



# Universidad Austral de Chile

Facultad de Ciencias de la Ingeniería  
Escuela de Ingeniería Civil en Informática

## **CLASIFICACIÓN DE IMÁGENES MÉDICAS PARA EL PRE-DIAGNÓSTICO DE ENFERMEDADES DE LA PIEL**

Proyecto para optar al título de  
Ingeniero Civil en Informática

**PROFESOR PATROCINANTE:**

Matthieu Vernier  
Doctor en informática

**PROFESOR CO-PATROCINANTE**

Fabián Ruiz  
Ingeniero Civil en Informática  
Magister en Informática

**PROFESOR INFORMANTE**

Walkiria Gallegos Fester  
Ingeniera Civil en Informática  
Magister en dirección general de empresas

**FRANCISCO JAVIER GONZÁLEZ GONZÁLEZ**

VALDIVIA – CHILE  
2022

# ÍNDICE

1	INTRODUCCIÓN.....	1
1.1	¿Qué es el cáncer de piel, cuál es su impacto en el mundo y en Chile?.....	1
1.2	Primera problemática en la salud pública chilena: la escasez de especialistas dermatólogos.....	2
1.3	Segunda problemática en la salud pública chilena: la derivación de pacientes y tiempos de espera.....	3
1.4	Oportunidad: mejorar la detección temprana del cáncer de la piel y reducir los errores en la derivación de pacientes utilizando tecnologías de información.....	5
1.5	Objetivo general del proyecto de título: Clasificar imágenes médicas para el pre-diagnóstico de enfermedades de la piel.....	6
2	Datos y pre procesamiento .....	7
2.1	Datasets disponibles para la clasificación de imágenes dermatológicas .....	7
2.2	Características de los datos utilizados .....	7
2.3	Clases desbalanceadas .....	8
2.4	Aumentación de datos .....	10
3	Integración del modelo de clasificación en un prototipo de aplicación móvil LUNA .....	11
3.1	Aplicaciones similares a LUNA.....	11
3.2	Requisitos funcionales y no funcionales .....	16
	Requisitos funcionales.....	16
	Aplicación móvil.....	16
	Seguridad .....	16
	Requisitos no funcionales.....	16
	Seguridad de los datos.....	16
3.3	Arquitectura de software LUNA .....	17
	Procesamiento de imágenes.....	18
	Colas de mensajes .....	19
	API REST .....	23
	Aplicación Móvil.....	23
	Workers analizadores .....	26
	Workers registradores.....	27
	Base de datos MongoDB .....	27
	Base de datos PostgreSQL .....	27
	Estructura de archivos del código fuente de LUNA.....	28
	Ejecución de LUNA en entorno local .....	29
4	Entrenamiento, validación y pruebas de modelos de clasificación de imágenes dermatológicas .....	31
4.1	Arquitecturas de red neuronal EfficientNet.....	31
4.2	Método de entrenamiento .....	34
4.3	Resultados y conclusiones post entrenamiento .....	36
5	Discusión y conclusión.....	39

6	REFERENCIAS .....	40
	ANEXO B: Estadísticas sobre el cáncer (Global Cancer Observatory).....	46

## ÍNDICE DE TABLAS

TABLA 1: ETAPAS DEL DESARROLLO DEL CÁNCER UTILIZADAS POR LOS DERMATÓLOGOS Y SU TASA DE SUPERVIVENCIA .....	1
TABLA 2: ANTERIORES PROYECTOS RELACIONADOS A LA TELEMEDICINA EN CHILE.....	5
TABLA 3: LISTA DE ARCHIVOS CSV CON METADATOS DE IMÁGENES PRESENTES EN EL DATASET SIIM-ISIC7	
TABLA 4: DESCRIPCIÓN DE LAS COLUMNAS PRESENTES EN ARCHIVOS CSV DE METADATOS.....	7
TABLA 5: DESCRIPCIÓN DE COMPONENTES DE ARQUITECTURA PROPUESTA.....	17
TABLA 6: DESCRIPCIÓN DE COMPONENTES EN RABBITMQ PARA ARQUITECTURA LUNA.....	21
TABLA 7: ENDPOINTS DISPONIBLES DE LA API REST PARA ARQUITECTURA LUNA .....	23
TABLA 8: TABLA USER.....	27
TABLA 9: TABLA ROLE.....	28
TABLA 10: ESTRUCTURA DE CARPETAS PARA EL CÓDIGO FUENTE DE COMPONENTES DE LUNA .....	28
TABLA 11: COMPONENTES ADICIONALES PARA ADMINISTRAR BASES DE DATOS .....	28
TABLA 12: TABLA COMPARATIVA DE RENDIMIENTO DE EFFICIENTNET SOBRE DATASET IMAGENET.....	32
TABLA 13: MÉTRICAS DE DESEMPEÑO LUNANET SOBRE DATASET SIIM-ISIC.....	37
TABLA 14: NÚMERO DE CASOS DE CÁNCER (NO MELANOMA) POR CADA 100.000 HABITANTES DURANTE EL AÑO 2020 .....	46
TABLA 15: TASA DE MORTALIDAD DE OTROS TIPOS DE CÁNCER (NO MELANOMA) POR CADA 100.000 HABITANTES DURANTE EL AÑO 2020 .....	46
TABLA 16: TOTAL DE CASOS Y TASA DE MORTALIDAD DE MELANOMA POR CADA 100.000 HABITANTES DURANTE EL AÑO 2020 .....	46

## ÍNDICE DE FIGURAS

FIGURA 1: DISTRIBUCIÓN DE ESPECIALISTAS DERMATÓLOGOS EN TERRITORIO CHILENO AÑO 2020	3
FIGURA 2: TIEMPOS DE ESPERA PARA PACIENTES DE CÁNCER, DESDE LA DERIVACIÓN HASTA EL PRIMER TRATAMIENTO	4
FIGURA 3: EJEMPLO DE DESBALANCE DE CLASES EN UN SISTEMA DE DETECCIÓN DE FRAUDE (FRAUDE VS NO FRAUDE).	9
FIGURA 4: EJEMPLO DE SUBMUESTREO (UNDER-SAMPLING) Y SOBRE MUESTREO (OVER-SAMPLING) SOBRE UN DATASET	9
FIGURA 5: INTERFACES DE APLICACIÓN UMSKINCHECK	12
FIGURA 6: INTERFACES APLICACIÓN MIISKIN	13
FIGURA 7: INTERFACES APLICACIÓN MEDGIC	13
FIGURA 8: INTERFACES APLICACIÓN SKINVISION	14
FIGURA 9: DIAGRAMA :DE A:RQUITECTURA LUNA	18
FIGURA 10: COMPONENTES PRESENTES EN RABBITMQ	20
FIGURA 11: FLUJO DE MENSAJES EN RABBITMQ PARA ARQUITECTURA LUNA	21
FIGURA 12: VISTA DE ANÁLISIS APLICACIÓN MÓVIL	24
FIGURA 13 : VISTA DE RESULTADOS APLICACIÓN MÓVIL	25
FIGURA 14 VISTA DE CUENTA APLICACIÓN MÓVIL	26
FIGURA 15: INICIO DE SERVICIOS LUNA EN ENTORNO LOCAL	29
FIGURA 16: INICIO DE APLICACIÓN MÓVIL	30
FIGURA 17: ESCALADO DE MODELOS POR DIMENSIÓN Y COMPUESTOA	32
FIGURA 18: COMPARATIVA DE MODELOS DE REDES NEURONALES RESPECTO A NÚMERO DE PARÁMETROS VS PRECISIÓN (IMAGENET)	33
FIGURA 19: COMPARATIVA DE MODELOS DE REDES NEURONALES RESPECTO A CANTIDAD DE OPERACIONES PUNTO FLOTANTE POR SEGUNDO (FLOPS) VS PRECISIÓN (IMAGENET)	34
FIGURA 20: DIAGRAMA DE DISTRIBUCIÓN DE DATASETS UTILIZANDO K-FOLD CROSS VALIDATION	35
FIGURA 21: CANTIDAD Y PROPORCIÓN DE MUESTRAS BENIGNAS VERSUS MALIGNAS PRESENTES EN EL DATASET SIIM-ISIC	36
FIGURA 22: RESULTADO PREDICCIONES CASOS BENIGNOS Y MALIGNOS DE LUNANET SOBRE DATASET SIIM-ISIC	37
FIGURA 1: DISTRIBUCIÓN DE ESPECIALISTAS DERMATÓLOGOS EN TERRITORIO CHILENO AÑO 2020 .....	3
FIGURA 2: TIEMPOS DE ESPERA PARA PACIENTES DE CÁNCER, DESDE LA DERIVACIÓN HASTA EL PRIMER TRATAMIENTO .....	4
FIGURA 3: EJEMPLO DE DESBALANCE DE CLASES EN UN SISTEMA DE DETECCIÓN DE FRAUDE (FRAUDE VS NO FRAUDE).....	9
FIGURA 4: EJEMPLO DE SUBMUESTREO (UNDER-SAMPLING) Y SOBRE MUESTREO (OVER-SAMPLING) SOBRE UN DATASET.....	9
FIGURA 5: INTERFACES DE APLICACIÓN UMSKINCHECK .....	12
FIGURA 6: INTERFACES APLICACIÓN MIISKIN .....	13
FIGURA 7: INTERFACES APLICACIÓN MEDGIC .....	13
FIGURA 8: INTERFACES APLICACIÓN SKINVISION.....	14
FIGURA 9: DIAGRAMA :DE A:RQUITECTURA LUNA .....	18
FIGURA 10: COMPONENTES PRESENTES EN RABBITMQ.....	20
FIGURA 11: FLUJO DE MENSAJES EN RABBITMQ PARA ARQUITECTURA LUNA .....	21
FIGURA 12: VISTA DE ANÁLISIS APLICACIÓN MÓVIL .....	24
FIGURA 13 : VISTA DE RESULTADOS APLICACIÓN MÓVIL .....	25
FIGURA 14 VISTA DE CUENTA APLICACIÓN MÓVIL .....	26
FIGURA 15: INICIO DE SERVICIOS LUNA EN ENTORNO LOCAL .....	29
FIGURA 16: INICIO DE APLICACIÓN MÓVIL .....	30

FIGURA 17: ESCALADO DE MODELOS POR DIMENSIÓN Y COMPUESTOA .....	32
FIGURA 18: COMPARATIVA DE MODELOS DE REDES NEURONALES RESPECTO A NÚMERO DE PARÁMETROS VS PRECISIÓN (IMAGENET) .....	33
FIGURA 19: COMPARATIVA DE MODELOS DE REDES NEURONALES RESPECTO A CANTIDAD DE OPERACIONES PUNTO FLOTANTE POR SEGUNDO (FLOPS) VS PRECISIÓN (IMAGENET) .....	34
FIGURA 20: DIAGRAMA DE DISTRIBUCIÓN DE DATASETS UTILIZANDO K-FOLD CROSS VALIDATION .....	35
FIGURA 21: CANTIDAD Y PROPORCIÓN DE MUESTRAS BENIGNAS VERSUS MALIGNAS PRESENTES EN EL DATASET SIIM-ISIC .....	36
FIGURA 22: RESULTADO PREDICCIONES CASOS BENIGNOS Y MALIGNOS DE LUNANET SOBRE DATASET SIIM-ISIC.....	37

## RESÚMEN

La detección temprana del cáncer de piel es de vital importancia para prevenir su ramificación, evitando un tratamiento agresivo o en el peor caso, un diagnóstico mortal. La exploración de la piel ayudada de la dermatoscopia ayuda a detectar estas lesiones. Una vez detectada la posible presencia de melanoma o carcinoma, un especialista dermatólogo toma una muestra (biopsia) para un análisis exhaustivo y realizar un diagnóstico

En Chile, la cantidad de especialistas dermatólogos es baja en comparación a la cantidad de habitantes, es más, la distribución de ellos por zona del país se presenta de forma desproporcionada. Esto se traduce finalmente en una problemática de escasez de especialistas y provoca la necesidad de optimizar todos los recursos disponibles para atender a los pacientes que necesiten de una atención.

El uso de inteligencia artificial para el análisis de imágenes tiene el potencial de convertirse en una gran ayuda para la detección temprana del cáncer a la piel, específicamente del melanoma y carcinoma. El análisis automático de imágenes provenientes de dermatoscopias presenta dos tareas principales: La detección de características propias de las enfermedades y la clasificación de la enfermedad según esas características.

Este documento propone una forma de integrar el análisis automático de imágenes médicas (dermatoscopias) para la detección de melanoma en un flujo donde participe tanto el paciente como los profesionales.

## **ABSTRACT**

Early detection of skin cancer is of vital importance to prevent its ramification, avoiding aggressive treatment or in the worst case, a fatal diagnosis. Skin examination aided by dermoscopy helps to detect these lesions. Once the possible presence of melanoma or carcinoma is detected, a dermatologist takes a sample (biopsy) for a thorough analysis and diagnosis.

