

**ASISTENCIA DIAGNÓSTICA CON VISIÓN ARTIFICIAL Y DEEP LEARNING DE LESIONES MELANOCÍTICAS EN IMÁGENES DERMATOSCÓPICAS CON INTERFAZ GRÁFICA PARA CARGA DE IMÁGENES**

* + 1. **Integrantes:**

Mario Miguel Barbosa Rengifo

José Francisco Martínez Amaya

Luis Fernando Ramírez Cardozo

Jhony Andrés Valderrama Molano

Proyecto transversal de las asignaturas:

Visión Computacional Con Deep Learning

Desarrollo de proyectos de inteligencia artificial

Profesores:

Juan Carlos Perafán Villota

Arturo Duque Marín

Universidad Autónoma de Occidente.  
Cali, 13 de marzo de 2021

# 1.2.1 Nombre del proyecto:

Asistencia diagnóstica con visión artificial y Deep learning de lesiones melanocíticas en imágenes dermatoscópicas con Interfaz gráfica para la carga de imágenes.

# 

# 1.2.2 Integrantes

* Mario Miguel Barbosa Rengifo
* José Francisco Martínez Amaya
* Luis Fernando Ramírez Cardozo
* Jhony Andrés Valderrama Molano

# 1.2.3 Planteamiento del problema

La dermatocospía es una técnica diagnóstica de relativo bajo costo y amplia accesibilidad, que permite mejorar el diagnóstico de lesiones dermatológicas en comparación con el ojo desnudo (sin amplificación). Los avances tecnológicos en los dispositivos móviles y las cámaras fotográficas han permitido la recolección de muchas imágenes de lesiones dermatológicas de una manera estandarizada y adecuada. Lo anterior hace que estas imágenes sean propicias para el entrenamiento de redes neuronales para la asistencia diagnóstica de lesiones cutáneas. (Ver figura 1)



**Figura 1. Ilustración de un aplicativo de asistencia diagnóstica dermatológica**

Muchos autores han logrado entrenar redes neuronales con el fin de detectar melanoma, un tipo de cáncer que es muy agresivo y cuyo diagnóstico no es sencillo. A pesar de los logros iniciales, los primeros estudios carecen de un tamaño de muestra adecuado y sus poblaciones no representaban otras lesiones cutáneas que se pueden confundir con el melanoma. Los recientes avances en las tarjetas gráficas y las técnicas de aprendizaje automático representan un incentivo para aproximar la complejidad de este problema diagnóstico y genera expectativa con respecto a la posibilidad de tener sistemas de diagnóstico automático que no requieren la intervención de un experto humano**1**⁠.

**Problema**

¿Es posible validar la utilidad del entrenamiento de modelos de CNN, como Inception-v4 e Inception-ResNet-V2 para tareas de detección y localización de lesiones melanocíticas en imágenes dermatoscópicas empleando un conjunto de datos diferente al empleado por los investigadores del estudio original?

# 1.2.4 Posible metodología usada para resolver el problema planteado

* + 1. Se realizará preprocesamiento del conjunto de datos, posteriormente se implementará al menos dos arquitecturas de redes neuronales convolucionales, se calculará las métricas (Recall, Accuracy), además de la matriz de confusión de cada modelo. Según lo relatado por otros autores existen arquitecturas con adecuado desempeño en el problema (Ver tabla 1). Estas arquitecturas son las RESNET, SQUEEZNET, DENSENET, INCEPTIONV3. Sin embargo, se desconoce el desempeño de nuevas arquitecturas como ResNet-50, Xception (2016), INCEPTION-V4 (2016), INCEPTION-RESNET-V2 (2016), ResNeXt-50 (2017) en la asistencia diagnóstica de melanoma por lo que su implementación podría mejorar las métricas de desempeño del modelo**2**⁠.
    2. **Tabla 1.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelo** | **(Precisión)** | **(F1-Score)** | **(ROC AUC)** |
| MODEL2(ResNet) | 94.24 | 94.22 | 98.61 |
| MODEL3(SqueezNet) | 97.40 | 94.57 | 99.77 |
| MODEL4(DenseNet) | 97.51 | 96.27 | 99.09 |
| MODEL5(Inceptionv3) | 98.19 | 95.74 | 99.23 |

# 1.2.5 Breve descripción teórica de cada tema usado en la metodología

Para satisfacer los requerimientos de complejidad de este proyecto y en concordancia con los temas abordados en este curso se escogieron los siguiente objetivo a desarrollar:

1. **Detección de lesiones melanocíticas en imágenes diagnosticas mediante modelos de CNN.**

