70€.

En cuanto a los recursos humanos para desarrollar este proyecto, se ha contemplado el coste de mercado de un desarrollador de código para la parte de algoritmia de Aprendizaje Profundo y el coste de mercado de un diseñador gráfico para el interfaz de la aplicación móvil, todo ello reflejado en la *Tabla 5 - Recursos humanos empleados en el proyecto*.

Función	Coste (€/hora)	Dedicación (horas)
Data Scientist	21	380
Diseñador Gráfico	25	100

Tabla 5 - Recursos humanos empleados en el proyecto

Con esta información, los honorarios totales del trabajo sumarían un total de 10.480 €, de los cuales 7.980 € corresponden al perfil de Data Scientist y 2.500 € corresponden al perfil de Diseñador Gráfico (Tokio School New Technology, 2022), suponiendo un coste general para todo el Trabajo Fin de Grado de 10.547,31 €.

Capítulo 5. SISTEMA DESARROLLADO

En este capítulo se detallan las características del proyecto llevado a cabo en el trabajo. Se empieza con una descripción de la construcción de las dos redes neuronales empleadas, así como el tratamiento del conjunto de los datos previo a su procesado y el enfoque desarrollado en cada red. Además, se ha aprovechado para reflejar parte de los resultados obtenidos en cada escenario. Por último, se detallan las partes del interfaz gráfico construido, sobre el que se plantea como trabajo futuro implementar el desarrollo algorítmico de este Trabajo Fin de Grado.

5.1 ANÁLISIS DEL DATASET Y ENFOQUES PLANTEADOS

Human Against Machines with 10000 Images es un conjunto de datos necesario para abordar la problemática que plantea este proyecto. Contiene 10015 imágenes dermatoscópicas de zonas corporales afectadas por las lesiones más importantes y peligrosas a día de hoy, sanitariamente hablando. A pesar de todo su contenido, es necesario un análisis del conjunto para evitar, en la medida de lo posible, saturar la capacidad de procesamiento ofertada por los recursos disponibles con los que se trabaja, reduciendo así los costes computacionales que suponen las redes neuronales que se quieren alimentar.

Es cierto que el *dataset* cuenta con muchos los detalles de cada lesión y afectado, pero como se ha mencionado, algunos datos no aportan relevancia en el alcance de este proyecto y serán retirados para reducir los recursos necesarios de procesado. En primer lugar, se han eliminado las imágenes con identificadores de lesión duplicados del conjunto (en total existen 7470 identificadores únicos). Este proceso consiste en retirar los identificadores (*lesion_id*) asignados a imágenes duplicadas, conservando únicamente aquel que corresponde a la imagen original. Se trata de imágenes obtenidas de la misma lesión, pero variando su ángulo, zum u otro parámetro de enfoque que inicialmente no requerimos como

información de partida¹⁵. Por otra parte, no se ha trabajado con la edad o sexo de los pacientes ni con la zona corporal de la localización de las lesiones pues no aporta relevancia en la detección persé de las afecciones, ergo no es necesario su procesamiento¹⁶.

Pasando a los enfoques planteados con los que se han tratado los datos en las redes neuronales, es primero necesario conocer la distribución de la información en el *dataset*. A continuación, se puede apreciar en *Figura 24 - Distribución de las imágenes en el conjunto de datos HAM10000* como existe un claro desequilibrio entre los cuatro grupos de lesiones

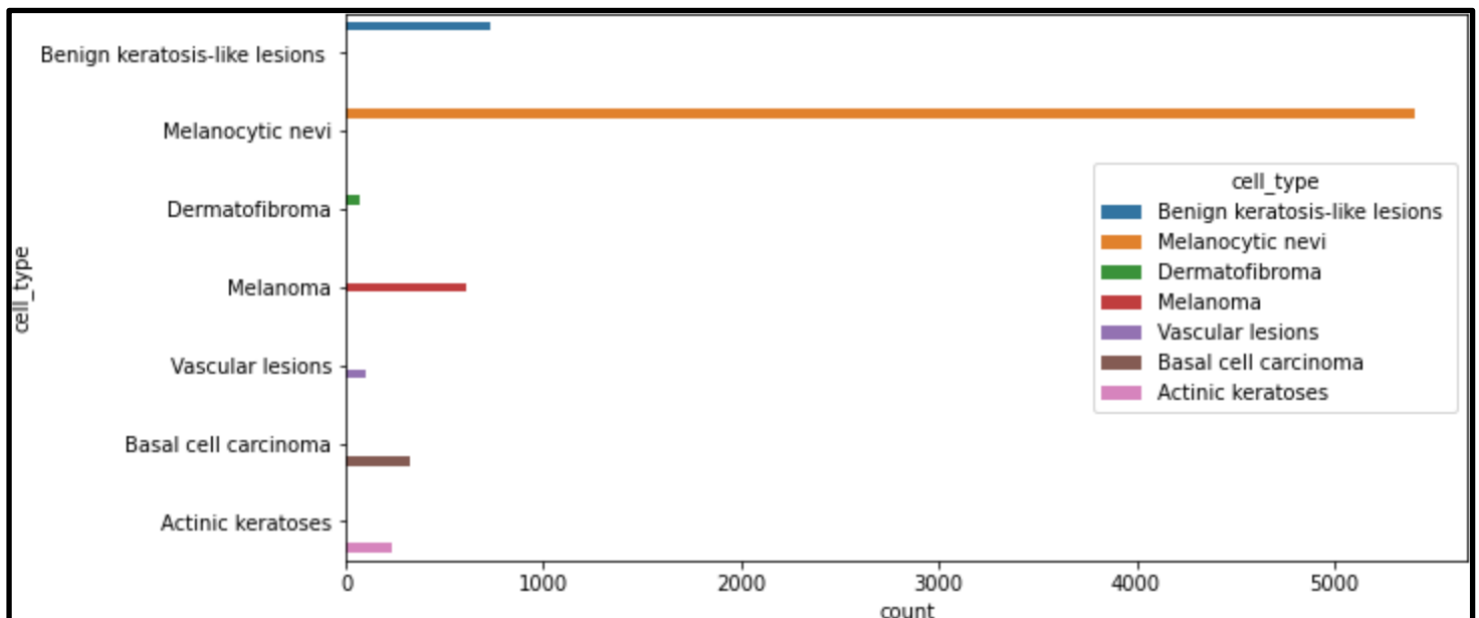


Figura 24 - Distribución de las imágenes en el conjunto de datos *HAM10000*

que podríamos categorizar como:

- ✚ Queratosis actínicas y carcinoma intraepitelial o enfermedad de Bowen (akiec) y Carcinoma de células basales (bcc).
- ✚ Lesiones benignas similares a la queratosis (bkl) y Melanoma (mel).
- ✚ Dermatofibroma (df) y Lesiones vasculares (vasc).

¹⁵ Posteriormente y una vez esté el conjunto se haya filtrado es cuando se valora la posibilidad de aumentar su tamaño, realizando variaciones en las imágenes originales.

¹⁶ A modo de sugerencia, se propone el procesamiento de esta información y su relación con las lesiones.

✚ Nevos melanocíticos (nv).

La alta presencia de imágenes en la clase *Melanocytic nevi (ny)* supone una desventaja para el resto de las clases, cuyo número de imágenes es claramente inferior. Por este motivo, se planteó realizar un primer enfoque que consistiera en establecer un número fijo de imágenes en cada clase, independientemente de las que cada una tuviera. En caso de que la clase tuviera menos imágenes que el valor genérico, se hacen réplicas de las existentes (y pertenecientes a esa misma clase) hasta alcanzar dicho valor. Si, por el contrario, el grupo posee más imágenes, simplemente no se tienen en cuenta las restantes. Para llevar a cabo esta implementación, fue necesario encontrar un equilibrio para, por una parte, establecer un número de imágenes razonable para evitar forzar el sobre aprendizaje en los modelos (pues alimentarlos con consecutivas réplicas implica tratar con imágenes idénticas), y por otra, visto que se ha hecho uso de los recursos ofrecidos por la versión gratuita de *Google Colaboratory*, no saturar la memoria *RAM* ofrecida en la sesión.

Respecto a la validación de la posterior clasificación de imágenes, se ha empleado como verdad absoluta (*ground truth*) la información aportada por el grupo de investigadores que participó en el recopilatorio de las imágenes dermatoscópicas y su tratamiento previo (metadatos disponibles en el *dataset*). Como se puede ver en la *Figura 25 -Distribución de los métodos de validación del conjunto de datos HAM10000*, las lesiones han sido validadas según:

- ✚ Seguimiento (*follow up*): la métrica seguida en este caso consiste en valorar si los nevos controlados (por dermatoscopia digital) muestran algún cambio durante 3 visitas de seguimiento o 1,5 años. De no ser así, se aceptaba como evidencia de benignidad biológica¹⁷.
- ✚ Histopatología (*histo*): diagnóstico histopatológico de las lesiones extirpadas (realizado por dermatólogos especializados).