In Chile, the number of dermatologist specialists is low in comparison to the number of inhabitants, moreover, the distribution of them by region of the country is disproportionate. This translates into a problem of shortage of specialists and causes the need to optimize all available resources to serve patients in need of care.

The use of artificial intelligence for image analysis has the potential to become a great aid in the early detection of skin cancer, specifically melanoma and carcinoma. The automatic analysis of dermoscopy images has two main tasks: the detection of disease-specific features and the classification of the disease according to those features.

This paper proposes a way to integrate the automatic analysis of medical images (dermoscopies) for melanoma detection in a flow involving both the patient and the professionals.



# 1 INTRODUCCIÓN

## 1.1 ¿Qué es el cáncer de piel, cuál es su impacto en el mundo y en Chile?

El cáncer de piel es el crecimiento fuera de control de células anormales en la epidermis, la capa más expuesta de la piel, causado por un daño irreparable al ADN lo que provoca mutaciones. Estas mutaciones llevan a las células de la piel a multiplicarse rápidamente y formar tumores malignos. Según la fundación Skin Cancer los principales tipos de cáncer de la piel son carcinoma de células basales, carcinoma de células escamosas, melanoma maligno y carcinoma de células de Merkel (Skin Cancer Foundation, 2021). El melanoma abarca menos del 1% de todos los tipos de cáncer de piel, sin embargo, es el más letal de todos, siendo el causante de la mayoría de las muertes por cáncer de piel (American Cancer Society, 2019). Una de las principales causas de esta enfermedad es la exposición a fuentes de rayos ultravioletas (rayos UV) y para poder prevenir esta enfermedad es necesario minimizar la exposición a ellas, ya sean fuentes naturales como el sol o artificiales como camas de bronceado. Otros factores de riesgo más importantes del melanoma son presentar lunares de grandes dimensiones, formas y colores atípicos, tener piel clara, pecas o cabello claro, que alguien en el historial familiar haya padecido melanoma, haber padecido de melanoma u otros cánceres de piel o ser de edad avanzada (American Cancer Society, 2019).

Para evaluar la extensión de la enfermedad, los doctores suelen utilizar una escala (Tabla 1) que define el progreso del cáncer (NIH, 2015) que se resume en la Tabla 1. Se observa un 98% de supervivencia en los melanomas diagnosticados in situ o localizado, cae a 64% cuando el melanoma es regional (Melanoma Research Alliance, 2020). Por lo tanto, la detección temprana del melanoma es un desafío importante.

Tabla 1: Etapas del desarrollo del cáncer utilizadas por los dermatólogos y su tasa de supervivencia

Definición	Comentario	Tasa de supervivencia a cinco años
In situ	Hay células anormales presentes, pero no se han diseminado al tejido cercano.	98%
Localizado	El cáncer se ha limitado al lugar en donde empezó, sin indicios de haberse diseminado.	98%
Regional	El cáncer se ha extendido a estructuras cercanas, como a los ganglios linfáticos, a tejidos u órganos.	64%

Distante	El cáncer se ha diseminado a partes distantes del cuerpo.	23%
----------	---	-----

## 1.2 Primera problemática en la salud pública chilena: la escasez de especialistas dermatólogos

Según datos del gobierno de Chile, aproximadamente 21.000 pacientes dermatológicos se encuentran registrados al año 2018 en el sector público. Al año 2021, este número va creciendo (+110% en los últimos 5 años)

### Cantidad y distribución de especialistas

Según registros de Estados Unidos, en el año 2017 existían 3.4 especialistas dermatólogos por cada 100.000 habitantes (Glazer y Rigel, 2017). En Francia, se registró en el año 2018 un total aproximado de 5.9 especialistas por cada 100.000 habitantes (Profil Médecin, 2020).

Para el año 2020 se encuentran registrados 576 especialistas dermatólogos registrados en el Registro Nacional de Prestadores Individuales de Salud (Superintendencia de Salud, 2021). Considerando la cantidad de habitantes del país en este año (Datosmacro, 2021), se calcula un total de 3.02 especialistas dermatólogos por cada 100.000 habitantes. Además, de la totalidad de especialistas mencionada anteriormente, la Figura 1 presenta la distribución de especialistas dermatólogos por región. Es importante notar que a pesar de que existen especialistas que no han informado su ubicación de trabajo, la concentración se acentúa en las regiones Metropolitana y de Coquimbo, para el resto de regiones, la cantidad de especialistas por habitantes es notablemente menor.



Figura 1: Distribución de especialistas dermatólogos en territorio chileno año 2020

En Chile, se han diseñado múltiples estrategias para mejorar el diagnóstico precoz del cáncer de piel. Estas incluyen identificación de poblaciones de alto riesgo para aplicar, en este grupo, programas especiales de educación y de revisión precoz, y en la población general, enseñar y estimular el autoexamen de la piel, aumentar la frecuencia del examen cutáneo completo en el contexto de una evaluación médica rutinaria.

En la población general, los pacientes juegan un rol importante en el diagnóstico del melanoma. Se ha reportado que más del 72% de los melanomas son detectados por el paciente o sus familiares.

Alertar a la población de los signos sospechosos de melanoma contribuye al diagnóstico a tiempo de las lesiones. Por ello, existe una necesidad de intensificar los esfuerzos para disminuir la mortalidad por melanoma en Chile mediante campañas de alerta poblacional, incentivando el autocuidado en relación a la exposición solar, el autoexamen periódico y el reconocimiento y consulta precoz por lesiones. Además, se debería reforzar durante la educación médica la participación en los esfuerzos de educación de la población, el valor del examen completo de la piel, el reconocimiento de las manifestaciones iniciales de un melanoma cutáneo, y la identificación de grupos de alto riesgo para su derivación oportuna a centros especializados (Sabatini-Ugarte et al., 2018).

### 1.3 Segunda problemática en la salud pública chilena: la derivación de pacientes y tiempos de espera

Los profesionales en la salud pública chilena destacan el problema de la derivación de pacientes como factor agravante en el fenómeno de saturación. Si bien no existen datos cuantitativos del fenómeno, actores del sector perciben que existe un mal uso o una mala priorización del tiempo de los dermatólogos especialistas debido a errores de derivación (Anexo A). Este fenómeno agrava la escasez de recursos humanos. En Chile, según

registros del gobierno del año 2018, existían un total de 21.000 pacientes dermatológicos registrados en el sector público. En el sistema de salud pública de la Región de Los Ríos se registró un total de 1200 pacientes, sin embargo, se contaba solamente con tres profesionales dermatológicos, ubicados en el Hospital Base de Valdivia. En el año 2020 siete profesionales dermatológicos se encuentran registrados en la Región de Los Ríos.

El sistema de salud pública de Inglaterra (Public Health England, 2014) estima que los tiempos de espera de los pacientes con cáncer debería seguir el siguiente estándar (Figura 2):

- Un máximo de dos semanas de espera desde una derivación urgente del médico de cabecera (según las reglas de espera de dos semanas) por sospecha de cáncer hasta la primera evaluación hospitalaria.
- Un máximo de un mes (31 días) de espera, desde el diagnóstico/decisión para tratar hasta el primer tratamiento definitivo.
- Un máximo de dos meses (62 días) de espera desde una derivación urgente al médico de cabecera (según las reglas de espera de dos semanas) por sospecha de cáncer hasta el primer tratamiento definitivo (31 días para niños).
- Un máximo de dos meses (62 días) de espera a partir de la decisión de un consultor de actualizar la urgencia de una derivación no TWW de un paciente que sospecha que tiene cáncer al primer tratamiento definitivo.

De lo anterior, se puede comprender la urgencia de la correcta derivación de un paciente con cáncer para poder tener su tratamiento, tomando en cuenta que los especialistas (dermatólogos en el caso del cáncer a la piel) son escasos

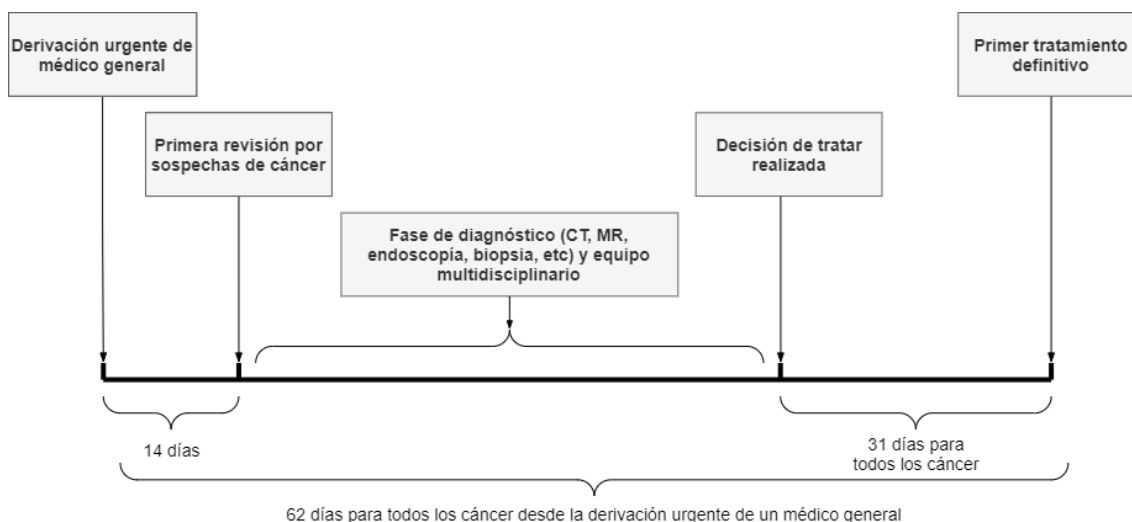


Figura 2: Tiempos de espera para pacientes de cáncer, desde la derivación hasta el primer tratamiento

## 1.4 Oportunidad: mejorar la detección temprana del cáncer de la piel y reducir los errores en la derivación de pacientes utilizando tecnologías de información

### Telemedicina en Chile

En sus inicios, la telemedicina se utilizó para llegar a poblaciones de difícil acceso o aisladas geográficamente para poder mejorar la accesibilidad. Hoy se piensa en además de mejorar la accesibilidad, mejorar la eficiencia y eficacia.

Existen antecedentes de proyectos relacionados a la telemedicina mostrados en la Tabla 2.

Tabla 2: Anteriores proyectos relacionados a la telemedicina en Chile

Año	Proyecto
1998	La Clínica Las Condes establece conexión con la Isla Juan Fernández vía email y modem.
2003	Comienza Proyecto Argonauta, el cual busca llevar la telemedicina en la Antártica
2004	Ministerio de Salud implementa servicio de Tele-electrocardiología a nivel nacional
2005	Ministerio de Salud crea el Centro de Asistencia Remota en Salud en la Región Metropolitana
2009	<ul style="list-style-type: none"><li>- Entra en operaciones Programa Galileo de Telemedicina Cardiovascular integrada para los Servicios de Salud de Talcahuano y Arauco</li><li>- El Ministerio de Salud comienza la implementación de Teledermatología.</li></ul>
2012	El Ministerio de Salud implementa teleradiología y teleasistencia a través de dispositivos móviles.
2017	Ministerio de Salud crea Red de Referencia de Telemedicina en Ataque Cerebro Vascular (ACV) en el Servicio de Salud Metropolitano Sur, luego de que en el año

	2016 se realizó la Primera Teletrombolisis en ACV con el Hospital El Pino
--	---

## **Tele dermatología**

En el año 2009 se comenzó una estrategia para facilitar el acceso a la dermatología desde la atención primaria, donde un profesional de la salud primaria envía fotografías y antecedentes clínicos a un especialista dermatólogo, quien evalúa los antecedentes y entrega orientación diagnóstica e indicaciones. Hasta el año 2017 se ha implementado en 29 servicios de salud del país y se han realizado 79801 consultas de esta manera.

Este trabajo de título busca aportar a la salud nacional, mostrando que es posible desarrollar un sistema de análisis automatizado y pre-diagnóstico de cáncer a la piel.

### **1.5 Objetivo general del proyecto de título: Clasificar imágenes médicas para el pre-diagnóstico de enfermedades de la piel**

Objetivo general:  
Clasificar imágenes médicas para el pre-diagnóstico de enfermedades de la piel, en particular el melanoma, a través del diseño e implementación de un prototipo de arquitectura de software y del entrenamiento y evaluación de modelos de redes neuronales.

#### **Objetivos específicos**

- Estructurar información del estado del arte sobre enfermedades de la piel y telemedicina en el contexto chileno y enfoques computacionales con redes neuronales para el análisis de imágenes médicas dermatológicas.
- Diseñar e implementar un prototipo de arquitectura de software que permita a un usuario tomar fotos de lunares y recibir retroalimentaciones generadas por un modelo de clasificación automática.
- Aplicar una metodología experimental de clasificación de imágenes que permita entrenar y evaluar modelos de redes neuronales, comparando su rendimiento y evaluando los factores de errores.