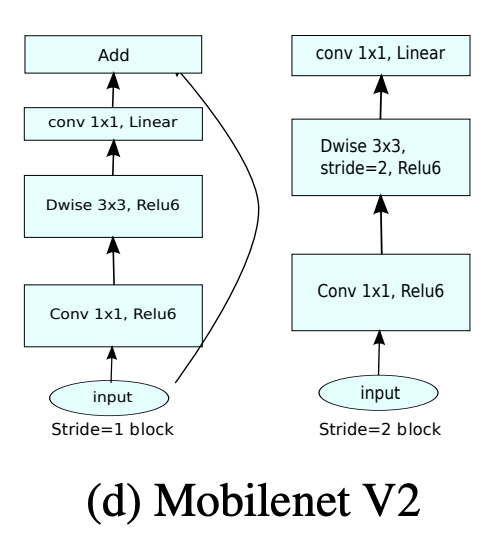
**Dataset**

Para este fin, se propone emplear el dataset: **“The HAM10000”** dataset, que contiene una gran colección de imágenes dermatoscópicas de lesiones cutáneas pigmentadas comunes. El dataset se encuentra disponible en: http://dx.doi.org/10.1038/sdata.2018.161

**Modelos**

Aunque aún desconocemos la arquitectura que se empleará para la tarea de localización de las lesiones, para la parte de detección consideramos emplear las versiones más recientes de los dos modelos, Inceptionv3 y ResNet, ya empleados para esta tarea, los cuales son Inception-v4 e Inception-ResNet-V2, respectivamente. Se espera que estos dos modelos mejoren significativamente los resultados reportados en la publicación original de la investigación y que se muestran en la tabla 1. Pero esta vez utilizando un nuevo dataset con imágenes similares. A continuación, mencionamos las características básicas de estos dos modelos:

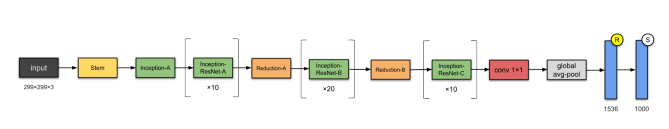
**MobileNetV2**

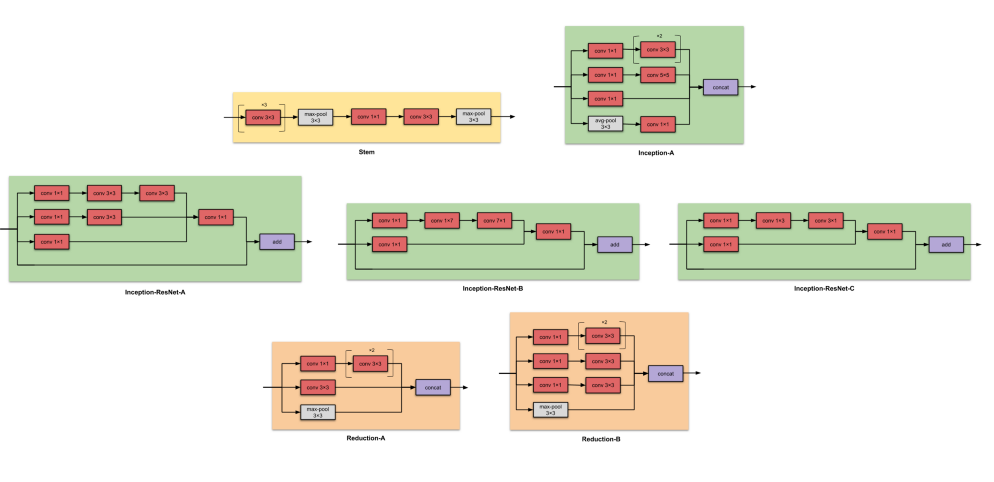
Es una arquitectura de redes neuronales convolucionales que buscan tener un buen desempeño en dispositivos móviles. Está basada en la estructura residual invertida donde las conexiones residuales son entre las “capas de cuello de botella”. La capa de expansión intermedia usa convoluciones livianas y poco profunda para filtrar características como fuente de no linealidad. En resumen, la arquitectura de la MobileNetV2 contiene inicialmente una capa convolucional con 32 filtros, a continuación 19 capas de cuello de botella. 

**Figura 2. Arquitectura del modelo MobileNetV2**

**Inception-ResNet-V2**

En el mismo artículo de la arquitectura de Inception-v4 los autores introdujeron una familia que mezcla Inception-ResNet-V1 y V2. Esta red tiene 56M de parámetros, convierte los módulos de incepción a bloque residuales de incepción, añade más bloques de incepción y añade un nuevo tipo de módulo de incepción(Incepción-A) después de un módulo fuente (steam)**3**.