¹⁷ No se etiquetaron otros diagnósticos benignos además de los nevos con este tipo de verdades básicas porque los dermatólogos no suelen controlar los Dermatofibromas Queratosis Seborreicas o Lesiones vasculares.

- ✚ Consenso (*consensus*): en aquellos casos benignos típicos sin histopatología ni seguimiento, se proporcionó una calificación de consenso de expertos, en caso de que todas las partes involucradas ofrecieran, de forma independiente, el mismo diagnóstico inequívoco de benignidad¹⁸.
- ✚ Confocal (*confocal*): La microscopía confocal de reflectancia es una técnica de imagen in vivo con una resolución a nivel casi celular. Ciertas Queratosis faciales benignas fueron verificadas por este método.

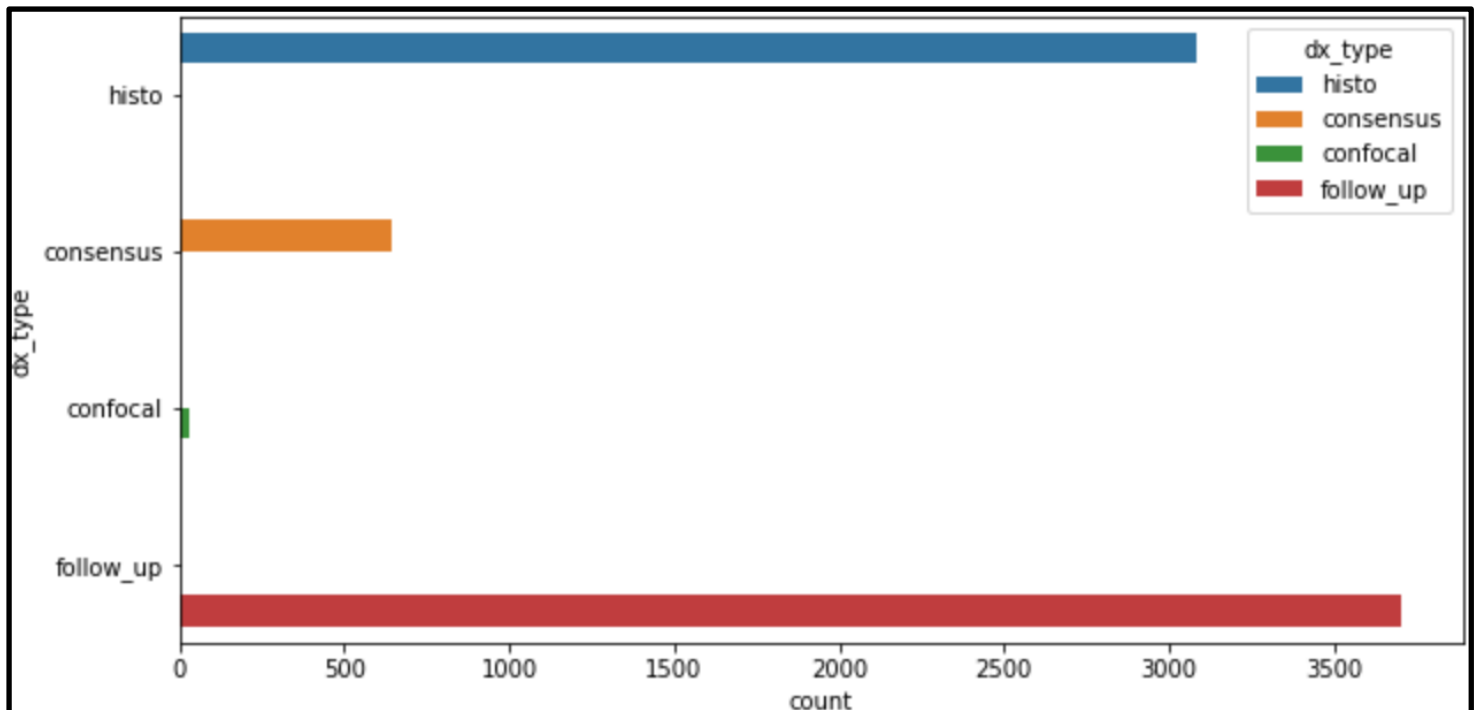


Figura 25 -Distribución de los métodos de validación del conjunto de datos HAM10000

Por último, se planteó un enfoque más realista que consiste en asignar pesos a todas las clases proporcionalmente a las imágenes de cada una y a las totales, es decir, dividir el número de imágenes totales (10015) entre el de cada clase. De esta forma, la clasificación

¹⁸ Las lesiones con este tipo de verdades básicas solían fotografiarse por motivos educativos y no necesitaban un seguimiento posterior o una biopsia para su confirmación (Tschandl, Rosendahl, & Kittler, 2018).

es más susceptible de ser correcta, pues de forma genérica se hace énfasis en aquellas clases que poseen un menor número de imágenes, pero, sin embargo, las implicaciones computacionales superan a las de la metodología previa, pues esta vez se alimenta a las Redes Neuronales con todas las imágenes disponibles en el conjunto de datos.

5.2 DISEÑO DE LAS REDES NEURONALES

A la hora de implementar el desarrollo de los algoritmos basados en Inteligencia Artificial, en ambos casos (CNN y ResNet) se utilizó un 80% de las imágenes totales para llevar a cabo el entrenamiento, dejando el 20% restante para su testeo. Dentro de la etapa de entrenamiento, 87.5% de las imágenes sirvieron de puro *training* (en torno al 70% del total dedicado a entrenamiento) mientras que 12.5% se utilizaron para *validation* (lo que equivale a un 10% del conjunto de entrenamiento).

En los siguientes apartados se detallan las características técnicas de cada red neuronal usada.

5.2.1 MODELO BASADO EN LA RED NEURONAL CONVOLUCIONAL

Como se ha descrito anteriormente, es común contar con imágenes que tengan una alta dimensionalidad, pues para alimentar de forma adecuada la red neuronal, cada píxel se considera una característica intrínseca que se moldea para funcionar con las capacidades comentadas de las CNNs. Además, el modelo también suele mantener un aprendizaje continuo para aquellos nuevos rasgos de mayor dimensionalidad que puedan ir apareciendo según se vaya avanzando en profundidad dentro de la red con cada capa.

Para el trabajo llevado a cabo, se elaboró un modelo CNN haciendo uso de la API secuencial que ofrece *Keras*, implementándolo también como base para la siguiente red neuronal (ResNet) que trataremos en el apartado posterior.

Empezando por la capa más superficial, su composición consta de una convolución de 2 dimensiones (Conv2D). Esta implementación permite asegurar una convolución

espacial en dos ejes sobre las imágenes que se procesan, es decir, operar sobre los ejes definidos en los datos. Los parámetros fijados para esta capa fueron:

- ✚ Número de filtros (*filters*): 32.
- ✚ Tamaño del filtrado (*kernel size*): 3x3.
- ✚ Función de activación (*activation*): ReLu o Unidad Lineal Rectificada, es decir, asignación de valores lineales para todos los valores positivos, y en caso de recibir valores negativos, estos se anulan (se asignan a cero).
- ✚ Relleno por defecto (*padding*): idéntico (*same*), para obtener un tamaño de salida acorde con el de la entrada. En este caso, con un filtro de 3x3 obtenemos $p=(3-1)/2=1$.
- ✚ Forma del contenido de entrada (*input shape*): al tratarse de las imágenes del conjunto de datos, se indica su alto, ancho y por último su dimensión que, como son imágenes a color, es tres (RGB).

Después de esta primera operación procedemos a aplicar MaxPool2D que elimina la información de la señal y actúa como un filtro de muestreo reducido¹⁹. Se centra en píxeles cercanos (vecinos) para elegir el valor máximo entre éstos y reducir así el sobreajuste. Su parámetro en este caso resulta ser:

- ✚ Tamaño de mezcla (*pool size*): 2x2.

Acto seguido, se aplica la operación de bloque central de Dropout, a modo de técnica de regularización, para ignorar aquellas neuronas seleccionadas de forma arbitraria en la etapa de entrenamiento del algoritmo y fijar sus valores a cero. De esta manera, se está obligando a la red neuronal a aprender las características de forma más distribuida, permitiendo mejorar la generalización y el sobreajuste que pueda aparecer durante la etapa de entreno. En este caso, una quinta parte de las neuronas son ignoradas durante el *training*.

¹⁹ Permite extraer las muestras sobre un conjunto reducido, disminuyendo en cierta medida las necesidades y costes computacionales.