## 2 DATOS Y PRE PROCESAMIENTO

### 2.1 Datasets disponibles para la clasificación de imágenes dermatológicas

#### SIIM-ISIC

La Sociedad para Imagenología Informática en Medicina (SIIM) junto con la Colaboración Internacional de Imagenología de la Piel (ISIC) han unido fuerzas para mejorar los diagnósticos de melanoma.

Este dataset contiene la mayor colección disponible públicamente de imágenes dermatoscópicas de lesiones a la piel con control de calidad. Es por esto que fue escogido para el entrenamiento de la red neuronal presente en LUNA.

### 2.2 Características de los datos utilizados

El dataset escogido (SIIM-ISIC) consta de una gran colección de imágenes de lesiones en la piel. Contiene un total de 88251 archivos y ocupa un tamaño de 116.16 GB. Las imágenes se encuentran en formato JPEG y TFRecord, con una resolución de 1024 x 1024 píxeles. Además, contiene metadatos complementarios para cada una junto con la clasificación de la lesión (“melanoma” o “no melanoma”) en formato CSV.

Los archivos CSV se detallan en la Tabla 3 y su contenido en la Tabla 4.

Tabla 3: Lista de archivos CSV con metadatos de imágenes presentes en el dataset SIIM-ISIC

Archivo	Descripción
train.csv - the training set	Dataset de entrenamiento
test.csv - the test set	Dataset de test

Tabla 4: Descripción de las columnas presentes en archivos CSV de metadatos

Columna	Descripción
image_name	Identificador único de la imagen. Corresponde al nombre de archivo
patient_id	id único del paciente al que corresponde la imagen

sex	Sexo del paciente
age_approx	Edad aproximada del paciente a la fecha de la captura de la imagen
anatom_site_general_challenge	Ubicación de la imagen en el cuerpo del paciente
diagnosis	Información detallada del diagnóstico (sólo para entrenamiento)
benign_malignant	Indicador de malignidad de la lesión presente en la imagen
target	Versión binaria (0 o 1) de la malignidad de la lesión

### 2.3 Clases desbalanceadas

En los datasets del mundo real, tiende a ocurrir que aparece un número de muestras de una clase dada de manera subrepresentada (ver Figura 3). Este es un problema ampliamente conocido, denominado “problema de clases desbalanceadas” el cual consta en aprender un concepto desde clases que presentan un pequeño número de muestras (Lemaître et al 2017) frente a otras.

El problema de clases desbalanceadas ha estado presente en diversas áreas tales como administración de telecomunicaciones, bioinformática, detección de fraude o diagnósticos médicos y ha sido considerado dentro de los 10 peores problemas en el área de la minería de datos y reconocimiento de patrones.

La presencia de clases desbalanceadas afecta de forma importante el proceso de aprendizaje de un modelo, lo que provocará que exista un sesgo a favor de la clase con el número mayor de muestras presentes en el dataset.



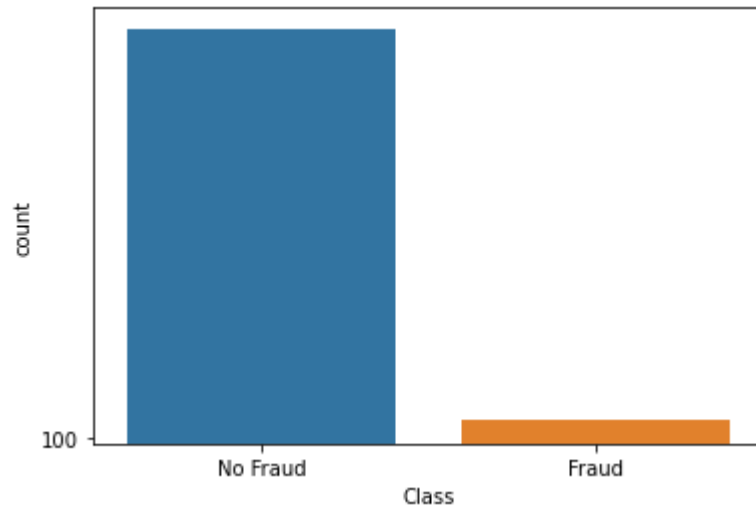


Figura 3: Ejemplo de desbalance de clases en un sistema de detección de fraude (Fraude vs No Fraude).

La forma común de lidiar con el desbalance de clases es equilibrar las muestras presentes aplicando un submuestreo (eliminar muestras) de las clases mayoritarias y un sobremuestreo (generar muestras) de las clases minoritarias (Figura 4).

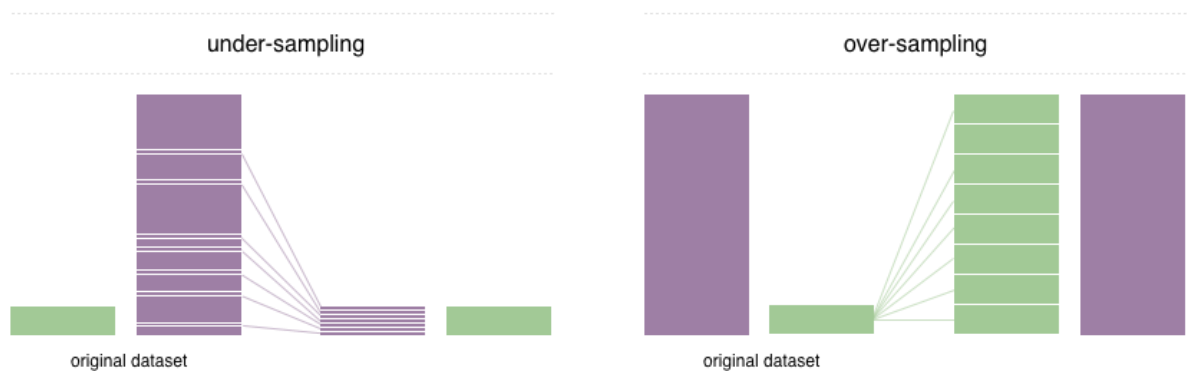


Figura 4: Ejemplo de submuestreo (under-sampling) y sobre muestreo (over-sampling) sobre un dataset

Aunque esto ayuda a balancear el set de datos, a la vez puede provocar un sobreajuste en torno a las clases minoritarias dada la duplicidad de información y pérdida de información al quitar muestras de las clases mayoritarias.

En la práctica, para poder abordar el problema del desbalance de clases, se utilizó una librería de Python llamada “Imbalanced Dataset Sampler” (<https://github.com/ufoyim/imbalanced-dataset-sampler>), la cual permite obtener muestras de forma “balanceada”.

## 2.4 Aumentación de datos

Muchas veces es necesario incrementar la cantidad de muestras en un set de datos y no siempre se pueden crear de forma sintética. La aumentación de datos consiste en generar nuevas muestras a partir de otras existentes, aplicando pequeñas transformaciones o utilizando modelos de deep learning para poder generar nuevos.

En el entrenamiento del modelo utilizado para este proyecto y con la finalidad de no caer en un sobreajuste, por cada una de las iteraciones de los k-folds, se ha aplicado las siguientes transformaciones por muestra:

- **Agregar cabello aleatoriamente:** De forma aleatoria se agregan vellos individuales sobre la imagen.
- **Recorte aleatorio:** Se recorta la imagen de forma aleatoria y la imagen resultante se escala al tamaño indicado.
- **Volteo horizontal aleatorio:** Dada una probabilidad, aleatoriamente se decide si voltear o no la imagen de forma horizontal
- **Volteo vertical aleatorio:** Dada una probabilidad, aleatoriamente se decide si voltear o no la imagen de forma vertical
- **Fluctuación de color aleatoria:** Cambia de forma aleatoria los valores de hue, brillo, contraste y saturación de la imagen.
- **Vista de microscopio:** Corta la imagen de forma circular, dejando un fondo negro simulando una vista de microscopio.

Estas transformaciones permiten que por cada iteración se genere un set de imágenes único e irrepetible.

### **3 INTEGRACIÓN DEL MODELO DE CLASIFICACIÓN EN UN PROTOTIPO DE APLICACIÓN MÓVIL LUNA**

Con el propósito de mostrar un caso de uso de cómo utilizar tecnologías de inteligencia artificial para desarrollar una solución informática que permita ayudar y agilizar procedimientos automatizados de detección y pre-diagnóstico de melanoma, se ha implementado un prototipo llamado LUNA. El escenario planteado es el de una entidad del sector de salud que necesite agilizar el proceso de detección de melanoma, analizando imágenes a partir de fotografías de celulares de sus doctores. A continuación, en este capítulo se muestran aplicaciones que intentan abordar este caso de uso, junto con la descripción de la plataforma LUNA.

#### **3.1 Aplicaciones similares a LUNA**

##### **UMSkinCheck**

UMSkinCheck (Figura 5) es una aplicación móvil que enseña al usuario a cómo realizar un autoexamen de su cuerpo en búsqueda de lunares y manchas sospechosas. Esta aplicación contiene una guía completa con instrucciones de localización de lunares, literatura sobre el cáncer de piel y como poder identificar lunares potencialmente cancerígenos.

Permite capturar y almacenar fotografías tomadas por el usuario, permitiendo realizar un mapeo de estas fotografías con áreas del cuerpo en un diagrama.

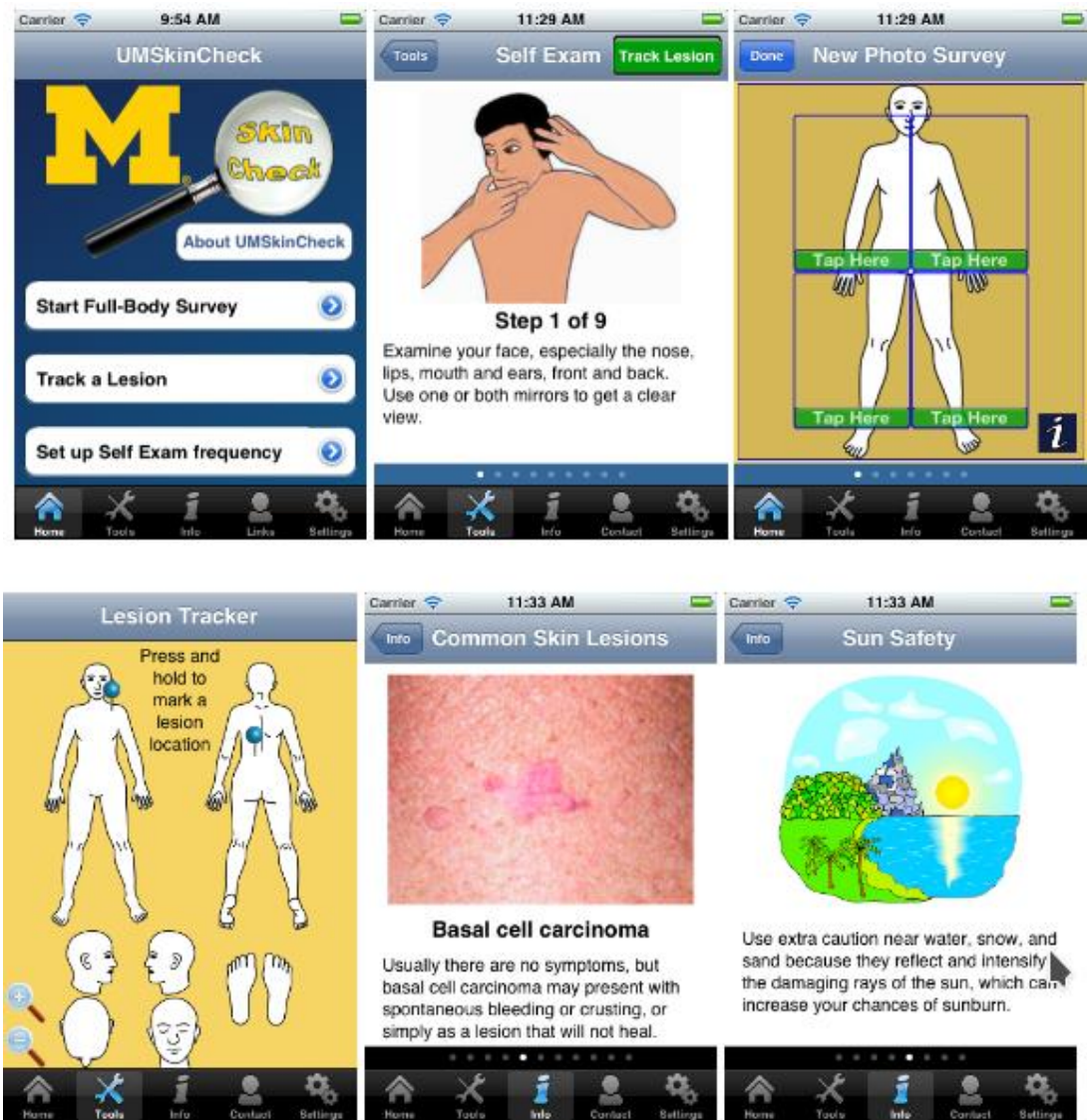


Figura 5: Interfaces de aplicación UMSkinCheck

### Miiskin

Miiskin (Figura 6) permite funciones similares a UMSkinCheck, pero con una modalidad básica y premium.