****

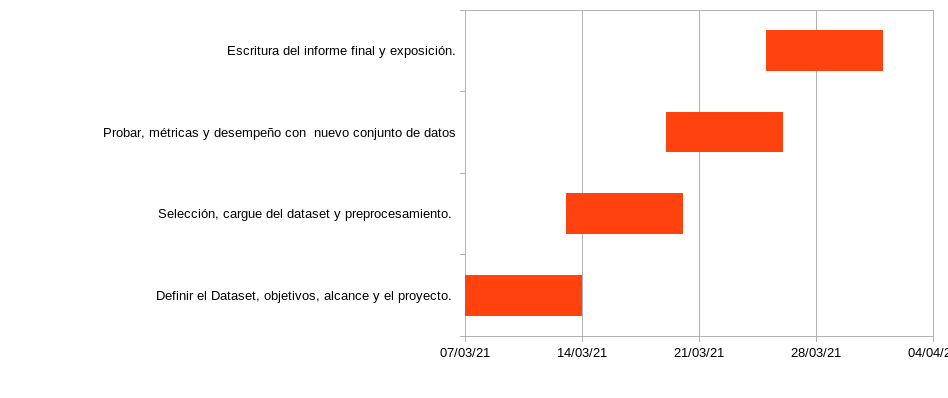


**Figura 3. Arquitectura del modelo Inception-ResNet V2**

# 1.2.6 Cronograma de actividades

|  |  |  |  |
| --- | --- | --- | --- |
| **Actividades** | **Fecha de inicio** | **Duración** | **Fecha final** |
| **Definir el Dataset, objetivos, alcance y el proyecto.** | 21-03-07 | 7 | 21-03-13 |
| **Selección, cargue del dataset y preprocesamiento.** | 21-03-13 | 7 | 21-03-19 |
| **Entrenamiento, test, desempeño y métricas con nuevo conjunto de datos** | 21-03-19 | 7 | 21-03-25 |
| **Escritura del informe final y exposición.** | 21-03-25 | 7 | 21-03-31 |

**Tabla 2. Cronograma de actividades**

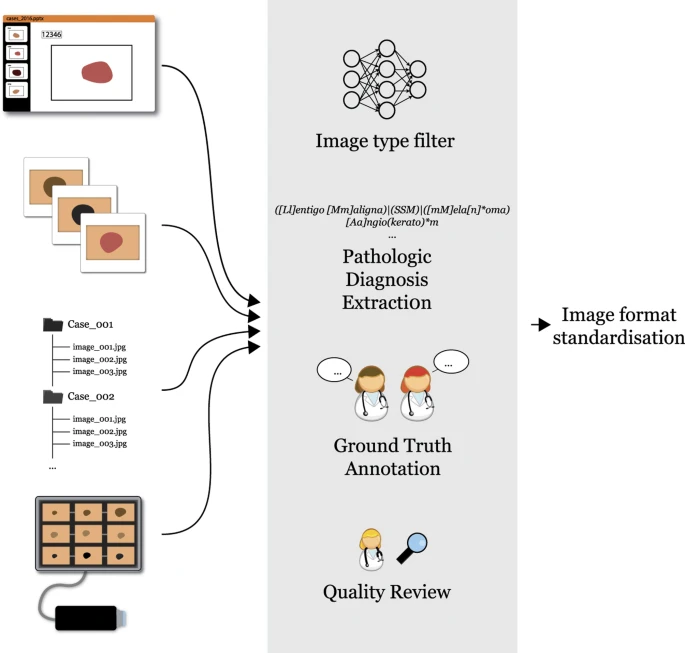


**Figura 4. Diagrama de Gantt del proyecto**

# 1.2.7 Resultados.

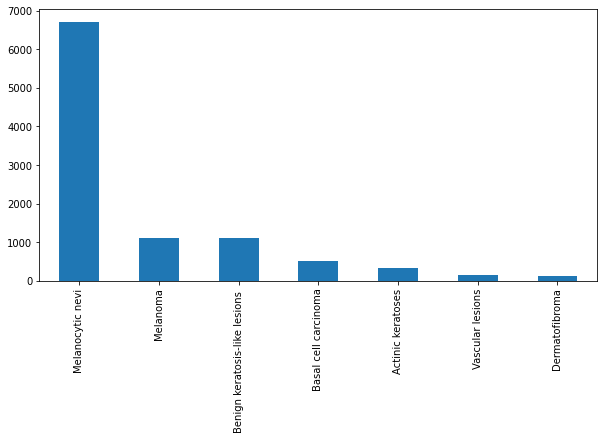
## 1.2.7.1 Descripción del conjunto de datos.