En la cuarta capa encontramos una nueva Conv2D para ofrecer mayor flexibilidad al modelo y para expresar transformaciones no lineales, reduciendo al máximo las pérdidas de información y facilitando de esta forma el aprendizaje. Esta vez, el número de filtros empleados se duplica respecto a la primera implementación de esta técnica, y los aplicativos que le siguen permanecen en la línea de los pasos mencionados previamente. En otras palabras, se aplica el mismo tipo de operaciones (aumentando el número de filtros utilizados) exceptuando los últimos tres puntos. En el antepenúltimo paso, esta vez el *pooling* no considera el valor máximo de los píxeles de entrada, sino que se centra en la media de los datos (manteniendo un tamaño de mezcla de 2x2) para aplicar el filtro de muestreo reducido, seguido de una capa de bloque central de aplanamiento (*flatten*). Su funcionalidad justamente permite aplanar los valores de entrada en un vector único de una sola dimensión.

Finalmente, la última capa de la red consiste en una capa densa (*dense*), en otras palabras, una capa regular completamente conectada en una red neuronal. El parámetro más importante en este caso es el número de clases objetivo que se quieren clasificar, y como bien se ha descrito en la definición del proyecto, se tratan de siete categorías (correspondientes a las siete lesiones del conjunto de datos). Sin embargo, la función de activación también tiene relevancia en el aplicativo; de forma predeterminada, se aplica como función de activación en esta capa una *SoftMax*, que como ya se ha dicho toma como entrada un vector de números reales para normalizarlo en una distribución de probabilidad. Esta distribución consiste en probabilidades (las mismas que números reales) proporcionales a la exponencial de los números que hayan servido como *input*²⁰. En nuestro preciso caso, es una funcionalidad muy acorde con la codificación de los datos para obtener las siete clases de categorías de lesiones cutáneas fijadas.

Con todos estos detalles, la foto del resumen de la implementación del primer modelo de Inteligencia Artificial desarrollado es la siguiente:

²⁰ Se realiza un mapeado del vector a la probabilidad de una salida dada en la clasificación binaria fijada.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
dropout (Dropout)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
conv2d_2 (Conv2D)	(None, 12, 12, 128)	73856
average_pooling2d (AveragePooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 7)	32263
=====		
Total params: 125,511		
Trainable params: 125,511		
Non-trainable params: 0		

Figura 26 - Resumen del modelo de la CNN

Una vez generado el modelo, los hiper parámetros configurados para su compilación consistieron en definir los siguientes aspectos, fijados de forma bastante genérica²¹ para conocer qué respuesta daba el algoritmo, y en función de esta, aplicar nuevos cambios:

- ✚ Tamaño del lote (entrenamiento y validación, batch size): 64.
- ✚ Optimizador (optimizer): Adam (con su tasa de aprendizaje, *learning_rate*, por defecto).
- ✚ Función de pérdida (loss function): Entropía categórica cruzada (*categorical_crossentropy*).
- ✚ Número de épocas (epochs): 30.

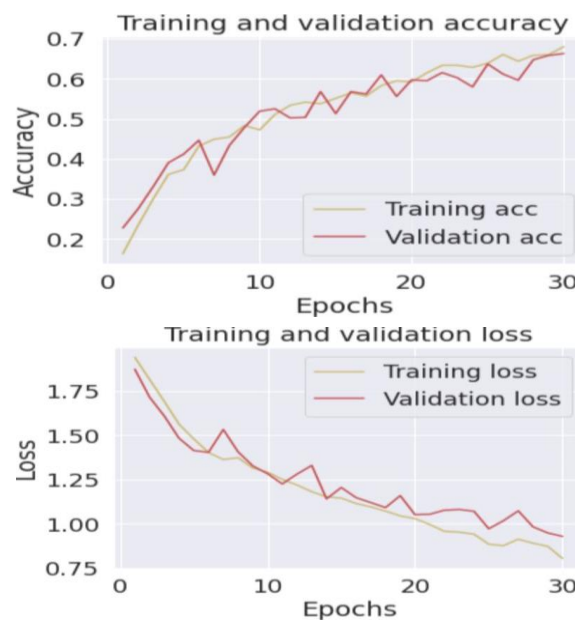


Figura 27 - Resultados del modelo basado en CNN con 30 épocas

La primera ejecución llevada a cabo con esta CNN presentó buenos resultados. En la *Figura 27 - Resultados del modelo basado en CNN con 30 épocas*, se puede apreciar como las curvas obtenidas, tanto para pérdidas como para acierto suponen un primer contacto que

²¹ Para realizar un mayor número de pruebas y un análisis de los resultados riguroso, no se realizó una búsqueda exhaustiva de los hiper parámetros. Se eligieron los valores típicos que han demostrado aportar mejores resultados.

se puede considerar como satisfactorio. Centrándonos más en la representación del acierto, por una parte, los valores más tempranos indican el comienzo del aprendizaje (en torno al 20%), con una validación por encima del entrenamiento, mientras que, por otra parte, a medida que avanzamos en el tiempo, la mejora se puede considerar prácticamente exponencial, quedando el valor de la última época ligeramente por debajo del 70%. Respecto a la curva de pérdida, el decremento es considerable, llegando a reducirse hasta un punto con tan solo 30 épocas.

Con este primer intento superado, se quiso empujar el modelo más allá para obtener un valor que superara la cota del 70% de acierto, pues se quería que el acierto para el aplicativo desarrollado fuera lo más alto posible, respetando los criterios de sobre aprendizaje. Para ello, se aumentó el número de épocas a un valor más estandarizado (100), lo que permitió lograr el incremento deseado:

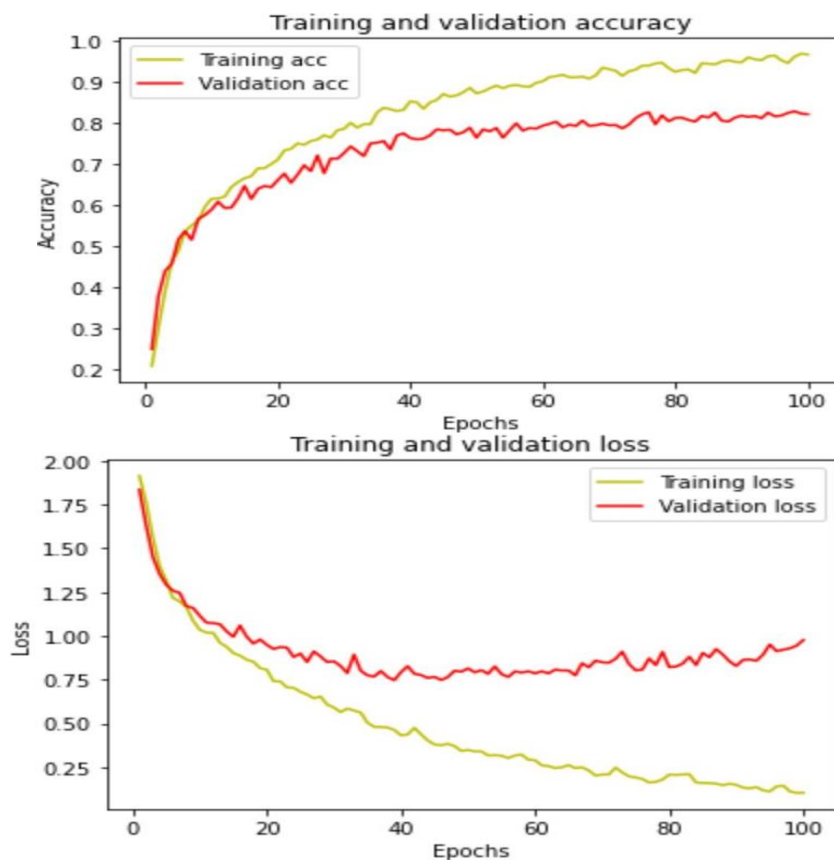


Figura 28 - Resultados del modelo basado en CNN con 100 épocas

Entonces, con la primera implementación por encima del 80% (en concreto 81,79%), se decidió practicar con la segunda implementación planteada, habiendo mejorado ya esta primera. Por un lado, en este primer caso, se fijó de forma manual el número de imágenes en cada clase a 700²², obteniendo un total de 4900 imágenes que procesar, replicando imágenes en aquellas clases cuyo valor total quedara por debajo del fijado y truncando las imágenes sobrantes en aquellas clases que lo superaban²³. Es por eso por lo que se aprovechó para redimensionar el tamaño de las imágenes, reduciendo las necesidades de procesamiento