Las funciones básicas permiten realizar un seguimiento de hasta 3 lunares y marcas en el cuerpo, permite también comparar imágenes de lunares a través del tiempo y envía recordatorios para realizar escaneos.

La versión premium permite realizar seguimiento ilimitado de lunares y marcas existentes en el cuerpo, detección automática de lunares y marcas a partir de fotografías grandes, mapeo automático de lunares y marcas además de proteger las fotografías con un pin de seguridad.

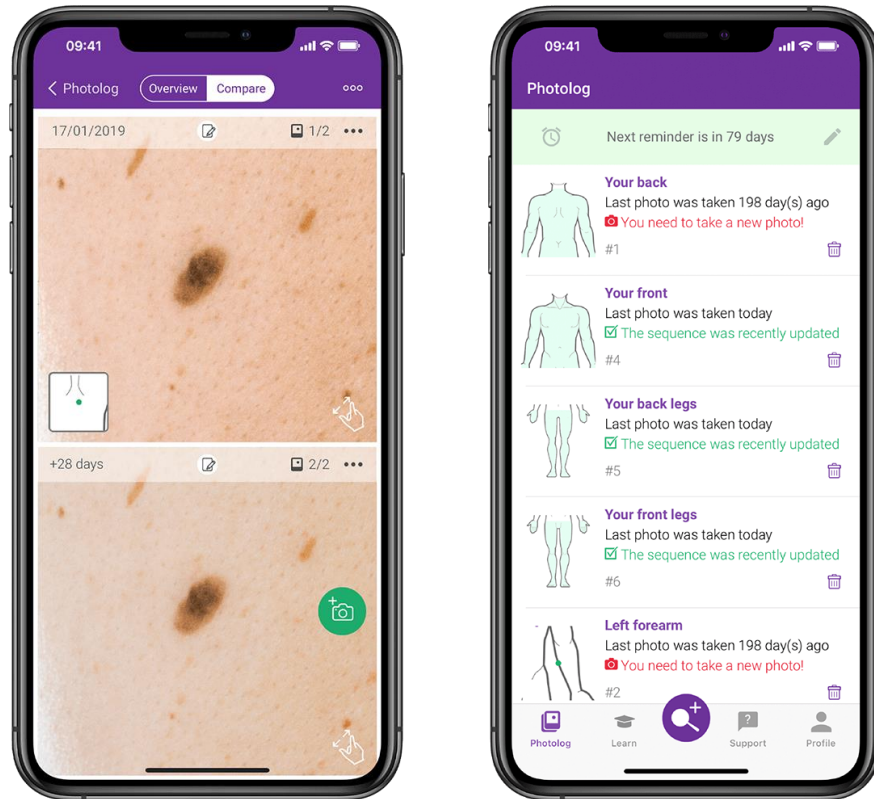


Figura 6: Interfaces aplicación Miiskin

## Medgic

Medgic (Figura 7) permite tomar fotografías y ser enviadas para revisión por una inteligencia artificial: Los resultados se envían a un servidor y quedan a la espera de ser revisados por profesionales (ver figura x).

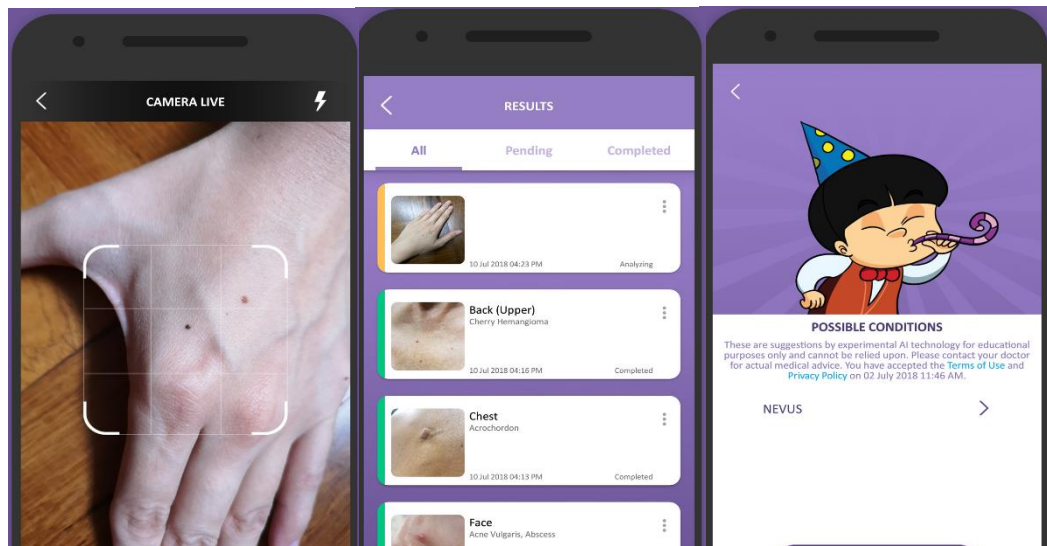


Figura 7: Interfaces aplicación Medgic

## SkinVision

SkinVision (Figura 8) es una aplicación de pago y ofrece una combinación de todas las aplicaciones anteriores: Permite realizar seguimiento por fotografías tomadas por el usuario, escaneo de imágenes utilizando inteligencia artificial, almacenamiento de fotografías e historial a través del tiempo, alertas en caso de detección de mediano o alto riesgo, análisis por especialistas en caso de detectar mediano o alto riesgo.

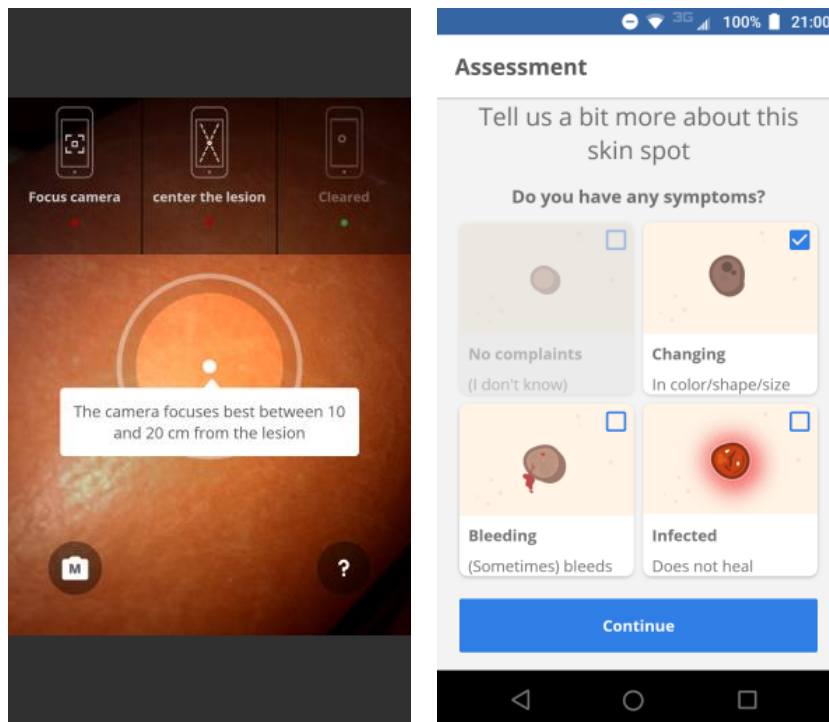


Figura 8: Interfaces aplicación SkinVision

De las aplicaciones encontradas e investigadas, sólo Medgic y Skinvision tienen similitudes con LUNA, al contar con un análisis automatizado ayudado de la inteligencia artificial.

## **3.2 Requisitos funcionales y no funcionales**

### **Requisitos funcionales**

#### **Aplicación móvil**

- REQF001: La aplicación móvil debe permitir el acceso sólo a usuarios registrados en el sistema
- REQF002: La aplicación móvil debe tener 4 secciones principales:
  - Login
  - Análisis de fotografías
  - Vista de solicitudes de análisis
  - Vista de detalle de análisis realizados
- REQF003: La vista de análisis de fotografías debe permitir seleccionar una fotografía desde la galería, un campo de texto que permita agregar una descripción (opcional) y un botón para poder enviarla a analizar, luego debe aparecer un registro de esa solicitud de análisis en la sección de solicitudes de análisis
- REQF004: La vista de solicitudes de análisis debe aparecer una tabla con todas las solicitudes realizadas, las columnas de esa tabla corresponden a los siguientes datos:
  - Fecha y hora de la solicitud.
  - Descripción en caso de existir.
  - Estado de la solicitud de análisis (En espera, Completado).
  - Un botón que permita ver el detalle de la solicitud.Al presionar el botón de detalle de la solicitud, se mostrará la vista de detalle de análisis. Además debe permitir filtrar los análisis por fecha.
- REQF005: La vista de detalle de solicitud de análisis debe mostrar la siguiente información:
  - Fecha y hora de la solicitud.
  - Una sección que permita mostrar u ocultar la imagen analizada, la imagen debe estar oculta de forma predeterminada.
  - Una sección que indique el estado del análisis (En espera, Completado).
  - Fecha y hora del análisis en caso de haberse realizado.
  - Resultado del análisis en caso de haberse realizado.

#### **Seguridad**

- REQF006: El acceso al sistema sólo está permitido para usuarios registrados
- REQF007: El registro de usuarios al sistema queda a cargo de un administrador

### **Requisitos no funcionales**

#### **Seguridad de los datos**

- REQNF001: Los registros de resultados de análisis no deben mostrar la información del usuario.



### 3.3 Arquitectura de software LUNA

Con el propósito de diseñar una arquitectura que sea escalable y capaz de soportar una gran cantidad de tráfico, se escogió un modelo distribuido basado en microservicios. Los componentes (Tabla 5) presentes deben permitir la autenticación de usuarios, el envío y procesamiento de imágenes, almacenamiento de ellas y posterior visualización de los resultados de los análisis.

Tabla 5: Descripción de componentes de arquitectura propuesta

Componente/servicio	Tecnología	Descripción
Aplicación Móvil	IONIC/React	Es el punto en donde el usuario interactúa con la plataforma. A través de ella se podrá solicitar análisis y ver sus resultados.
API REST	NodeJS	Punto de entrada al sistema, a través de ella, la aplicación móvil podrá autenticarse, solicitar análisis de imágenes y ver resultados de análisis anteriores.
Base de datos administrativa	PostgreSQL	Almacena los registros necesarios para la administración de usuarios
Almacenamiento de imágenes y resultados	MongoDB	Almacena las imágenes junto con los resultados de los análisis
Gestión de solicitudes de análisis de fotografías	RabbitMQ	Ayuda a gestionar la carga de información, a través de mensajes y colas. Permite que varios procesos consumidores puedan trabajar de forma paralela
Análisis de imágenes	Python	Procesos que pueden ejecutarse en forma paralela (worker), cuya tarea es recibir información de RabbitMQ, analizar las

		imágenes y enviar los resultados a través del mismo
Registro de resultados	Python	Procesos que pueden ejecutarse en forma paralela (worker), cuya tarea es obtener información de las imágenes junto con los resultados de análisis, asociarlos a un usuario y guardarlos en las bases de datos correspondientes

### Procesamiento de imágenes

Es importante que el análisis de imágenes se realice de forma asíncrona y debe existir un mecanismo que permita y asegure el análisis de cada imagen enviada a análisis. Esto se puede lograr generando un encolamiento de estos procesos para que puedan ser ejecutados a medida que los recursos se encuentren disponibles.

Los servicios que componen la arquitectura propuesta (ver Figura 9) son:

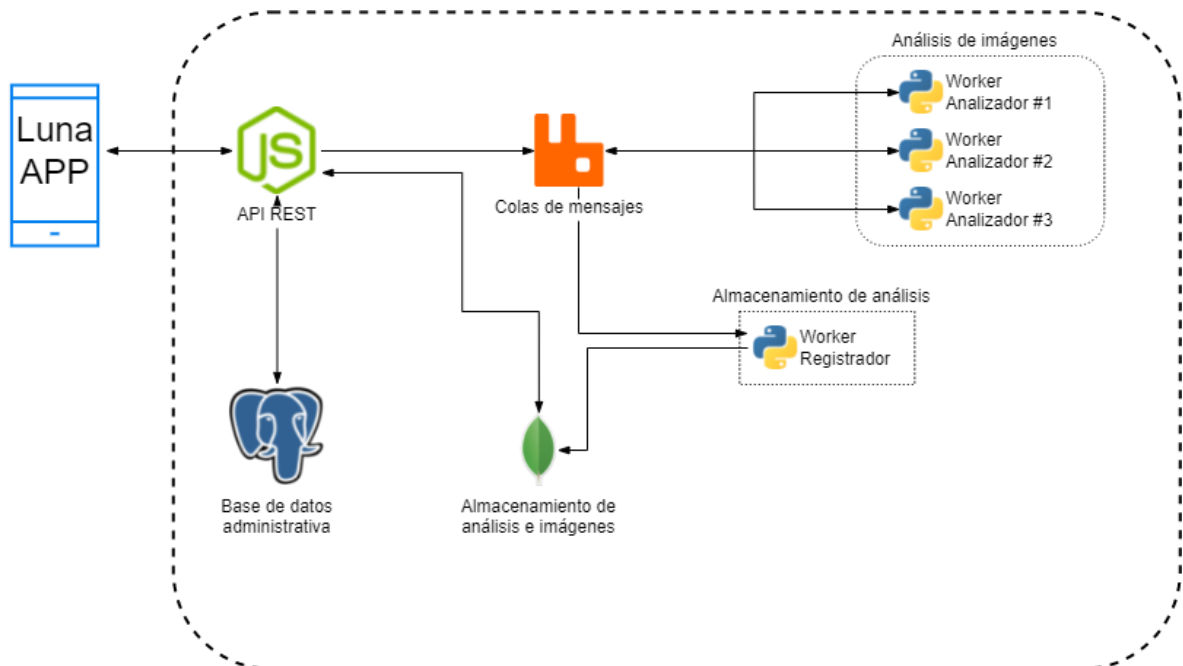


Figura 9: diagrama de arquitectura LUNA

A continuación, se detalla cada uno de los componentes que forman parte de la arquitectura LUNA

### Colas de mensajes

Existen tecnologías que permiten la implementación de colas de mensajes. RabbitMQ es una de ellas y es la que se escogió para actuar como bróker de mensajería.