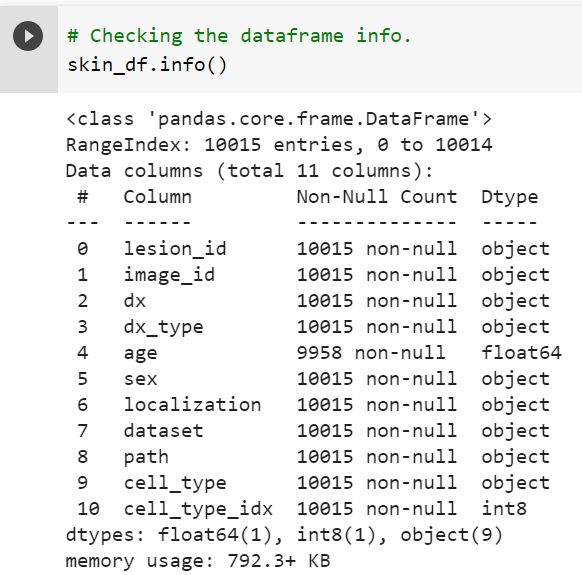
El entrenamiento de redes neuronales para la asistencia diagnóstica de lesiones pigmentadas es obstaculizado por el tamaño de muestra de los conjuntos de datos(dataset) y por su poca diversidad en términos de representación de lesiones. El conjunto de datos HAM10000(“Human Against Machine with 10000 training images”) es el resultado de una recolección de imágenes dermatoscopicas a partir de diferentes poblaciones. Dada la diversidad de fuentes se realizó una limpieza de las imagenes. El conjunto final de datos consiste en 10015 imagenes de dermatoscopias liberadas para el entrenamiento académico de modelos de aprendizaje automático, los diagnósitcos se confirmaron por la toma de una biopsia, seguimiento de las lesiones en lois pacientes, consenso de expertos o microscopía confocal.

****

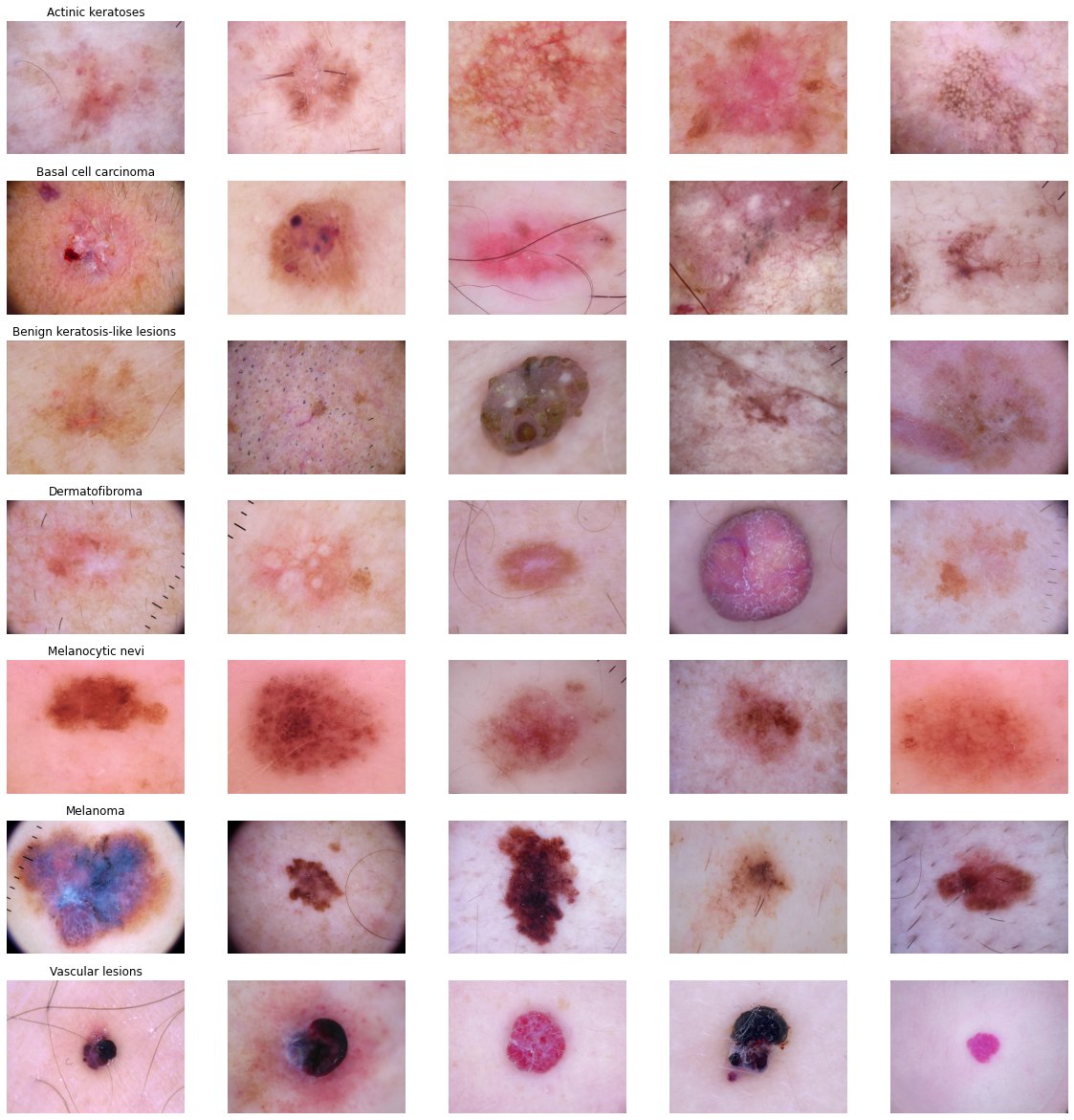
## 1.2.7.1 Preprocesamiento del conjunto de datos.