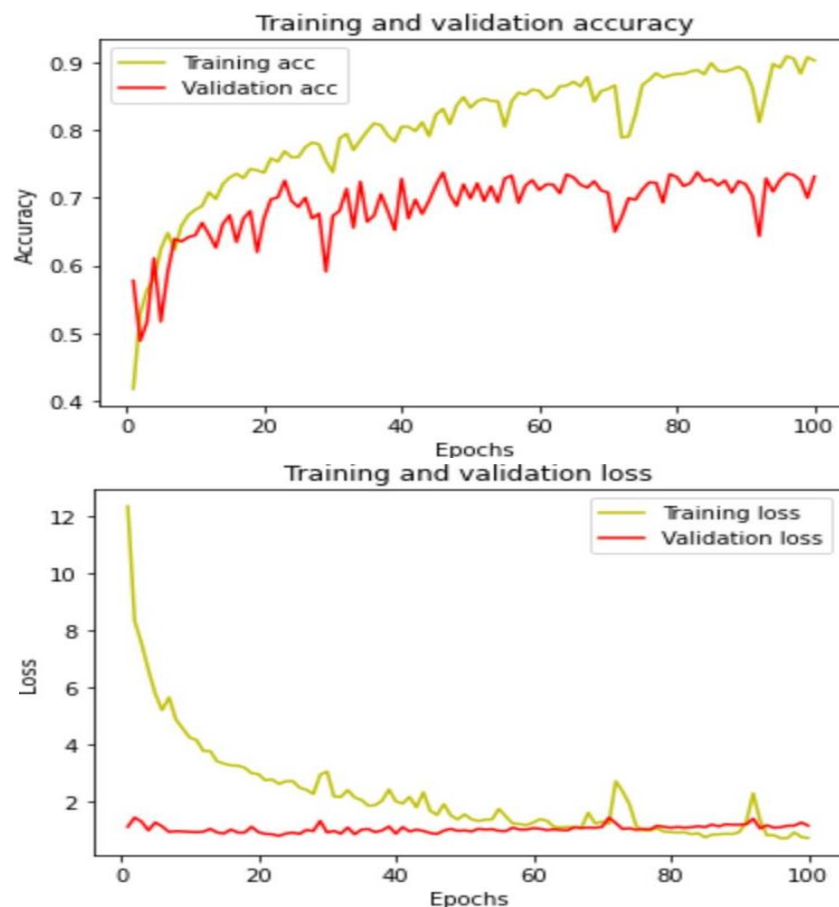


Figura 29 - Resultados del modelo basado en CNN con 100 épocas y todas las imágenes disponibles en el conjunto de datos

²² Se estableció el valor tras evaluar los resultados con incrementos en rangos de 100 imágenes, encontrando un equilibrio entre coste computacional y recursos disponibles en 700 imágenes.

²³ Como se menciona con más detalle en la respectiva sección, tener que replicar imágenes idénticas en una misma clase puede desencadenar en resultados falsamente positivos, pudiendo resultar en un porcentaje de acierto muy elevado (por encima del 90%) pero no tan sólido, por ejemplo.

de los recursos disponibles de la versión de *Google Colaboratory* con la que se trabajó (sacrificando calidad y resolución de imagen por otra parte) y, lo más importante, pudiendo trabajar con toda la información disponible en el conjunto de datos.

En la *Figura 29 - Resultados del modelo basado en CNN con 100 épocas y todas las imágenes disponibles en el conjunto de datos* podemos apreciar el resultado obtenido. Se puede ver a la vista como, habiendo aumentado el número de épocas, el valor del porcentaje de acierto durante el entrenamiento se dispara a medida que se avanza en el tiempo y, sin embargo, la validación queda por debajo. A pesar de ello, esta constatación no resulta preocupante, pues si bien con 4900 imágenes se superó el umbral del 80% de acierto, aunque con todo el conjunto de datos se esté ligeramente por debajo (precisamente 73,12%), se puede afirmar que en este último caso el algoritmo clasifica mejor²⁴.

5.2.2 MODELO BASADO EN LA RED NEURONAL RESIDUAL

Para el aplicativo de este apartado, se ha hecho uso de la Red Neuronal Residual ResNet50, que como se mencionó en el respectivo capítulo, consta de 50 capas como bien indica su nombre (He, Zhang, Ren, & Sun, 2015).

En primer lugar, se definió el modelo base tomando la arquitectura de ResNet50, previamente entrenado con los pesos de ImageNet e indicando el tamaño de los datos de entrada definido. Después de este punto, se añadieron capas adicionales a la red permitiendo alinear los números de los parámetros de entrada y las salidas del algoritmo. En concreto, se realizaron las siguientes operaciones una vez se cargó el modelo base, en el orden que procede:

- ✚ Convolución de dos dimensiones (con 64 filtros, tamaño de filtrado de 3x3, función de activación ReLu o Unidad Lineal Rectificada y relleno por defecto).

²⁴ Al haberse tratado con todos los datos, el algoritmo sabe clasificar mejor que no en el caso previo, pues había sido entrenado con prácticamente la mitad de los datos (donde parte de éstos eran idénticos).

- ✚ Mezcla de dos dimensiones también según el criterio máximo (MaxPool2D), con *pool_size* de 2x2.
- ✚ Dropout del 40% de las neuronas de la red, forzando al algoritmo a aprender de forma más distribuida, permitiendo mejorar la generalización y el posible sobreajuste que pudiera haber durante la etapa de entreno.
- ✚ Aplanamiento (*flatten*) permitiendo transformar los valores de entrada en un único vector de una dimensión.
- ✚ Una capa densa con 128 neuronas y, de nuevo, ReLu como función de activación.
- ✚ Dropout del 40% de las nuevas neuronas.
- ✚ Capa densa con el número de clases objetivo a clasificar y la función de activación *softmax*.

Con las capas posteriores a ResNet50 establecidas, se fijaron los hiper parámetros adicionales como el optimizador Adam (con tasa de aprendizaje fijada a 0,001 pero variable durante el entrenamiento) y la función de pérdida *categorical_crossentropy* (permitiendo la categorización en una sola etiqueta)²⁵. De esta forma, la compilación del modelo resultó tal y como se presenta en la *Figura 30 - Resumen del modelo de ResNet50*.

²⁵ Suele ocurrir cuando sólo una categoría es aplicable a cada conjunto de los datos procesados. En otras palabras, un ejemplo sólo puede pertenecer a una clase, como es el caso para el clasificador de este proyecto.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
resnet152 (Functional)	(None, 5, 4, 2048)	58370944
conv2d_3 (Conv2D)	(None, 5, 4, 64)	1179712
max_pooling2d_2 (MaxPooling 2D)	(None, 2, 2, 64)	0
dropout_2 (Dropout)	(None, 2, 2, 64)	0
flatten_1 (Flatten)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_3 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 7)	903
=====		
Total params: 59,584,455		
Trainable params: 59,433,031		
Non-trainable params: 151,424		

Figura 30 - Resumen del modelo de ResNet50

Pasando a la etapa de entrenamiento, la novedad que se implementó en este modelo fue que se alimentó en cada etapa con datos aumentados, es decir, realizando modificaciones sobre el conjunto de datos original para reforzar el aprendizaje del algoritmo. Una forma de tratar los datos desequilibrados como en nuestro caso es el sobre muestreo aleatorio²⁶. Sin embargo, los investigadores llegaron a la conclusión de que ésta no es la mejor práctica, pues la distribución de las clases cambiaría significativamente: las clases más pequeñas serían

²⁶ Muestrear de nuevo las muestras menos frecuentes para ajustar su cantidad en comparación con las muestras predominantes.

mucho menos variables que las clases más grandes y constarían de muchas imágenes similares con pequeñas transformaciones afines. De esta forma, se decidió luchar contra el desequilibrio fijando los pesos correspondientes a cada clase para que el modelo fuera más sensible a los desequilibrios naturales presentes en el conjunto de datos²⁷. Por lo tanto, se realizaron:

- ✚ Rotaciones aleatorias del 20%.
- ✚ Desplazamientos horizontales aleatorios del 20%.
- ✚ Desplazamientos verticales aleatorios del 20%.
- ✚ Desplazamientos aleatorios del 10%.
- ✚ Ampliaciones aleatorias del 10%.

En este caso se definió el número de épocas a 50 en primera instancia, un tamaño en cada pasada de 128 imágenes y una tasa de aprendizaje basada en *ReduceLROnPlateau* y permitiendo disminuir la tasa de aprendizaje durante el entrenamiento, partiendo del optimizador Adam. Este método permite alcanzar el mínimo global de una función de pérdida de forma más eficiente, pues como se ha mencionado anteriormente en la función del optimizador, la tasa de aprendizaje comienza en 0,001, y disminuye en un factor de 0,5 si la precisión (*val_acc*) no mejora después de 3 épocas (elemento de *patience*).