Un broker de mensajería es un software en el cual se definen colas para que otras aplicaciones puedan conectarse y así poder enviar y recibir mensajes. Un mensaje puede contener cualquier tipo de información: un mensaje de texto, información sobre un proceso o una tarea que deba ser iniciada en otra aplicación, la cual puede estar alojada en un servidor distinto. El software que administra las colas almacena los mensajes hasta que una aplicación se conecte y los consuma desde dicha cola. La aplicación que recibe el mensaje lo procesa para efectuar las acciones correspondientes.

Los componentes presentes en RabbitMQ son:

- **Productores:** Entidad que publica mensajes hacia un exchange
- **Exchanges:** Se encargan de recibir los mensajes y en base a reglas de ruteo (bindings) enviarlos a una o varias colas. Existen distintos tipos de exchanges:
  - **Directo:** Un exchange directo envía mensajes basado en un “routing key”, el cual es un atributo agregado a la cabecera del mensaje por parte del productor. Este tipo de exchange leerá la routing key y enviará el mensaje a las colas que presenten un binding key con exactamente el mismo valor.
  - **Por tópicos:** Este tipo de exchange rutea los mensajes a las colas basados en comodines entre la “routing key” del mensaje y un patrón definido en la “binding key” de la cola. Donde sea que calce dicha expresión el mensaje será enviado a esa cola.
  - **Fanout:** Un exchange de tipo fanout copiará un mensaje recibido y lo enviará a todas las colas que estén conectadas a él. No hace distinción entre un mensaje u otro.
- **Colas:** Reciben mensajes desde los exchanges y los almacena hasta que sean consumidos por un consumidor.
- **Consumidores:** Consumen mensajes de colas específicas.

Los componentes descritos y tipos de exchanges pueden ser identificados en la Figura 10 a continuación:

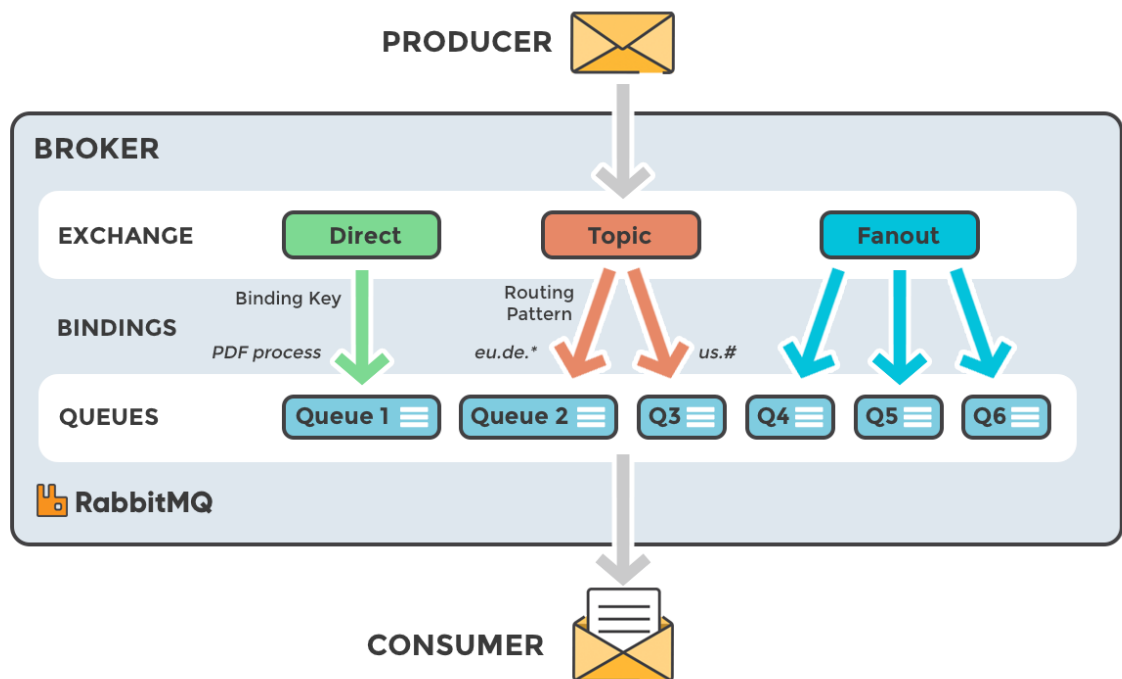


Figura 10: Componentes presentes en RabbitMQ

Para el escenario previsto en la arquitectura propuesta, el flujo de un mensaje en RabbitMQ (ver Figura 11) y sus componentes (Tabla 6) se describe de la siguiente manera:

1. La solicitud de análisis es recibida por la API REST, la cual creará un mensaje y lo enviará al exchange llamado "send\_to\_process".
2. El exchange enviará el mensaje a la cola llamada "process\_pending".
3. Los workers de análisis consumirán los mensajes que se encuentren en la cola "process\_pending", para luego realizar el análisis.
4. Los workers de análisis enviarán el nuevo mensaje con el resultado del análisis al exchange "process\_completed".
5. El exchange "process\_completed" enviará los mensajes recibidos a la cola del mismo nombre.
6. Los workers registradores consumirán mensajes de la cola "process\_completed", extraerán la información y almacenarán los resultados en la base de datos mongoDB.

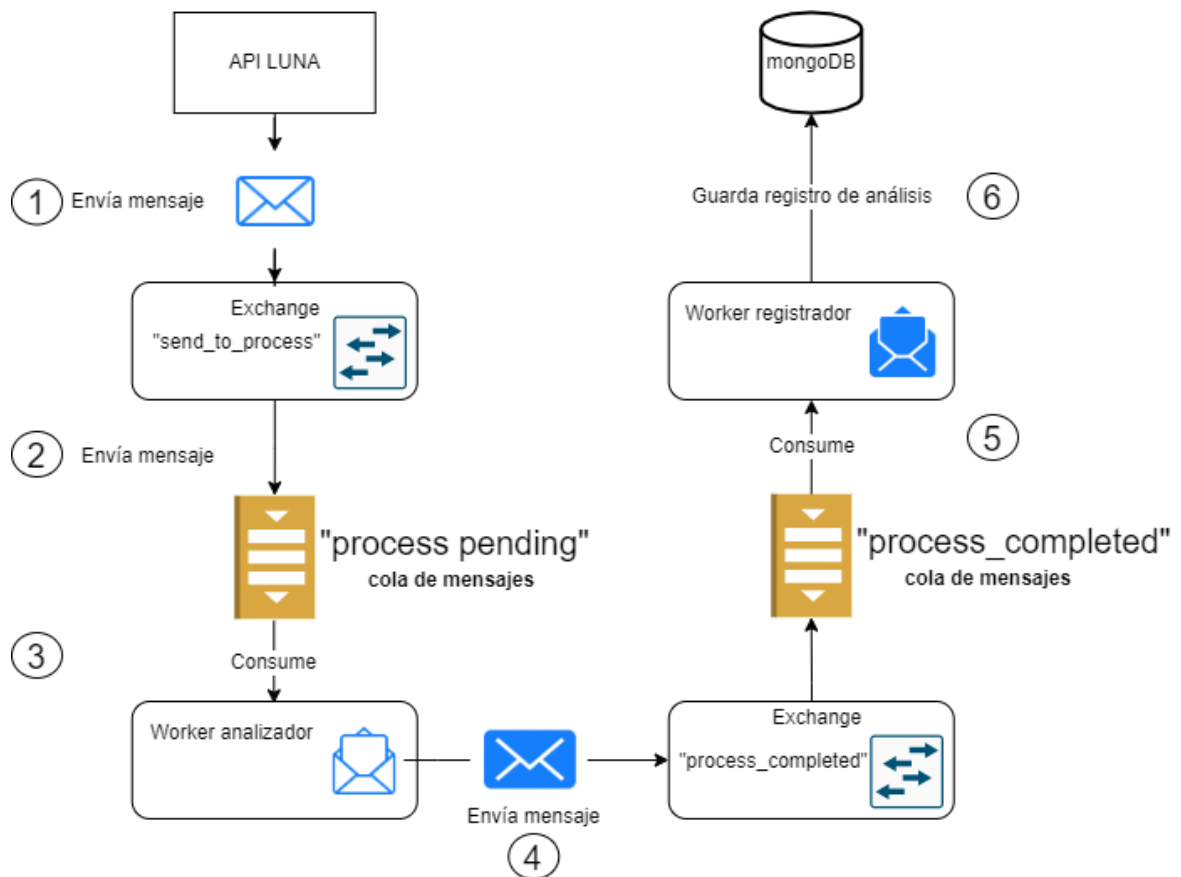


Figura 11: Flujo de mensajes en RabbitMQ para arquitectura LUNA

Tabla 6: Descripción de componentes en RabbitMQ para arquitectura LUNA

Componente	Tipo de componente	Descripción
send_to_process	Exchange - fanout	Recibe los mensajes de solicitudes de análisis y los envía a la cola "process_pending"
process_completed	Exchange - fanout	Recibe los mensajes de análisis realizados y los envía a la cola "process_completed"
process_pending	Cola conectada a exchange send_to_process	Almacena mensajes con imágenes pendientes de análisis

process_completed	Cola Cola conectada a exchange process_completed	Almacena mensajes de análisis realizados
LUNA API	Productor	Recibe solicitudes de análisis por parte del usuario para luego crear un mensaje que contenga la imagen para ser enviado al exchange “send_to_process”
Worker analizador	Consumidor	Consume mensajes de la cola “process_pending” para luego analizar la imagen contenida. El resultado del análisis es enviado como un nuevo mensaje al exchange “process_completed”
Worker registrador	Consumidor	Consume mensajes de la cola “process_completed”, extrae la información del análisis y lo escribe en la base de datos MongoDB

## API REST

Este servicio corresponde a una API REST, donde el usuario podrá autenticarse para poder analizar imágenes, obtener lista de resultados y ver el detalle de un análisis. La autenticación se realiza a través de JWT.

La autenticación se realiza enviando una solicitud HTTP de tipo POST con el nombre de usuario y contraseña. La respuesta de la API serán los datos del usuario y el token de acceso. Este token debe ser incluido en todas las otras consultas realizadas a la API.

A continuación, en la Tabla 7 se listan los endpoints disponibles y una breve descripción

Tabla 7: Endpoints disponibles de la API REST para arquitectura LUNA

Endpoint	Requiere autenticación?	Descripción
/signin	no	Inicio de sesión
/api/auth/signup	si	Creación de cuenta.
/logout	si	Cerrar sesión.
/api/imageProcess	si	Solicitud de análisis de imagen.
/api/me/getAnalysisList	si	Obtención de lista de análisis.
/api/getAnalysis/:id	si	Obtención de detalle de análisis a partir del id.
/api/me	si	Obtención de datos del usuario.

Las fotografías deben ser recibidas en formato string BASE64 por un servicio API REST, el cual validará la autenticación del usuario para luego generar un mensaje que será enviado a un exchange que espera imágenes pendientes de análisis.

El mensaje enviado tendrá la información de la imagen y el identificador del usuario correspondiente siguiendo el siguiente formato:

```
{
  image: string,
  userId: integer,
}
```

## Aplicación Móvil

Se ha desarrollado una aplicación móvil utilizando la tecnología IONIC. IONIC es un framework SDK open source para el desarrollo de aplicaciones móviles híbridas, esto quiere decir que el mismo código fuente se puede utilizar para generar una aplicación

para distintos sistemas operativos como Android e IOS. Se puede ejecutar hasta en un navegador web. IONIC se integra con frameworks de frontend tales como Angular, React y Vue.