Para realizar análisis de las imágenes se realizó un análisis descriptivo del conjunto de datos tomando como fuente el archivo de metadata publicado en el repositorio. Se observó que la categoría más frecuente fue la de NV(nevus melanocítico) y que la otras categorías. Además la edad presenta pocos datos pérdidos.

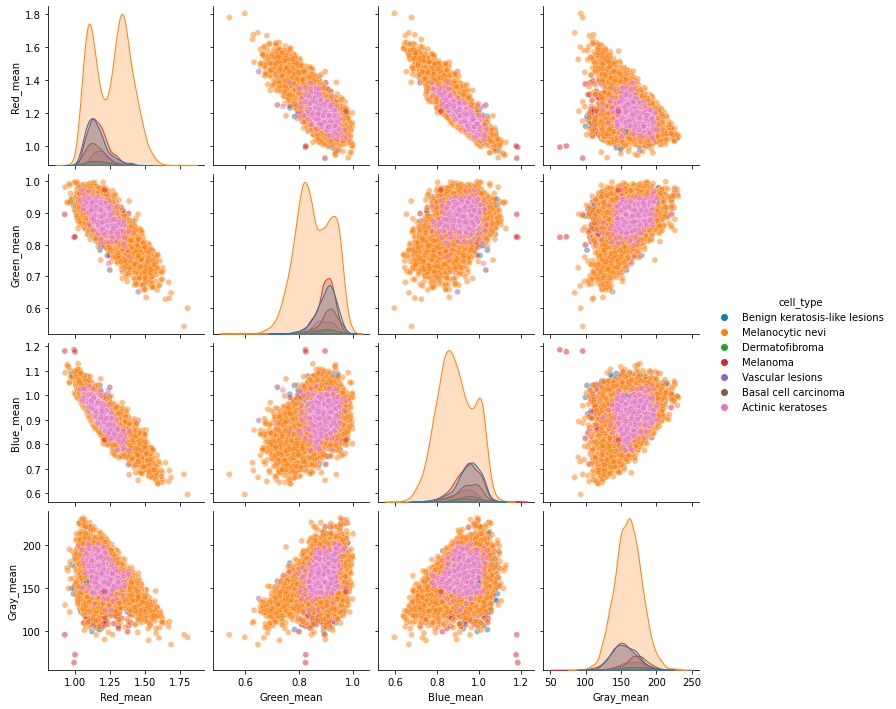
****

****

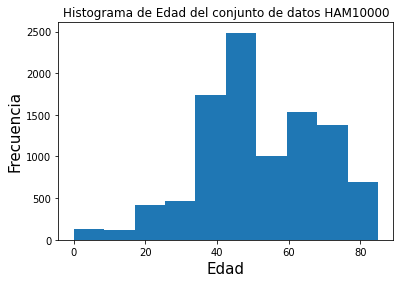
Se visualizan algunas imágenes por cada categoría. Se puede observar que las imágenes de la categoría Melanoma cumplen con las características clínicas de dichas lesiones en el “abcd”, es decir son asimétricas(a), con bordes irregulares(b), coloración no homogénea(c) y diámetro considerable(d). Las imágenes de lesiones vasculares son “rojizas o azulosas”. Las correspondientes a la categoría dermatofibroma son poco coloridas y pueden presentar nodulaciones no necesariamente diagnósticas de la categoría. Un ruido especial de este tipo de imagen lo constituye la presencia de cabellos los cuales podrían retirarse manualmente o con un modelo de detección.