Mediante estos detalles, la respuesta del modelo fue muy satisfactoria, a pesar de que previamente se tuvo que ajustar la tasa de aprendizaje probando diferentes valores y el número de imágenes en cada pasada, para no saturar los recursos computacionales. Igualmente, mediante la *Figura 31 - Árbol de decisión en el entrenamiento de una* , se analizó cada resultado para evaluar qué opción resultaba más conveniente.

Las posibilidades planteadas en el árbol de decisión son las siguientes:

²⁷ A pesar de emplear esta técnica, el mejor método consiste en recoger más datos.

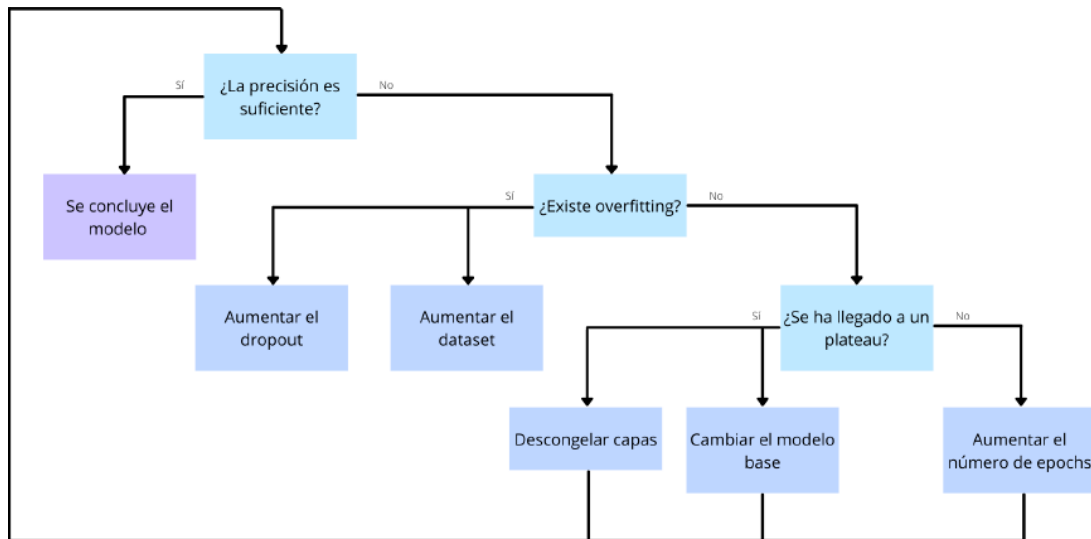


Figura 31 - Árbol de decisión en el entrenamiento de una red neuronal

- ✚ Incrementar el número de épocas: en principio es recomendable usar un número relativamente alto (entre 30 y 70) aunque es cierto que, si el modelo llega a una zona de estancamiento, aumentar el número de *epochs* no implica mejorar los resultados obtenidos.
- ✚ Incrementar el dropout: con un sobre aprendizaje cuasi inexistente, no resulta muy efectivo el cambio.
- ✚ Sustituir el modelo base: en esta escenario, un cambio de este calibre puede conllevar un incremento en el rendimiento. Para el caso que nos ocupa, se puede implementar una arquitectura parecida a la ResNet u otra CNN distinta, como VGG16, una ResNet con más niveles, o incluso un planteamiento distinto.
- ✚ Incrementar el tamaño del conjunto de datos: de forma generalizada, incrementar el conjunto de datos siempre suele estar acompañado de mejoras en el rendimiento global del modelo. A pesar de este hecho, no siempre resulta viable hacerlo pues para este proyecto, por ejemplo, tan solo se ha podido contado con 10015 imágenes²⁸ (de las cuales cerca de 2500 estaban repetidas). Por otra parte, un mayor número de datos

²⁸ Otra alternativa es otorgar un mayor porcentaje al conjunto de entrenamiento sin reducir el conjunto de validación, aunque esto podría provocar que el modelo desarrollara sobre aprendizaje.

que procesar implica también incrementar el tiempo de ejecución y los recursos computacionales necesarios para el procesado, limitados en la sesión empleada de *Google Colaboratory*.

- + Descongelar capas del modelo base: de esta manera el conjunto de datos con el que se entrena originalmente la red se puede adaptar mejor a los nuevos datos. En caso de encontrarnos frente a un estancamiento (en esa situación, es difícil que el modelo siga aprendiendo), descongelar alguna de las capas de la red original puede incrementar la precisión.

5.3 DISEÑO DEL INTERFAZ GRÁFICO

El alcance de este Trabajo Fin de Grado, además de los algoritmos desarrollados para la clasificación de siete afecciones en la piel, incluye el diseño del interfaz gráfico de la aplicación móvil, capaz de analizar imágenes tomadas con la cámara del dispositivo (haciendo uso de la linterna de este si fuera necesario), de manera que en el futuro sea posible desarrollar dicha aplicación. Esta combinación permite disponer de un entorno de trabajo idóneo para la identificación de lesiones cutáneas, tal y como se planteó en el inicio del proyecto.

A la vista de los resultados obtenidos, la creación de una aplicación básica que clasifique estas lesiones y pueda identificar las más peligrosas dentro del grupo resulta viable, pudiéndose seguir la propuesta planteada en *Figura 32 - Propuesta del interfaz gráfico para la aplicación de clasificación de lesiones cutáneas Skin App*. Sin embargo, como se ha explicado antes, el conjunto de datos con el que se han entrenado los modelos resulta en cierto modo limitante, por lo que se considera que será necesario mantener los algoritmos alimentados de forma recurrente. Con este objetivo en mente, se sugiere aprovechar la propuesta desarrollada para, junto al clasificador de lesiones, extraer la información destacable de cada imagen.

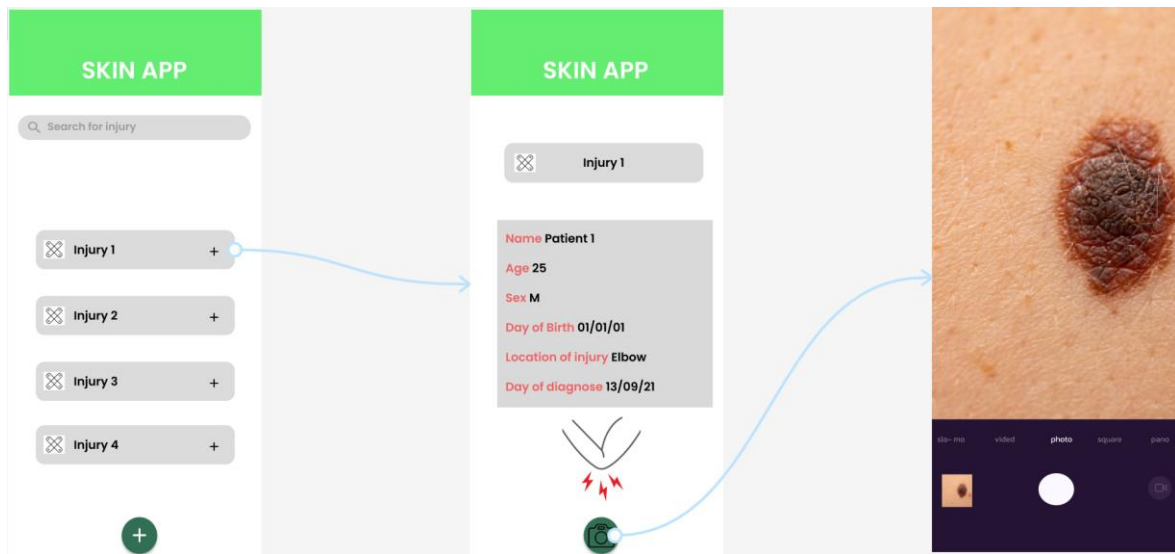


Figura 32 - Propuesta del interfaz gráfico para la aplicación de clasificación de lesiones cutáneas *Skin App*

Capítulo 6. ANÁLISIS DE RESULTADOS

A pesar de haber desvelado parte de los resultados en el capítulo anterior, en esta sección se recalcarán los más significativos del proyecto, haciéndose un análisis crítico de los mismos.

Después de las pruebas realizadas con cada implementación, de forma general se puede afirmar que los modelos desarrollados tienen una mejor respuesta cuando se trabaja con todo el conjunto de datos, todo ello conllevando un mayor coste computacional a su vez. Si bien en el primer enfoque se ha conseguido superar el umbral del 80% de acierto, los resultados de la clasificación no terminan de ser convincentes para todas las clases. La *Figura 33 - Error cometido por el modelo CNN desarrollado en el primer enfoque* da fe de ello: se puede ver que, en las clases correspondientes a Queratosis Actínica, Dermatofibroma y Melanoma (etiquetadas como 0, 3 y 4 respectivamente), el error cometido a la hora de clasificar llega a superar el 50%, lo que implica que el algoritmo se equivoca en torno a la mitad de las veces.