Para este proyecto, se ha optado por utilizar la integración con el framework React, ya que de los disponibles (Angular, React, Vue) fue el que permitía desarrollar la aplicación de manera más rápida.

La aplicación consta de tres vistas principales: análisis de imagen, resultados de análisis y cuenta.

La vista de análisis (ver Figura 12) permite tomar una fotografía o seleccionar una de la galería para ser enviada hacia la API para su análisis previa autenticación. Una vez tomada o seleccionada la imagen, se puede enviar a analizar presionando el botón “analizar imagen”. En caso contrario, es posible cancelar la toma de imagen presionando el botón “cancelar”. Es necesario estar autenticado en la aplicación para poder acceder a esta vista.

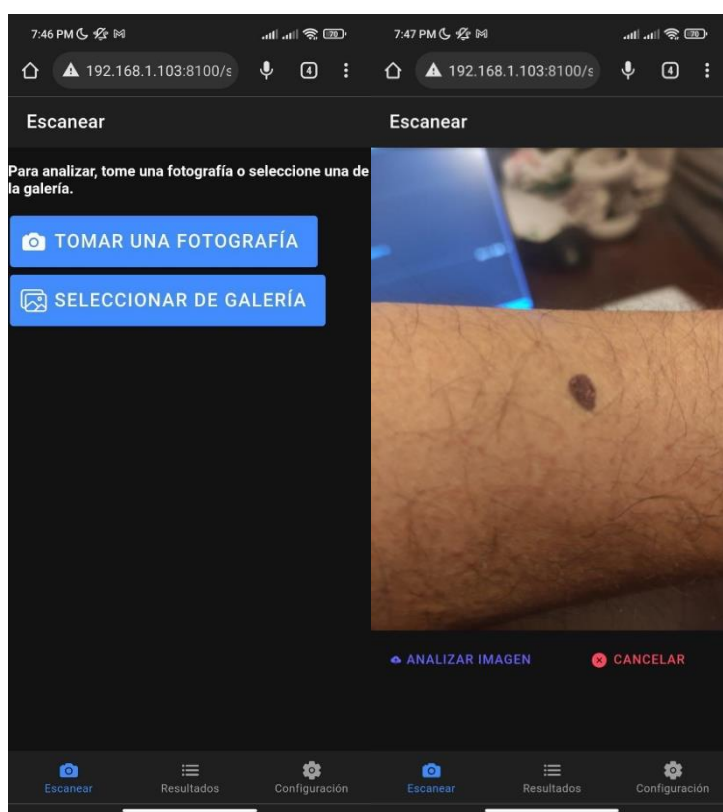


Figura 12: Vista de análisis aplicación móvil



La vista de resultados (ver Figura 13) presenta un listado con todos los análisis realizados por el usuario. El usuario puede presionar sobre un elemento de la lista para poder visualizar el detalle del análisis. Es necesario estar autenticado en la aplicación para poder acceder a esta vista.

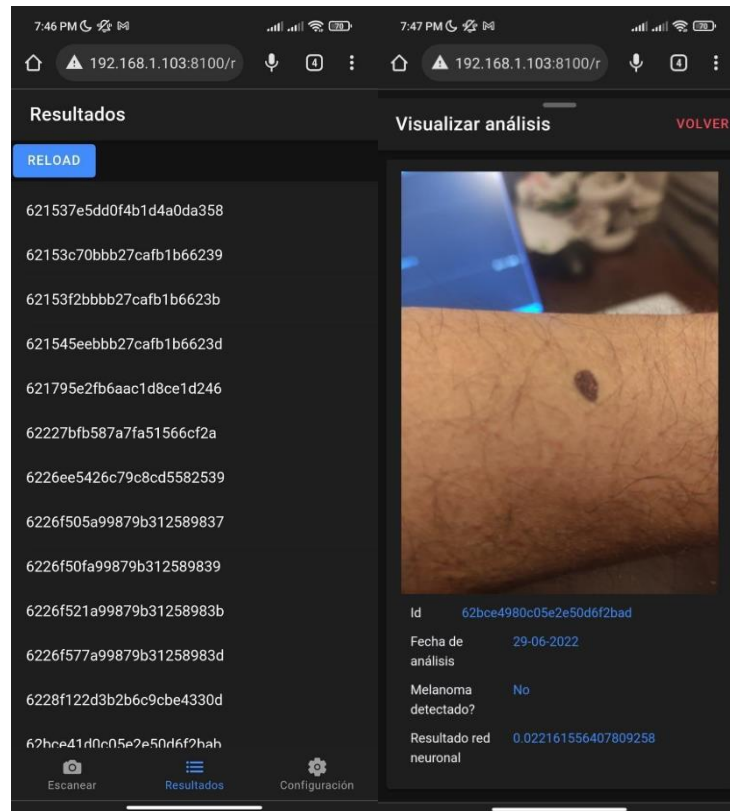


Figura 13 : Vista de resultados aplicación móvil

La vista de cuenta (ver Figura 14) muestra información del usuario autenticado: nombre de usuario, e-mail y roles asignados. También permite iniciar o cerrar sesión.

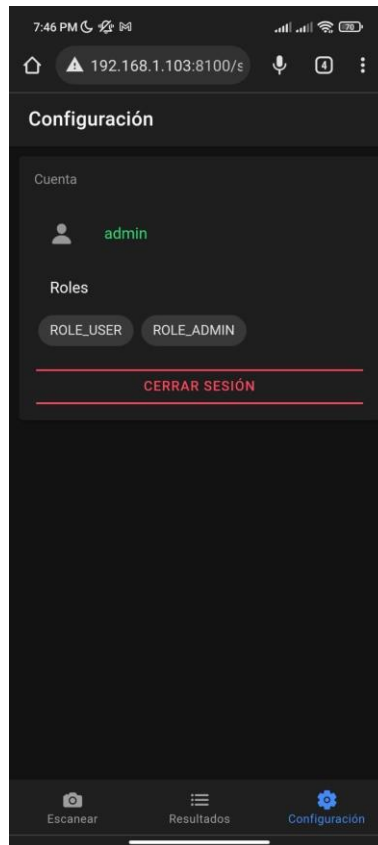


Figura 14 Vista de cuenta aplicación móvil

Un token JWT obtenido tras autenticarse es enviado a la API REST en todas las consultas realizadas. Este token es almacenado temporalmente en la aplicación mientras se encuentre en ejecución.

### Workers analizadores

Los workers analizadores son programas escritos en el lenguaje Python, los cuales permanecen ejecutándose de manera continua, consumiendo mensajes de la cola de imágenes pendientes de análisis. Estos workers leen el mensaje desde la cola, obtienen la imagen, ejecutan el análisis y envían el resultado dentro de un nuevo mensaje hacia exchange de imágenes analizadas. Una vez enviado este mensaje, el worker queda disponible para poder analizar la siguiente imagen. El nuevo mensaje enviado sigue la siguiente estructura:

```
{
  userId: int
  image:string
  creationDate: string
  analysisDate: string
  result: float
  isMelanoma: booleano
}
```

### Workers registradores

Los workers registradores consumen mensajes de la cola de imágenes analizadas y escriben los registros en una base de datos MongoDB. Una vez almacenado el análisis, este worker queda disponible a la espera de más mensajes.

### Base de datos MongoDB

En esta base de datos se almacenan los resultados de análisis de las imágenes enviadas por los usuarios. El formato de los registros es el siguiente:

```
{
  userId: int
  image:string
  creationDate: string
  analysisDate: string
  result: float
  isMelanoma: booleano
}
```

### Base de datos PostgreSQL

En esta base de datos se almacenan los datos de los usuarios, será utilizada por el servicio de API REST al momento de realizar el registro de usuarios y autenticación. Los datos de usuario se almacenan en la tabla “user” (Tabla 8), que contiene la información de id de usuario, nombre y contraseña encriptada. Además, la tabla “role” (Tabla 9) almacena los roles asignados al usuario, estos pueden ser “admin” o “user”. El diseño del mockup no contempla una interfaz administrativa, por ello el registro y asignación de roles a usuarios se debe realizar a través de la API Rest y modificaciones en base de datos.

Las tablas presentes en esta base de datos son:

Tabla 8: Tabla user

Campo	Tipo de dato
id	Entero
username	String
password	String
email	String

Tabla 9: Tabla role

<b>Campo</b>	<b>Tipo de dato</b>
id	Entero
name	String

### **Estructura de archivos del código fuente de LUNA**

La plataforma se desarrolló bajo un repositorio único. El código fuente de cada componente se encuentra dentro de una carpeta como detalla la Tabla 10 a continuación:

Tabla 10: Estructura de carpetas para el código fuente de componentes de LUNA

<b>Componente</b>	<b>Carpeta</b>	<b>Puerto de acceso</b>
Aplicación Móvil	luna_app	8100 en modo local
API REST	luna_api	8090
Worker Analizador	luna_analyzer_worker	-
Worker Registrador	luna_data_worker	-
RabbitMQ	luna_rabbit	-

Además de los componentes antes mencionados, se agregaron instancias de “Adminer” y “Mongo Express” (ver Tabla 11), que ayudarán a administrar las bases de datos PostgreSQL y MongoDB respectivamente.

Tabla 11: Componentes adicionales para administrar bases de datos

<b>Componente</b>	<b>Puerto de acceso</b>
Adminer	8092
Mongo express	8091

## Ejecución de LUNA en entorno local

Todos los componentes, salvo la aplicación móvil se desarrollaron para ser desplegados localmente utilizando Docker. El comando “docker-compose up -d” inicializará y desplegará los servicios de LUNA (Figura 15).

```
→ luna_platform git:(ionic_app) docker-compose up -d
[+] Running 8/8
:: Container luna_platform-rabbitmq-1      Healthy
:: Container luna_platform-adminer-1       Running
:: Container luna_platform-mongo_express-1 Running
:: Container luna_platform-mongo-1         Running
:: Container luna_platform-worker_data-1   Running
:: Container luna_platform-postgres-1      Running
:: Container luna_platform-worker_analyzer_1-1 Running
:: Container luna_platform-luna_api-1      Running
→ luna_platform git:(ionic_app) █
```

Figura 15: Inicio de servicios LUNA en entorno local

Para iniciar la aplicación móvil, se debe ejecutar el comando “ionic serve --external --no-open”, tal como se muestra en la Figura 16.

La aplicación se iniciará de forma local como un servidor web, entonces es posible acceder a ella navegando a la dirección ip del servidor local en el puerto 8100 utilizando un smartphone.

```

→ luna_app git:(ionic_app) ionic serve --external --no-open
> react-scripts start
[react-scripts] Attempting to bind to HOST environment variable: 0.0.0.0
[react-scripts] If this was unintentional, check that you haven't mistakenly set it in your shell.
[react-scripts] Learn more here: https://cra.link/advanced-config
[react-scripts] (node:456382) [DEP_WEBPACK_DEV_SERVER_ON_AFTER_SETUP_MIDDLEWARE] DeprecationWarning:
'onAfterSetupMiddleware' option is deprecated. Please use the 'setupMiddlewares' option.
[react-scripts] (Use 'node --trace-deprecation ...' to show where the warning was created)
[react-scripts] (node:456382) [DEP_WEBPACK_DEV_SERVER_ON_BEFORE_SETUP_MIDDLEWARE] DeprecationWarning
: 'onBeforeSetupMiddleware' option is deprecated. Please use the 'setupMiddlewares' option.
[react-scripts] Starting the development server...
[react-scripts] You can now view luna_app in the browser.
[react-scripts]   Local:      http://localhost:8100
[react-scripts]   On Your Network: http://192.168.0.27:8100
[react-scripts] Note that the development build is not optimized.
[react-scripts] To create a production build, use npm run build.

[INFO] Development server running!

Local: http://localhost:8100
External: http://192.168.0.27:8100, http://172.19.0.1:8100

Use Ctrl+C to quit this process

[react-scripts] assets by chunk 1.74 MiB (id hint: vendors) 45 assets
[react-scripts] + 21 assets
[react-scripts] cached modules 6.17 MiB (javascript) 31.3 KiB (runtime) [cached] 480 modules
[react-scripts] webpack 5.69.1 compiled successfully in 1230 ms
[react-scripts] Files successfully emitted, waiting for typecheck results...
[react-scripts] Issues checking in progress...
[react-scripts] No issues found.

```

Figura 16: Inicio de aplicación móvil

Una vez la aplicación se inicie con éxito, desde el teléfono celular se debe ingresar a la dirección IP del equipo que sirve la aplicación en el puerto 8100.

Dentro de la aplicación

## **4 ENTRENAMIENTO, VALIDACIÓN Y PRUEBAS DE MODELOS DE CLASIFICACIÓN DE IMÁGENES DERMATOLÓGICAS**

### **4.1 Arquitecturas de red neuronal EfficientNet**

EfficientNet (Tan y V.Le) es un modelo de código abierto que nace a partir de una propuesta de Google para el escalado compuesto (compound scaling) de modelos de redes neuronales, utilizando coeficientes compuestos sobre sus dimensiones (ancho, profundidad y resolución).