****

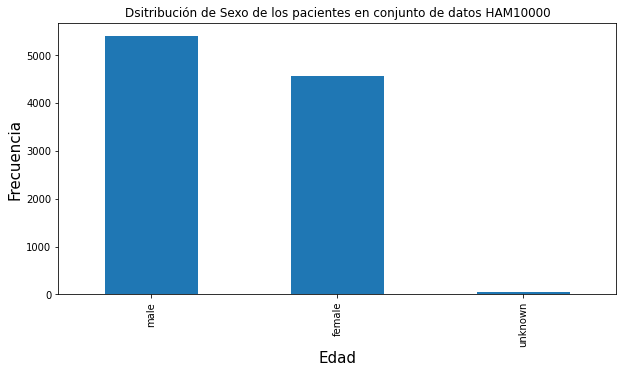
Al evaluar la distribución de las intensidades de color por categoría se observa un comportamiento bimodal en el rojo para la categoría de “nevo melanocítico”, esta distribución en menos pronunciada al evaluar la distribución en los otros colores. Esto puede corresponder a un artefacto de la captación de las imágenes o a una característica propia de la lesión, sin embargo otras aproximaciones han encontrado hallazgos concordantes con que sea una característica propia del melanoma al entrenar con conjunto de datos más pequeños. Otros dos aspectos a tener en cuenta son la evaluación de la asimetría con segmentaciones manuales y la reducción de ruido detectando y la vasculatura de la lesión(4).

****

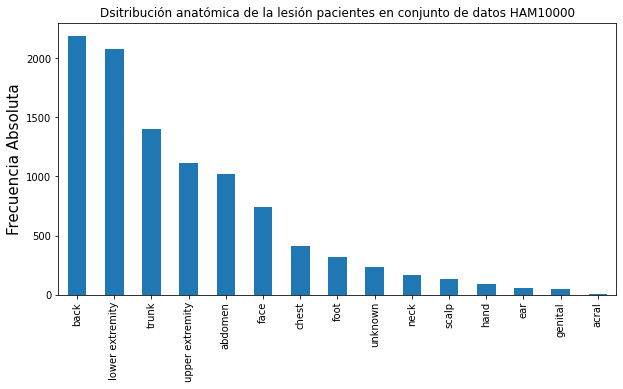
La distribución de la edad en el dataset representa mejor la población entre 60 y 80 años con una curva sesgada a la derecha.

****

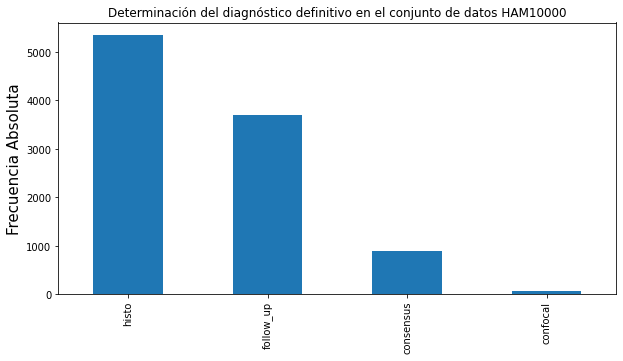
La distribución por sexo es muy similar con mayor representación de hombres que mujeres.

****

La distribución anatómica de la lesión más frecuente es la espalda, el tronco y las extremidades inferiores.

****

El diagnóstico definitivo de la lesión se realizó en la mayoría de los casos con la biopsia de la lesión(categoría “histo”). Sin embargo, una proporción importante del dataset tuvo como metodología de diagnóstico el seguimiento.

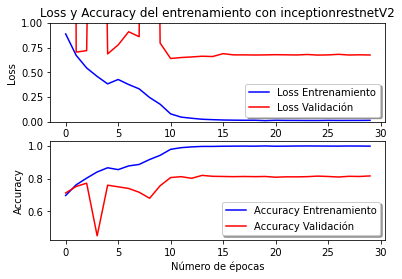


# 1.2.8 Entrenamiento de la red y métricas.

## 1.2.8.1 InceptionResNetV2

Para el entrenamiento de la red se procedió a cambiar el tamaño de las imagenes originalmente en una resolución de 800x600, y se redujo a 75x75, por razones de presupuesto computacional. Luego de esto se parametrizó el campo objetivo del dataframe construido con el diagnóstico. Para la optimización del modelo se empleó el algoritmo ADAM con una tasa de aprendizaje(learning rate) de 0.001 y como función de pérdida la entropía cruzada categoría(cross entropy loss). Se usaron los pesos entrenados de la red con el dataser imagenet, finalemnte la última capa se desconecto para reemplazar con las categorías de los diagnósticos del Dataset.