Claramente, el haber establecido el mismo número de imágenes en todas las clases ha desfavorecido aquellas cuyo valor quedaba por debajo, y esto queda reflejado en los resultados de clasificación. La explicación para este hecho reside en el sobre aprendizaje: en el 57,14% de los datos totales empleados no ha sido necesario replicar imágenes, pero en el 42,86% restante sí, lo que ha supuesto que en tres de las siete clases de nuestro conjunto de datos nos encontráramos con un error de clasificación significativo, pues el porcentaje de datos es considerablemente alto. Justamente es en estas tres clases en la que las imágenes totales en cada una, previo al proceso de replicado, es bastante bajo, y por ello, el sobre aprendizaje y la incorrecta clasificación en esas clases destaca.

Se estableció el valor a 700 imágenes por clase después de varias etapas de prueba, tanto con los medios empleados para la computación de los datos y la generación de los modelos (*Google Colaboratory*), como con las respuestas obtenidas, quedando este valor como equilibrio entre ambas necesidades.

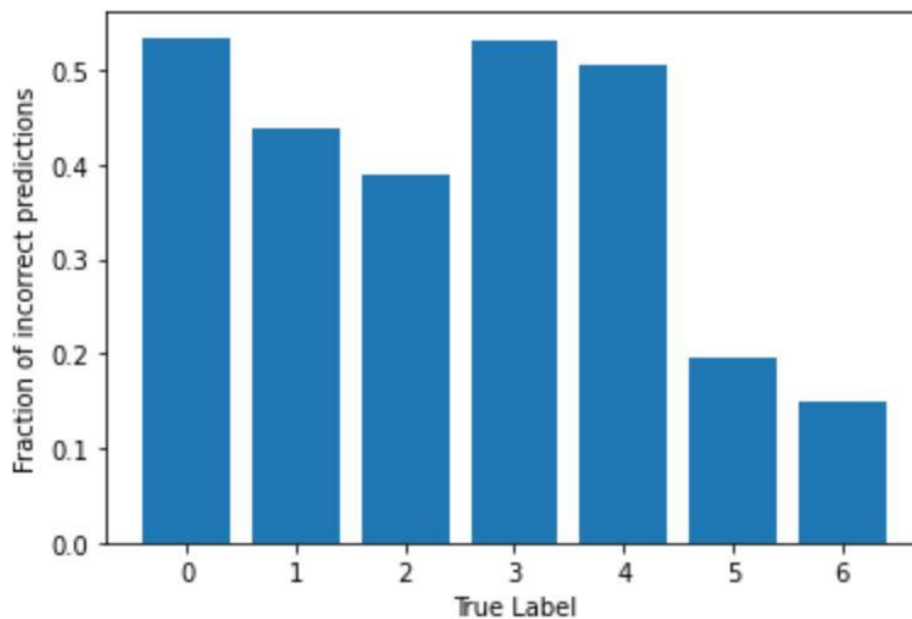


Figura 33 - Error cometido por el modelo CNN desarrollado en el primer enfoque

Por otra parte, en el segundo enfoque utilizado, se han empleado todos los datos disponibles en el conjunto de datos con el que se ha trabajado. Esta vez, el no haber tenido que alimentar el modelo con replicas exactas de las imágenes de varias clases ha implicado reducir ligeramente el acierto del modelo hasta un 73,12%, pero con un error muy poco significativo. Tan solo se comete un 30% de fallo en las clases Carcinoma de células basales, Queratosis benigna, Melanoma y Nevos melanocíticos (etiquetadas como 1, 2, 4 y 5 respectivamente).

La *Figura 34 - Error cometido por el modelo CNN desarrollado en el segundo enfoque* permite, además, ver como el error de clasificación en las clases Dermatofibroma y Lesiones vasculares (etiquetadas como 3 y 6 respectivamente) se ha reducido significativamente, destacando que ambas clases juntas suman un 2,56% de los datos totales

(las más pequeñas). En este punto es importante recordar que esto ha sido posible precisamente debido a la distribución de pesos seguida para el entrenamiento del modelo (segundo enfoque), según las imágenes presentes en cada clase y las totales en el conjunto.

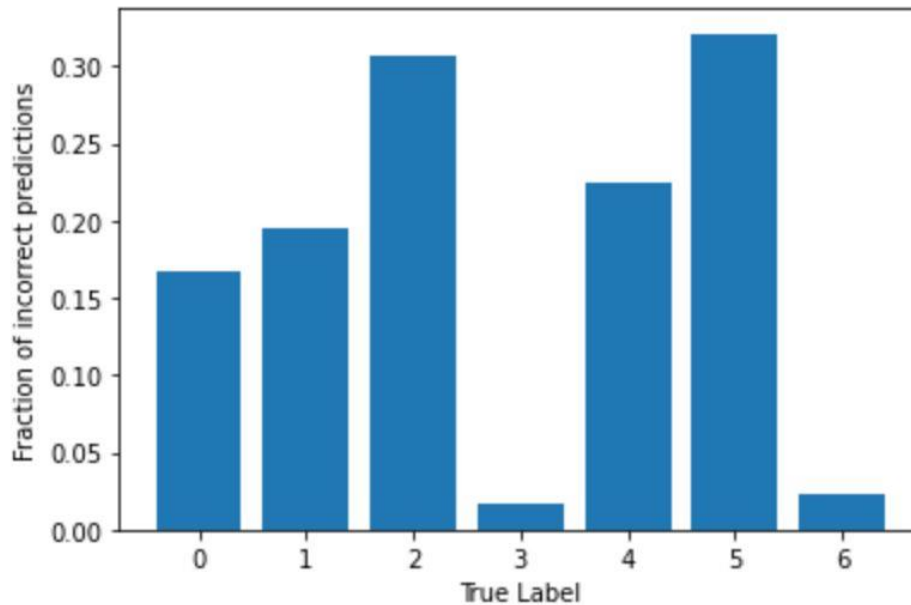


Figura 34 - Error cometido por el modelo CNN desarrollado en el segundo enfoque

Vistas las respuestas del modelo basado en una Red Neuronal Convolutiva ante los dos enfoques, con la Red Neuronal Residual se decidió procesar directamente todo el conjunto de datos, habiendo realizado previamente el filtrado necesario de las muestras duplicadas, al igual que en el caso anterior. Adicionalmente, esta vez se aplicaron también modificaciones en las imágenes (*data augmentation*) para reforzar el aprendizaje de la red, como se ha mencionado al final del capítulo anterior. De esta manera, se han intentado anticipar los resultados esperados (en principio similares a los de la *Figura 34 - Error cometido por el modelo CNN desarrollado en el segundo enfoque*), para así partir de un acierto mayor y con un menor error de clasificación.

Como refleja la *Figura 35 - Error cometido por el modelo ResNet desarrollado*, las operaciones realizadas en las imágenes (rotaciones horizontales y verticales, ampliaciones o cambios de luminosidad) permitieron obtener unos valores muy gratificantes, empezando

con un acierto temprano por encima del 70% hasta alcanzar un 80,6%. En cuanto a la tasa de error, esta fue prácticamente inexistente en todas las clases por los motivos previamente mencionados.

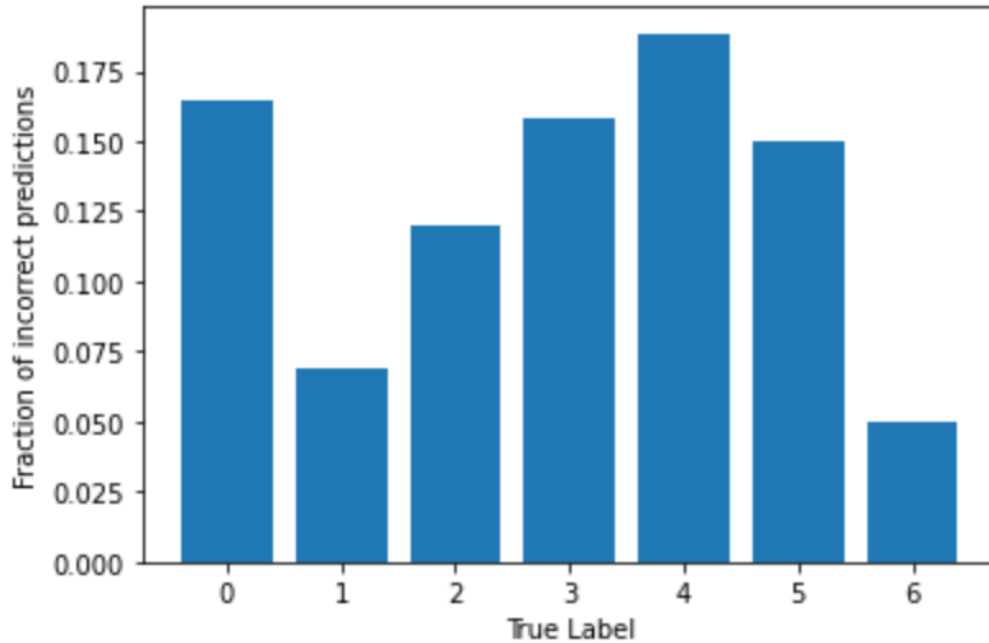


Figura 35 - Error cometido por el modelo ResNet desarrollado

Resulta evidente, a la vista de los resultados obtenidos, que cuando no se dispone de los datos suficientes para trabajar con modelos de Aprendizaje Profundo (para clasificación de imágenes en este caso), los datos aumentados aportan un valor considerable para mejorar el rendimiento del modelo, como se acaba de demostrar.