Para realizar la tarea de clasificación de imágenes de lunares se ha escogido un modelo basado EfficientNet, específicamente el modelo EfficientNet-B1. La implementación de este modelo está disponible en el paquete de python “efficientnet-pytorch” y puede ser instalado utilizando la herramienta de descarga PIP.

Tradicionalmente, la forma común de realizar un escalado de un modelo es aumentando su profundidad, ancho o también, pero de forma no tan común como las anteriores, la resolución de imágenes. Es posible escalar un modelo en dos dimensiones de forma arbitraria, pero esto requerirá un ajuste manual y tedioso que sin embargo, no asegura una eficiencia ni precisión óptimas.

La profundidad de un modelo se refiere directamente al número de capas que éste posea, su ancho corresponde al número de neuronas presentes en cada una de sus capas y su resolución representa la resolución de las imágenes que son procesadas.

En la Figura 17 se puede apreciar el escalamiento de un modelo en base a sus dimensiones tanto individuales como de forma compuesta.

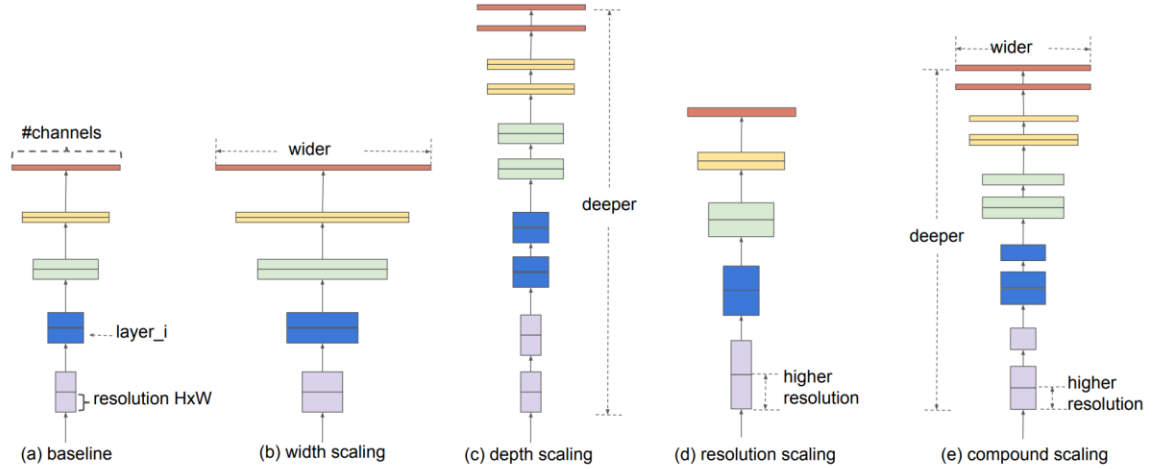


Figura 17: Escalado de modelos por dimensi3n y compuesto

Figura X: escalado de modelos por dimensi3n y compuesto

Gracias a este m3todo de escalado compuesto es posible mejorar la eficiencia y precisi3n de un modelo de manera 3ptima. El proceso consta inicialmente con una red neuronal base (EfficientNet B0) e ir ajustando los coeficientes del escalamiento compuesto. El resultado de estas modificaciones dan 3rigen a las variantes B1, B2, B3, B4, B5, B6, B7 (ver Tabla 12).

Tabla 12: Tabla comparativa de rendimiento de EfficientNet sobre dataset ImageNet

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPs	Ratio-to-EfficientNet
<b>EfficientNet-B0</b>	<b>77.1%</b>	<b>93.3%</b>	<b>5.3M</b>	<b>1x</b>	<b>0.39B</b>	<b>1x</b>
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
<b>EfficientNet-B1</b>	<b>79.1%</b>	<b>94.4%</b>	<b>7.8M</b>	<b>1x</b>	<b>0.70B</b>	<b>1x</b>
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
<b>EfficientNet-B2</b>	<b>80.1%</b>	<b>94.9%</b>	<b>9.2M</b>	<b>1x</b>	<b>1.0B</b>	<b>1x</b>
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
<b>EfficientNet-B3</b>	<b>81.6%</b>	<b>95.7%</b>	<b>12M</b>	<b>1x</b>	<b>1.8B</b>	<b>1x</b>
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
<b>EfficientNet-B4</b>	<b>82.9%</b>	<b>96.4%</b>	<b>19M</b>	<b>1x</b>	<b>4.2B</b>	<b>1x</b>
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
<b>EfficientNet-B5</b>	<b>83.6%</b>	<b>96.7%</b>	<b>30M</b>	<b>1x</b>	<b>9.9B</b>	<b>1x</b>
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
<b>EfficientNet-B6</b>	<b>84.0%</b>	<b>96.8%</b>	<b>43M</b>	<b>1x</b>	<b>19B</b>	<b>1x</b>
<b>EfficientNet-B7</b>	<b>84.3%</b>	<b>97.0%</b>	<b>66M</b>	<b>1x</b>	<b>37B</b>	<b>1x</b>
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

We omit ensemble and multi-crop models (Hu et al., 2018), or models pretrained on 3.5B Instagram images (Mahajan et al., 2018).



El estudio sobre el escalamiento compuesto, demuestra que es posible crear modelos de redes neuronales con escalamientos óptimos, que tendrán un mejor rendimiento frente a otros modelos los cuales han sido escalados arbitrariamente. Los modelos EfficientNet muestran mejor precisión (Top-1 Accuracy) a un menor número de parámetros (ver Figura 18) y a un menor costo de procesamiento (FLOPS) frente a modelos escalados arbitrariamente de forma tradicional (ver Figura 19).

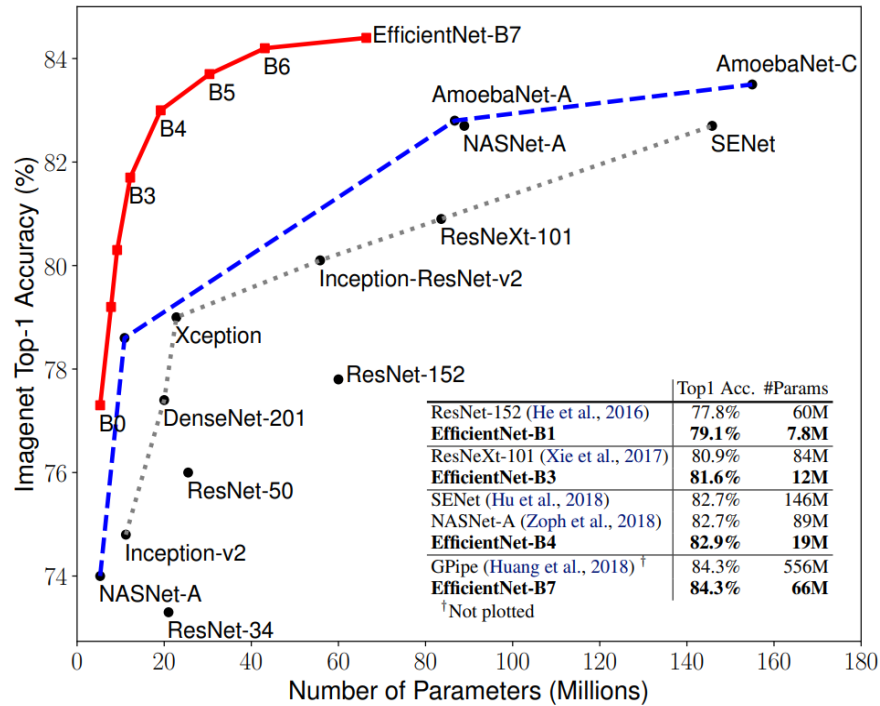


Figura 18: comparativa de modelos de redes neuronales respecto a número de parámetros vs precisión (ImageNet)

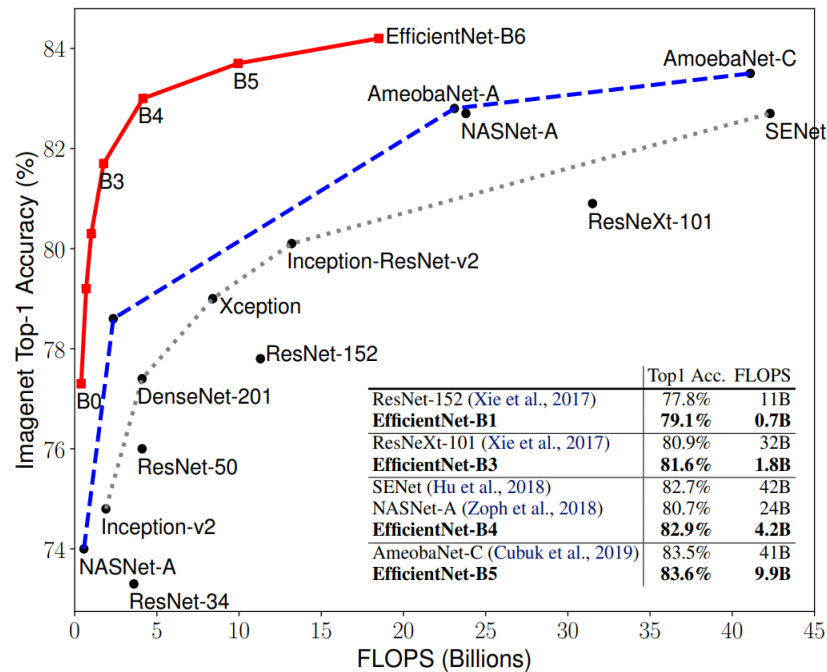


Figura 19: comparativa de modelos de redes neuronales respecto a cantidad de operaciones punto flotante por segundo (FLOPS) vs precisión (ImageNet)

## 4.2 Método de entrenamiento

Se ha implementado un notebook de jupyterlab (utilizando Python como lenguaje de programación) para realizar el entrenamiento del modelo (*ver anexo X*). El método empleado para el entrenamiento se basa en una técnica de validación llamada en inglés k-Fold Cross-Validation (Figura 20), la cual permite evaluar modelos inteligencia artificial sobre set de datos limitados.

El método k-Fold cross validation consiste en:

- Mezclar imágenes del dataset aleatoriamente.
- Dividir el dataset completo en k-grupos iguales.
- Por cada k-grupo:
  - Tomar el k-grupo y considerarlo como un dataset de evaluación.
  - Tomar el resto de grupos como dataset de entrenamiento.
  - Entrenar el modelo utilizando el dataset de entrenamiento y evaluarlo con el de evaluación.
  - Guardar los puntajes de evaluación y descartar el modelo
- Resumir el rendimiento del modelo utilizando una muestra o promedio de los puntajes de evaluación

En la implementación se itera por cada uno de los k grupos (k=10). Por cada iteración se crea un dataset de entrenamiento y otro de evaluación, para luego realizar el entrenamiento por una cantidad de épocas definidas (e=10). Por cada

época el modelo evalúa cada una de las imágenes, realizando una predicción, calculando la pérdida (loss), y verificando la precisión (accuracy), almacenando los valores (pesos) del modelo cuando este valor supera al máximo obtenido hasta el momento. Al finalizar el proceso se rescatan los pesos presentes al obtener la mejor precisión descartando el resto.

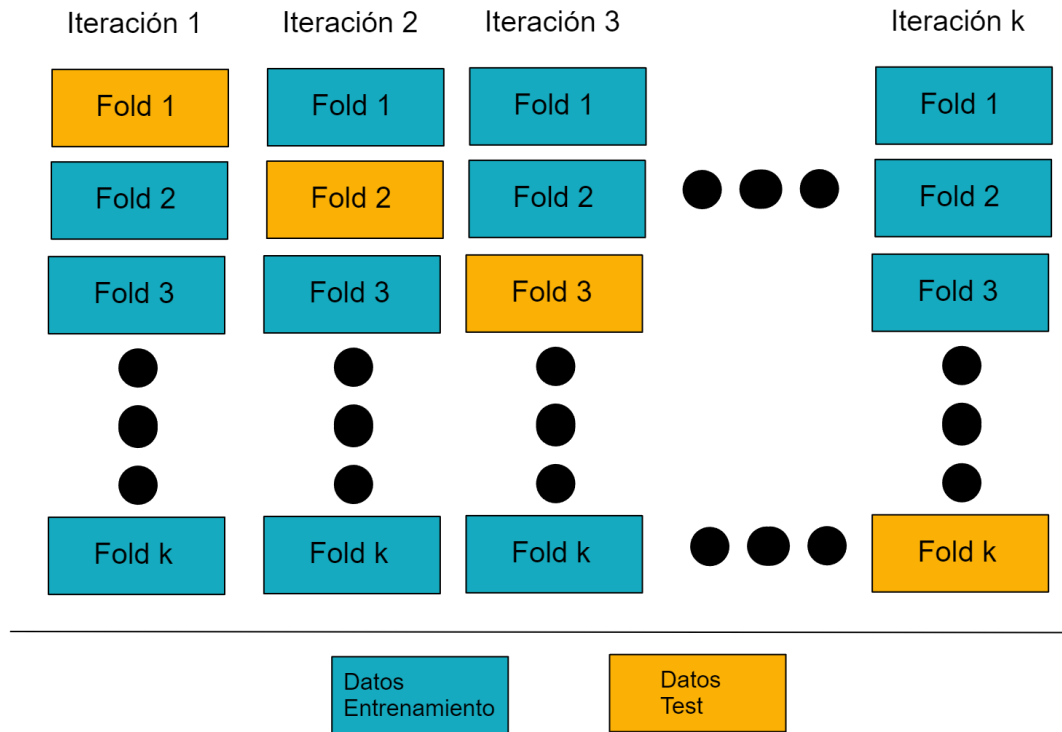


Figura 20: Diagrama de distribución de datasets utilizando k-fold cross validation

### 4.3 Resultados y conclusiones post entrenamiento

Una vez completada la fase de entrenamiento y con los mejores pesos para el modelo entrenado, es momento de probarlo contra los datos de prueba del dataset SIIM-ISIC.