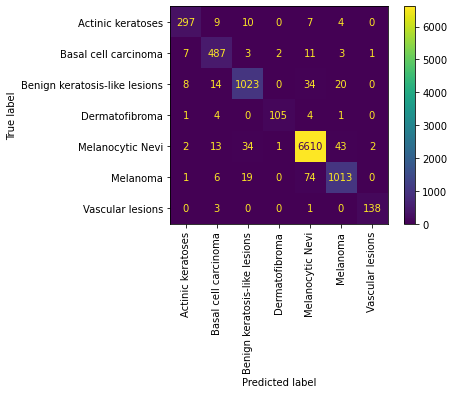
Con estos valores se procedió a realizar una evaluación de las métricas de Loss y accuracy como se puede ver a continuación:

****

Se entrenó con 30 épocas, pero identificamos que el valor mínimo de pérdida ocurrió en la época 26, además después de esta época presentó un nuevo incremento del valor de loss. Como se puede observar en la gráfica el valor del Accuracy(desempeño) en el conjunto de validación es menor al del conjunto de entrenamiento lo que nos permite concluir que no existe sobreajuste.

**Matriz de confusión**

Al evaluar la matriz de confusión del modelo se observa que tanto el recall, accuracy y F1-score se encuentran por encima del 90%. El mayor valor ocurrió en la categoría del nevus melanocítico, de ínteres para el proyecto los valores de específicos para el diagnóstico de melanoma fueron 93%,91% y 92%. El valor de Kappa fue de 0.93 lo cual se interpreta como acuerdo sustancial(mayor a 0.8).

****

**precision recall f1-score support**

**Actinic keratoses 0.94 0.91 0.92 327**

**Basal cell carcinoma 0.91 0.95 0.93 514**

**Benign keratosis-like lesions 0.94 0.93 0.94 1099**

**Dermatofibroma 0.97 0.91 0.94 115**

**Melanocytic Nevi 0.98 0.99 0.98 6705**

**Melanoma 0.93 0.91 0.92 1113**

**Vascular lesions 0.98 0.97 0.98 142**

**accuracy 0.97 10015**

**macro avg 0.95 0.94 0.94 10015**

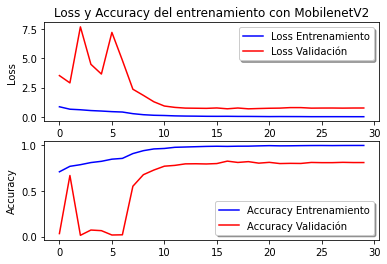
**weighted avg 0.97 0.97 0.97 10015**

## 1.2.8.2 MobilenetV2

Para el entrenamiento de la red se procedió a cambiar el tamaño de las imagenes originalmente en una resolución de 800x600, y se redujo a 96x96, por razones de presupuesto computacional y por ser la una entrada válida a la red mobilenetV2. Luego de esto se parametrizó el campo objetivo del dataframe construido con el diagnóstico. Para la optimización del modelo se empleó el algoritmo ADAM con una tasa de aprendizaje(learning rate) de 0.001 y como función de pérdida la entropía cruzada categoría(cross entropy loss). Se usaron los pesos entrenados de la red con el dataser imagenet, finalemnte la última capa se desconectó para reemplazar con las categorías de los diagnósticos del Dataset.

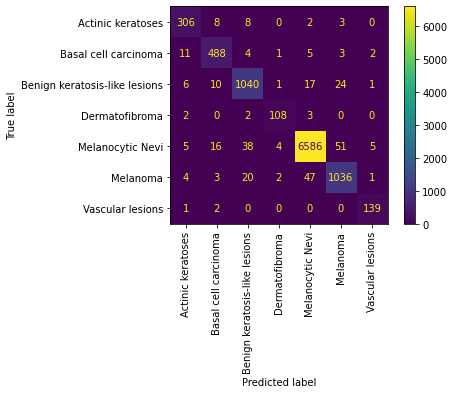
Con estos valores se procedió a realizar una evaluación de las métricas de Loss y accuracy como se puede ver a continuación:

Se entrenó con 30 épocas, pero identificamos que el valor mínimo de pérdida ocurrió en la época 29, además después de esta época presentó un nuevo incremento del valor de loss. Como se puede observar en la gráfica el valor del Accuracy(desempeño) en el conjunto de validación es menor al del conjunto de entrenamiento lo que nos permite concluir que no existe sobreajuste.

****

**Matriz de confusión**

Al evaluar la matriz de confusión del modelo se observa que tanto el recall, accuracy y F1-score se encuentran por encima del 90%. El mayor valor ocurrió en la categoría del nevus melanocítico, de ínteres para el proyecto los valores de específicos para el diagnóstico de melanoma fueron 93%,93% y 93%. El valor de Kappa fue de 0.93 lo cual se interpreta como acuerdo sustancial(mayor a 0.8).