Capítulo 7. CONCLUSIONES Y TRABAJOS FUTUROS

A lo largo de este Trabajo Fin de Grado se han planteado una serie de modelos con diferentes arquitecturas y se han comparado entre sí, concluyendo en la construcción del clasificador de lesiones en la piel. Además, se ha diseñado un interfaz gráfico que puede servir de base para el desarrollo de una aplicación móvil completamente funcional de esta temática. Por lo tanto, haciendo un recapitulativo de los objetivos definidos en el alcance del proyecto, se establecieron los siguientes aspectos:

- ✚ La creación del entorno de trabajo y del clasificador de imágenes dermatoscópicas.
- ✚ La implementación de un enfoque secuencial a partir de una Red Neuronal Convolutiva y una Red Neuronal Residual, para su posterior comparación.
- ✚ El diseño gráfico del interfaz para el entorno de trabajo.

Por encima de este listado, quedan los objetivos sociales que pretende cumplir el desarrollo. En primer lugar, la detección temprana del cáncer de piel para su posible tratamiento, pues como se ha mencionado, en caso contrario resulta incurable (Sanofi, 2020). Por otro lado, facilitar la accesibilidad a un diagnóstico temprano, que, en muchas ocasiones, supone un coste importante para los pacientes y es difícil de asumir según cada entorno.

7.1 MODELOS DE APRENDIZAJE PROFUNDO DESARROLLADOS

El pilar más importante para la creación del entorno de trabajo y que se ha realizado en este proyecto ha consistido en el desarrollo de los dos modelos de Aprendizaje Profundo mencionados en el segundo punto del listado anterior, cada uno con su correspondiente implementación y todos ellos siguiendo un enfoque secuencial.

De esta forma, como se ha venido comentando a lo largo de este documento, resultará útil en caso de llevar a cabo el desarrollo de la aplicación móvil completamente funcional

un suministro más extenso de imágenes que permitan al algoritmo reforzar su potencial. Se marca como pendiente esta tarea para la futura implementación.

Como era de esperar, la implementación de la Red Neuronal Residual (previamente entrenada) ha supuesto una mejora en la precisión del respectivo modelo, visto que se parte de una estructura más robusta (sólo se modifican ciertos niveles de la red, típicamente los últimos y menos profundos) que en el caso del modelo de la Red Neuronal Convolutiva. Haber aplicado *Transfer Learning* ha sido el punto que más ha mejorado los resultados del proyecto. Adicionalmente, es bienvenida la introducción de nuevas imágenes en el conjunto de entrenamiento que sigan la misma estructura del conjunto de datos empleado, con el fin de robustecer todavía más la precisión de las arquitecturas desarrolladas.

En la siguiente tabla, se puede ver el resumen general de los resultados de los modelos desarrollados:

<i>Modelo</i>	<i>Precisión</i>
CNN (primer enfoque)	81,79%
CNN (segundo enfoque)	73,12%
ResNet50	80,2%

Tabla 6 - Resultados obtenidos con los modelos desarrollados

7.2 MEJORAS GENERALES

En la *Figura 36 - Visión general de las lesiones en el conjunto de datos HAM10000* se puede ver el aspecto que presentan las imágenes del conjunto de datos usado. Como se aprecia en las distintas clases, todas ellas tienen fotografías dermatoscópicas que fueron tomadas en pacientes con un color de piel muy parecido. Se puede decir en cierta medida entonces que, los resultados obtenidos en todos los modelos quedan sesgados al mismo y, que este factor supone una desventaja, pues todas las imágenes son de piel más bien rosa y de tonalidad clara únicamente.

De este modo y en relación con el punto tratado en el apartado anterior, se propone llevar a cabo el incremento de las imágenes en el nuevo *dataset* con tonalidades menos claras, dentro de la medida de lo posible, para lograr la misma clasificación que en este proyecto, pero en un rango de muestras más amplio.

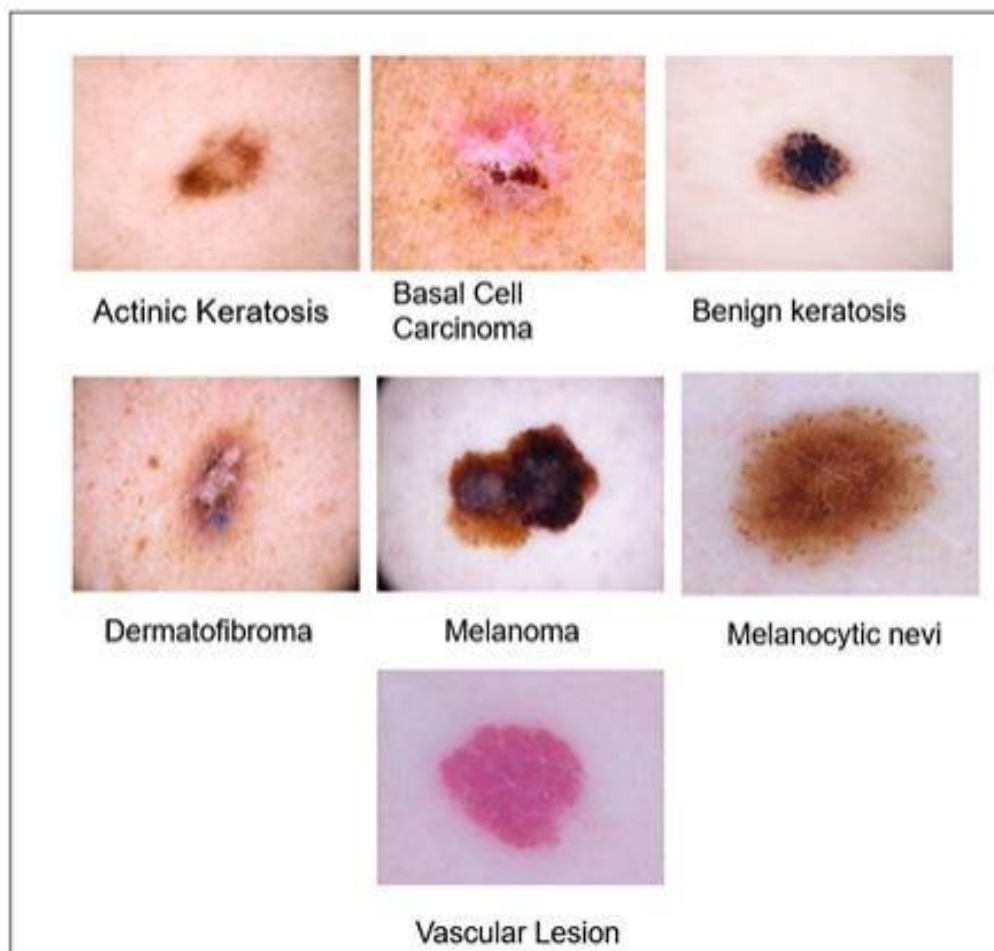


Figura 36 - Visión general de las lesiones en el conjunto de datos *HAM10000*

7.3 CONSECUCIÓN DE LOS OBJETIVOS

Por último, con este desarrollo queda más cerca la intención de poder hacer más accesible un primer análisis de la medición de las lesiones en la piel de los pacientes, pudiendo así alertarles en el respectivo caso para proceder a un estudio de más alto nivel y

con mayor elaboración. Esto supone un paso en la dirección adecuada para la creación de la aplicación móvil que pueda analizar afecciones en la dermis de los diferentes pacientes. Siendo el desarrollo de código abierto, queda disponible a cualquier persona que lo desee para su consulta y aportación.

Capítulo 8. BIBLIOGRAFÍA

American Cancer Society. (14 de Agosto de 2019). Etapas del cáncer de piel tipo melanoma. *American Cancer Society*, 1.

