

**Facultad de ingeniería**

**Visión por computadora**

**PROYECTO 4 SKIN CANCER MNIST: HAM10000**

**Autores:**

**Jairo Andrés Carrasco Gárate**

**Diego Benjamin Muñoz Zapata**

**Juan Pablo Olmedo Saavedra**

**Sebastián Ignacio Pinto Espinosa**

**Sebastián Andrés Quiroz Saavedra**

**Addán Ignacio Sáez Rodríguez**

**Profesor:**

**Billy Mark Peralta Marquez**

**Santiago**, **Chile**

**2025**

**ÍNDICE**

[1. DEFINICIÓN DE PROYECTO 3](#_heading=h.5t9jadtna7na)

[1.1 Contexto 3](#_heading=h.ns9mirww5zcd)

[1.2 Importancia 3](#_heading=h.b244cnkzgt9a)

[1.3 Desafíos y limitaciones 4](#_heading=h.7ra7sffqdrsw)

[2. DESCRIPCIÓN DE DATOS 5](#_heading=h.3et2vm459veq)

[1. Identificación de Imágenes y Lesiones 5](#_heading=h.mj4r71wyap7z)

[2. Diagnóstico (dx) 5](#_heading=h.p1fbovqxbtlt)

[3. Tipo de Verificación Diagnóstica (dx\_type) 6](#_heading=h.bxjy41l31knc)

[4. Edad de los Pacientes 6](#_heading=h.6pdp6z1lik59)

[5. Sexo 6](#_heading=h.5uhryo7co7i)

[6. Localización Anatómica 7](#_heading=h.xohbr8cnktie)

[3. DISEÑO DE EXPERIMENTOS 8](#_heading=h.tvuqt2j1kt9z)

[4. EVALUACIÓN 9](#_heading=h.jbiprdlolu05)

[5. REFERENCIAS 10](#_heading=h.yjaui7audm2d)

# DEFINICIÓN DE PROYECTO

El proyecto consiste en el desarrollo y fortalecimiento de herramientas de inteligencia artificial para el diagnóstico automatizado de lesiones cutáneas pigmentadas utilizando un conjunto de datos **Skin Cancer MNIST: HAM10000.**

## 1.1 Contexto

El diagnóstico del cáncer de piel, es un proceso que requiere el análisis visual por parte de especialistas. Los especialistas dermatólogos aunque son expertos, la carga de trabajo y la sutileza de algunas lesiones pueden hacer que el diagnóstico sea difícil y propenso a errores. El conjunto HAM1000 (“Human Multiclass Skin Cancer Dataset”), fue creado por Philipp Tschandl, donde se encuentran imágenes de lesiones de piel que han sido confirmadas histopatológicamente o por seguimiento clínico.

## 1.2 Importancia

El desarrollo de un modelo preciso para clasificar el cáncer de piel es crucial por varias razones:

i. Detección temprana y mejora de los resultados: La detección temprana del cáncer de piel es super importante y vital para el lograr un tratamiento exitoso. Un modelo o herramienta que reconozca e identifique las lesiones sospechosas puede disminuir el tiempo de diagnóstico, lo que por consiguiente lleva a un tratamiento más oportuno y aumento en la tasa de supervivencia.

ii. Apoyo a los profesionales de la salud: El modelo no reemplaza al especialista, sino que actúa como un “segundo diagnóstico”. Ayuda a priorizar casos con alta probabilidad de malignidad, lo que puede permitir a los médicos enfocarse en los casos más críticos.

iii. Reducción de la carga de trabajo: La automatización de una parte del proceso de detección permitirá reducir la carga de trabajo de los especialistas de la salud.

iv. Investigación y desarrollo: El proyecto sirve como una plataforma para explorar y comparar diferentes arquitecturas de redes neuronales, colaborando a la investigación en ia aplicada a la salud.

## 1.3 Desafíos y limitaciones

Según Bupa Seguros explica que a pesar de las ventajas, la implementación de la IA en oncología aún enfrenta desafíos como:

i. La dependencia de datos de calidad: Sin datos adecuados, los algoritmos pueden fallar.

ii. Interpretabilidad de los resultados: Los especialistas deben comprender cómo la IA llega a sus resultados y conclusiones para poder confiar.

iii. Aspectos éticos y regulatorios: Además es esencial garantizar la seguridad y privacidad de los pacientes.

# DESCRIPCIÓN DE DATOS

El conjunto de los datos **HAM10000** está conformado por 10.015 imágenes dermatoscópicas que fueron recopiladas a partir de distintas poblaciones utilizando varias técnicas de adquisición de almacenamiento de datos. Este enfoque nos ayudará a tener una mejor representación clínica de estos datos y nos permitirá entrenar nuestro modelo con un mayor grado de inteligencia.

**Estructura del Dataset**

### **1. Identificación de Imágenes y Lesiones**

* **lesion\_id:** 7.470 valores únicos, permite rastrear una misma lesión a través de múltiples imágenes.
* **image\_id:** 10.015 valores únicos, identificador de cada imagen individual.

### **2. Diagnóstico (dx)**

El dataset abarca **7 categorías diagnósticas** de lesiones pigmentadas:

* **nv (nevos melanocíticos):** 6.700 casos (categoría más abundante).
* **mel (melanoma):** 1.100 casos.
* **bkl (lesiones queratósicas benignas):** 1.100 casos.
* **bcc (carcinoma basocelular):** 500 casos.
* **akiec (queratosis actínicas ):** 300 casos.
* **vasc (lesiones vasculares):** 150 casos.
* **df (dermatofibroma):** 100 casos.

### **3. Tipo de Verificación Diagnóstica (dx\_type)**

* **histo (histopatología):** referencia principal, más del 50% de los casos.
* **consensus (consenso de expertos)**
* **confocal (microscopía confocal in vivo)**
* **follow\_up (seguimiento clínico)**

### **4. Edad de los Pacientes**

* La edad abarca desde recién nacidos (0 años) hasta pacientes de 85 años.
* Los grupos más representados se concentran entre 40 y 60 años, con picos en:  
  + **45 años:** 1.299 casos
  + **50 años:** 1.187 casos
  + **55 años:** 1.009 casos

### **5. Sexo**

* **hombres)**
* **mujeres**
* **desconocido**

### 

### 

### **6. Localización Anatómica**

**Se identifican 15 regiones corporales, con predominancia en:**

* **Espalda (back): 2.192 casos**
* **Extremidades inferiores : 2.077 casos**
* **Tronco : 1.404 casos**
* **Extremidades superiores: 1.118 casos**
* **Abdomen: 1.022 casos**
* **Cara: 745 casos**

**Otras regiones menos frecuentes incluyen cuero cabelludo (128), mano (90), oreja (56), genital (48) y acral (7 casos).**

# DISEÑO DE EXPERIMENTOS

# EVALUACIÓN

# REFERENCIAS

K Scott Mader, Skin Cancer MNIST: HAM10000

<https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>

Bupa Seguros

<https://www.segurosbupa.cl/blog/la-ia-en-el-diagnostico-temprano-del-cancer>