****

**precision recall f1-score support**

**Actinic keratoses 0.91 0.94 0.92 327**

**Basal cell carcinoma 0.93 0.95 0.94 514**

**Benign keratosis-like lesions 0.94 0.95 0.94 1099**

**Dermatofibroma 0.93 0.94 0.94 115**

**Melanocytic Nevi 0.99 0.98 0.99 6705**

**Melanoma 0.93 0.93 0.93 1113**

**Vascular lesions 0.94 0.98 0.96 142**

**accuracy 0.97 10015**

**macro avg 0.94 0.95 0.94 10015**

**weighted avg 0.97 0.97 0.97 10015**

# 1.2.9 Discusión.

Al comparar con otros artículos que realizaron análisis similares al del presente trabajo, las métricas se ubican por encima del 90-95%(tanto para recall, accuracy y F1-score). Sin embargo, el hecho de que una parte importante del diagnóstico se realizara con seguimiento de la lesión indica que la característica de “ground true” puede no ser correcta y esto añadiría incertidumbre epistémica al modelos entrenado. Desde el punto de vista metodológico esto se compensa midiendo el acuerdo entre el modelo y los datos reportados con el estadístico kappa el cual presenta un valor de 0.93 que indica un alto grado de concordancia de las predicciones del modelo con los diagnósticos obtenidos en el dataset. A pesar de esta métrica la validación debe pasar por una validación geográfica(que incluya personas con otros tonos de piel por ejemplo y una frecuencia distinta de las lesiones) y una validación temporal(es decir medir el desempeño o el entrenamiento en otros años para que el efecto cohorte de estas lesiones no interfiera con el modelo).

Por otro lado es importante mencionar que se dejó todo el entrenamiento a la red sin tratar de establecer la utilidad de ciertos conceptos técnicos conocidos en la literatura médica(como por ejemplo la asimetría de la lesión como posible característica de melanoma). Además, otro aspecto que no se tuvo en cuenta es explorar que posibles variables evaluaba el modelo y de esta forma optimizar nuevos entrenamientos. Desde el punto de vista de identificación de elementos ruidosos tampoco se realizó una limpieza de las imágenes de artefactos que no generan información diagnóstica(como por ejemplo la remoción de pelos o superficies irregulares como presentan las lesiones de las orejas). Finalmente desde el punto de vista de visión computacional clásica tampoco se empleó ningún algoritmo para detectar puntos de interés.

A pesar de las limitaciones y los múltiples elementos no evaluados, la visión artificial con deep learning constituye una herramienta de asistencia de toma de decisiones clínicas que podría optimizar los recursos de los sistemas de salud por medio de la clasificación y optimización del orden de consulta de los pacientes, además de tener el potencial de disminuir las barreras de acceso a consulta al no subestimar la gravedad de las lesiones o permitir que personas con acceso a un celular o una cámara tengan una primera impresión diagnóstica.

# 1.2.10 Conclusiones.

1. Las redes InceptionrestnetV2 y MobilenetV2 al ser entrenadas con el dataset HAM10000 pudieron generar un modelo de clasificación con métricas de discriminación y exactitud adecuadas.
2. Las redes InceptionrestnetV2 y MobilenetV2 pueden ser una opción para implementar un sistema de ayuda de decisiones diagnósticas.
3. Se debe mejorar la infraestructura de sistemas de información de los hospitales y centros de atención clínica de tal manera que permitan construir de manera continua diferentes modelos de ayuda de decisiones clínicas que se actualicen con base en nuevas características clínicas y datos de la historia clínica.
   * 1. **1.2.7 Bibliografía.**
4. Tschandl P, Rosendahl C, Kittler H. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data [Internet]. 2018;5:1–9. Available from: http://dx.doi.org/10.1038/sdata.2018.161
5. Kadampur MA, Al Riyaee S. Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Informatics Med Unlocked [Internet]. 2020;18(November 2019):100282. Available from: https://doi.org/10.1016/j.imu.2019.100282
6. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI Conf Artif Intell AAAI 2017. 2017;4278–84.
7. Pathan S, Aggarwal V, Gopalakrishna Prabhu K, Siddalingaswamy PC. Melanoma Detection in Dermoscopic Images using Color Features. Biomed Pharmacol J. 2019;12(1):107–15.
8. Kadampur MA, Al Riyaee S. Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Informatics Med Unlocked [Internet]. 2020;18(November 2019):100282. Available from: https://doi.org/10.1016/j.imu.2019.100282