Aptoide. (s.f.). Obtenido de <https://moleexplore.es.aptoide.com/app>

Arellano, A. R. (23 de Mayo de 2018). *Academia Española de Dermatología y Venerología*. Obtenido de <https://aedv.es/ederma-una-aplicacion-seguimiento-las-lesiones-piel-sospechosas/>

Boesch, G. (2022). *Viso.AI*. Obtenido de Deep Residual Networks (ResNet, ResNet50) – Guide in 2022: <https://viso.ai/deep-learning/resnet-residual-neural-network/>

Calzado, A., & Geleijns, J. (24 de Diciembre de 2010). Tomografía computarizada. Evolución, principios técnicos y aplicaciones. *Didac*, 18.

DiarioFarma. (23 de Mayo de 2018). Obtenido de <https://www.diariofarma.com/2018/05/23/una-aplicacion-util-para-la-autoexploracion-de-piel-contra-el-melanoma>

Flores, D. W. (s.f.). *Cinvestav*. Obtenido de Perceptrón multicapa: https://www.tamps.cinvestav.mx/~wgomez/notas/RP/Notas_Clase12.pdf

G.Cid. (11 de Febrero de 2022). *El Confidencial*. Obtenido de https://www.elconfidencial.com/tecnologia/ciencia/2020-02-11/apps-cancer-piel-aplicaciones-peligro-inseguras_2448988/

Géron, A. (2017). Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. *O'Reilly*.

Gluon. (2021). *Gluon*. Obtenido de Prepare the ImageNet dataset: https://cv.gluon.ai/build/examples_datasets/imagenet.html

- Google. (s.f.). *Colaboratory: Frequently Asked Questions*. Obtenido de Google:
<https://research.google.com/colaboratory/faq.html>
- Hameed, N., M.Shabut, A., K.Ghosh, M., & Hossain, M. (18 September 2019). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. 1-18.
- He, K., Zhang, X., Ren, S., & Sun, J. (10 de Diciembre de 2015). *Cornell University*. Obtenido de Deep Residual Learning for Image Recognition:
<https://arxiv.org/abs/1512.03385>
- I.E.S. Suel. (s.f.). *Ciencias Naturales*. Obtenido de IESSuel:
http://www.iessuel.es/ccnn/interactiv/s_nervioso/s_nervioso_y_endocrino_06.htm
- Keshav. (s.f.). *Vidyasheela*. Obtenido de Introduction To Gradient descent algorithm (With Formula): <https://vidyasheela.com/post/introduction-to-gradient-descent-algorithm-with-formula>
- MHealth Belgium. (s.f.). *SkinVision by SkinVision BV*. Obtenido de <https://mhealthbelgium.be/apps/app-details/skinvision>
- Molexplore. (6 de febrero de 2022). Obtenido de <https://molexplore.com/es/>
- Moran, M. (2021). *Objetivos de Desarrollo Sostenible*. Obtenido de <https://www.un.org/sustainabledevelopment/es/health/>
- Naciones Unidas. (2021). *Objetivos de Desarrollo Sostenible*. Obtenido de <https://www.un.org/sustainabledevelopment/es/inequality/>
- Piccinini, G. (2004). THE FIRST COMPUTATIONAL THEORY OF MIND AND BRAIN: A CLOSE LOOK AT MCCULLOCH AND PITTS'S "LOGICAL CALCULUS OF IDEAS IMMANENT IN NERVOUS ACTIVITY". *Scinapse*, 195.

- Prabhu. (4 de Marzo de 2018). Understanding of Convolutional Neural Network (CNN) - Deep Learning.
- Rahaman, M. M. (Julio de 2020). *ResearchGate*. Obtenido de https://www.researchgate.net/figure/An-example-of-an-overfitted-accuracy-loss-curve-while-training-DenseNet201-without-data_fig3_343584572
- Raschka, S., & Mirjalili, V. (2017). Clasificar imágenes con redes neuronales convolucionales profundas, Python Machine Learning, aprendizaje automático con Python, scikit-learn y TensorFlow. *Marcombo*, 515-558.
- Russakovsky, O., Deng, J., & Su, H. (1 de Setiembre de 2014). *Cornell University*. Obtenido de ImageNet Large Scale Visual Recognition Challenge: <https://arxiv.org/abs/1409.0575>
- Sachan, A. (2021). *CV-Tricks.com*. Obtenido de Detailed Guide to Understand and Implement ResNets: <https://cv-tricks.com/keras/understand-implement-resnets/>
- Sah, E. (29 de Diciembre de 2018). *Medium*. Obtenido de Deep Learning - In simple words...: <https://medium.com/@eshant.sah/deep-learning-in-simple-words-d6e027468836>
- Sanofi. (18 de Julio de 2020). Obtenido de <https://campus.sanofi.es/es/noticias/2020/tecnologia-aplicada-en-el-diagnostico-del-cancer-de-piel/>
- SkinVision. (6 de febrero de 2022). Obtenido de <https://www.skinvision.com/articles/how-machine-learning-detects-skin-cancer/>
- Steve. (31 de Enero de 2020). Building ResNet from scratch in PyTorch.
- Tensorflow. (2017). *Transfer learning and fine-tuning*. Obtenido de https://www.tensorflow.org/tutorials/images/transfer_learning

- Tensorflow. (2021). *Convolutional Neural Network (CNN, “Red neuronal convolucional (CNN) / TensorFlow Core.* Obtenido de <https://www.tensorflow.org/tutorials/images/cnn>
- TensorFlow. (s.f.). *Why TensorFlow.* Obtenido de TensorFlow: <https://www.tensorflow.org/about>
- Thinkage. (12 de Febrero de 2017). *Github.* Obtenido de Deep Learning 2 - Introduce the activation functions for neural network: <https://schwalbe10.github.io/thinkage/2017/02/12/activation.html>
- Tokio School New Technology. (11 de Febrero de 2022). *Tokio School.* Obtenido de <https://www.tokioschool.com/noticias/sueldo-programador-python-descubre-su-salario/>
- Tschandl, P., Rosendahl, C., & Kittler, H. (14 de Agosto de 2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 1-9.
- Uq.edu.au. (22 de Noviembre de 2016). *Action potentials and synapses.* Obtenido de <https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/action-potentials-and-synapses>
- ViewNext. (19 de Agosto de 2020). *ViewNext An IBM Subsidiary.* Obtenido de El transfer learning y las redes convolucionales: <https://www.viewnext.com/transfer-learning-y-redes-convolucionales/>
- Wikimedia. (10 de Diciembre de 2018). *Keras.* Obtenido de WIKIPEDIA La enciclopedia libre: <https://es.wikipedia.org/wiki/Keras>
- Wikipedia. (3 de Setiembre de 2021). *Fundación Wikimedia.* Obtenido de WordNet: <https://es.wikipedia.org/wiki/WordNet>

Wikipedia Contributors. (21 de Julio de 2021). *Gradient descent*. Obtenido de Wikipedia:
https://en.wikipedia.org/wiki/Gradient_descent

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

En esta sección se trata la relación del proyecto de fin de grado con los Objetivos de Desarrollo Sostenible (ODS), aprobados en el año 2015 en la Agenda 2030 para el Desarrollo Sostenible. Concretamente, encaja con los Objetivos de Desarrollo Sostenible:

- ✚ 3 (Moran, 2021): aumentar la esperanza de vida y reducir algunas de las causas de muerte comunes y erradicar por completo una gran variedad de enfermedades y abordar un gran número de problemas de salud, tanto constantes como emergentes, en particular con el punto 3.4 descrito como reducir en un tercio la mortalidad prematura por enfermedades no transmisibles mediante la prevención y el tratamiento y promover la salud mental y el bienestar.
- ✚ 10 (Naciones Unidas, 2021): reducir las desigualdades (y la desigualdad dentro de los países) y garantizar que nadie se queda atrás, en particular con el punto 10.3, garantizando la igualdad de oportunidades y reducir la desigualdad de resultados.

A lo largo del presente documento, se ha recalcado reiteradamente la importancia de la detección temprana del cáncer de piel (melanoma maligno) para poder hacer frente a la enfermedad, pues de no ser así resulta incurable (Sanofi, 2020). Sabemos que tan solo en nuestro país, esta enfermedad es la responsable de más del 90% de las muertes por cáncer de piel, además de ser uno de los más costosos (considerándose su diagnóstico, tratamiento y seguimiento (DiarioFarma, 2018)). Se espera pues que, gracias a este desarrollo, en el caso de llevar a cabo la implementación de una aplicación móvil basada en el mismo, todos los usuarios dispongan del acceso que les corresponde para un primer diagnóstico.

ANEXO II

Enlace al cuaderno elaborado con *Google Colaboratory* con el código y detalles del proyecto para la ejecución de los modelos expuestos en este documento:

https://colab.research.google.com/drive/1ZqQ7IK48GD_5avwGHeWrA_IMNMCwG-Ub?usp=sharing