Con un total de 10.932 imágenes etiquetadas, existe un 1,8% de casos malignos (200 imágenes) contra un 98, 2% de casos benignos (10.732 imágenes). Estos datos se representan en los gráficos de la Figura 21:

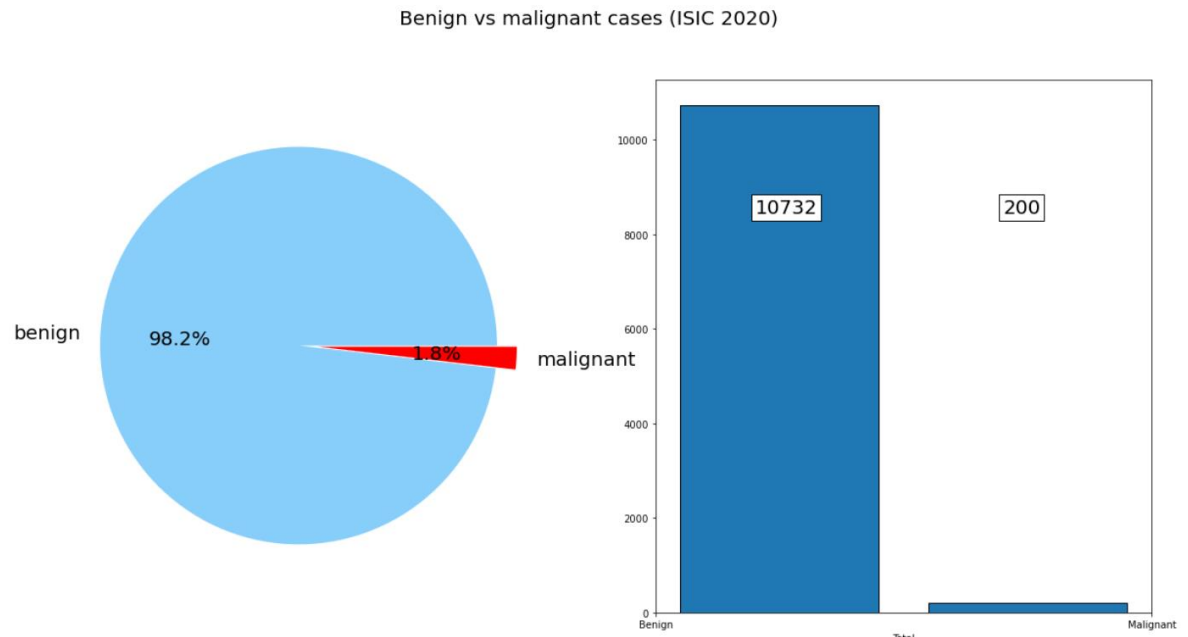


Figura 21: Cantidad y proporción de muestras benignas versus malignas presentes en el dataset SIIM-ISIC

Tras obtener las predicciones hechas por LUNANET sobre el mismo dataset (ver Figura 22) , es posible observar que ha detectado un 96,1% de casos malignos (10.502 imágenes) y un 3,9% de casos benignos (430 imágenes). Es notoria la diferencia entre la realidad (etiquetas del dataset) versus las predicciones del modelo, a simple vista pareciera que se ha cometido un error al interpretar el resultado de predicción, pero es necesario analizar las métricas de desempeño del modelo para llegar a una conclusión que cuente con un respaldo teórico.

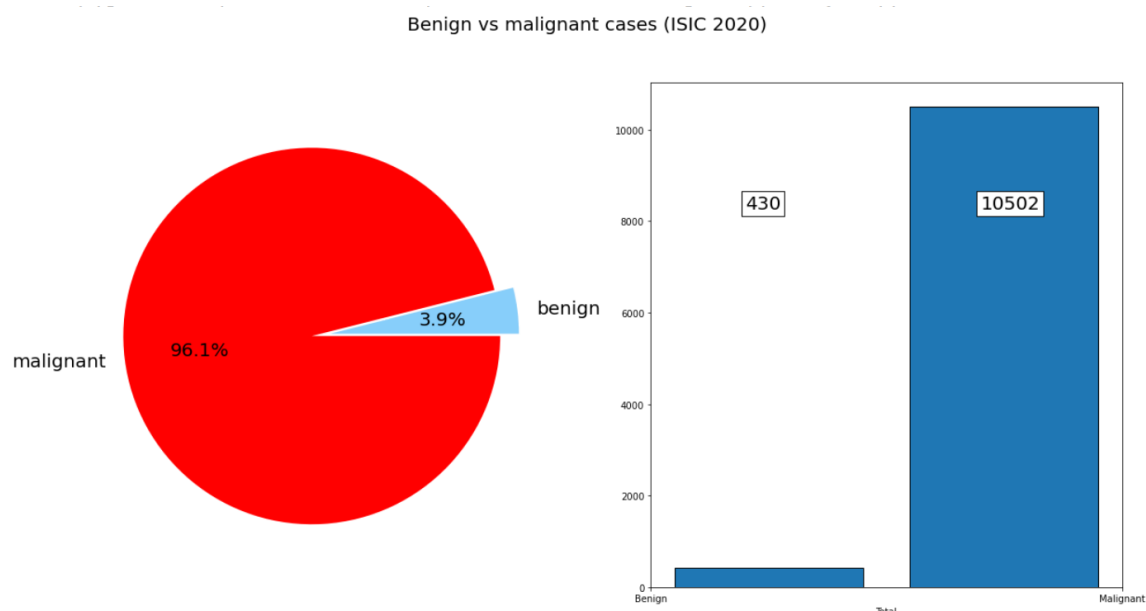


Figura 22: Resultado predicciones casos benignos y malignos de LUNANET sobre dataset SIIM-ISIC

Gracias a la librería sklearn es posible obtener las métricas de rendimiento del modelo LUNANET (ver Tabla 13) proporcionando como entrada las predicciones realizadas:

Tabla 13: Métricas de desempeño LUNANET sobre dataset SIIM-ISIC

	Precisión	Recall
<b>Benigno</b>	0,93	0,89
<b>maligno</b>	0,05	0,08

Las métricas de precisión (precision) y exhaustividad (recall) nos indican que tan bien o mal es el desempeño de nuestro modelo:

El valor de precisión nos indica las veces que la muestra y la predicción coinciden: Cuantos casos benignos eran realmente benignos y cuantos casos malignos eran realmente malignos.

Podemos ver que para los casos benignos el valor es 0,93, lo que indica que el 93% de las veces que el modelo indicó que la imagen correspondía a un caso benigno, lo era. Sin embargo, para los casos malignos el valor fue 0,05, es decir que sólo el 5% de las veces que el modelo predijo que una imagen correspondía a un caso maligno acertó. Observando el valor de precisión se puede concluir que el modelo no es capaz de detectar correctamente los casos malignos, y no se puede confiar en él en esos casos.

En cuanto al valor de recall indica la cantidad de casos que el modelo es capaz de identificar correctamente, es decir, para los casos benignos corresponde a 0,89, lo que

indica que de 100 casos benignos el modelo fue capaz de detectar 89. Para los casos malignos, el valor es 0,08, lo que quiere decir que de 100 casos malignos, sólo fue capaz de detectar 8.

Podemos concluir, con los valores de precisión y recall, que nuestro modelo es un buen detector de casos benignos, pero un mal detector de casos malignos. Si nos interesa detectar melanomas, no podremos confiar en los resultados de este modelo.

## 5 DISCUSIÓN Y CONCLUSIÓN

Existen enfermedades que pese a ser mortales, como en el caso del melanoma, pueden ser tratadas si son detectadas a tiempo. Es urgente la necesidad de optimizar tiempos y recursos frente a una evidente escasez de profesionales disponibles para que las personas que necesiten ser atendidas puedan hacerlo a tiempo.

El campo de la inteligencia artificial es bastante amplio, sin embargo, el análisis de imágenes enfocado a la salud es un foco de investigación cada vez más recurrente. Estos estudios han permitido desarrollar tecnologías que permiten analizar y detectar enfermedades de forma desatendida y automatizada. El desarrollo de este proyecto ha demostrado que es posible utilizar estas tecnologías y poder aplicarlas para resolver problemáticas reales, construyendo soluciones que aporten a la sociedad, como en el caso concreto de la salud pública de nuestro país.

El objetivo principal de este proyecto fue cumplido en su totalidad al diseñar e implementar el sistema prototipo LUNA, el cual es capaz de procesar fotografías enviadas por los usuarios utilizando modelos de redes neuronales. Se mostraron casos de proyectos y avances relacionados a la telemedicina en nuestro país, y a pesar de que el modelo entrenado no cumplió con las expectativas, se diseñó un método de entrenamiento para un modelo de red neuronal que sea capaz de clasificar imágenes médicas. Esto da por cumplido los objetivos secundarios al entregar el conocimiento adquirido a lo largo de este documento junto con el código fuente del software desarrollado.

## 6 REFERENCIAS

What Is Skin Cancer?. Accedido el 5 de mayo, 2020, desde

<https://www.skincancer.org/skin-cancer-information>

¿Qué es el cáncer de piel tipo melanoma?. Accedido el 5 de mayo, 2020, desde

<https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/acerca/que-es-melanoma.html>

Factores de riesgo para el cáncer de piel tipo melanoma. Accedido el 5 de mayo, 2020,

desde <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/causas-riesgos-prevencion/factores-de-riesgo.html>

Melanoma Survival Rates. Accedido el 15 de mayo, 2020, desde

<https://www.curemelanoma.org/about-melanoma/melanoma-staging/melanoma-survival-rates/>

Estadificación del cáncer. Accedido 10 junio, 2020, desde

<https://www.cancer.gov/espanol/cancer/diagnostico-estadificacion/estadificacion>

Chiffres clés : Dermatologue. Accedido 15 de junio, 2020, desde

<https://www.profilmedecin.fr/contenu/chiffres-cles-medecin-dermatologue>

Chile: Economía y demografía. Accedido 15 de junio, 2020 desde

<https://datosmacro.expansion.com/paises/chile>

Registro Nacional de Prestadores Individuales de Salud. Accedido 15 junio, 2020, desde

<https://rnpi.superdesalud.gob.cl>

Sabatini-Ugarte N., Molgó M. & Vial G. (2018).

Melanoma en Chile ¿Cuál es nuestra realidad?, desde

<https://doi.org/10.1016/j.rmcl.2018.06.006>

Glazer, M., Rigell, D., Analysis of Trends in Geographic Distribution of US

Dermatology Workforce Density,

<https://jamanetwork.com/journals/jamadermatology/fullarticle/2607375>



## ANEXOS

### **Anexo A: Entrevista a Claudio Valdebenito, Servicio de Salud Valdivia 2019.**

Entrevista con clientes/usuarios – Venture design

**Nombre entrevistado:** Claudio Valdebenito

**Rol entrevistado:** Jefe Subdepto. Tecnologías de Información y Comunicaciones, SSV

Preguntas:

**1. ¿Qué es lo que más/menos le gusta de su trabajo?**

Lo que más me gusta es la posibilidad de innovar, de resolver problemas reales de las personas. Una de las motivaciones por las que soy funcionario público es acortar la brecha entre la salud pública y privada.

Lo que no me gusta es que existen variables que obstaculizan en lugar de permitir avanzar en cuanto al quehacer de uno, por ejemplo cuando cambia el gobierno, el ministro de salud o cambio de jefatura generalmente vienen con enfoques, ventas, estrategias diferentes y se desecha lo realizado para comenzar de nuevo.

**2. ¿Cuáles son las tareas cotidianas que deben desarrollarse en su cargo?**

Estoy al servicio de la jefatura de mi unidad (unidades de operaciones, unidad de mantención de sistemas, unidad de mesa de ayuda, equipo de interoperabilidad, unidad de desarrollo de sistemas). Esas jefaturas están con sus planes de trabajo, sus mecánicas, su gente, sus recursos con sus proyectos, mantenciones o temas operativos. Cuando ellos tienen problemas que escapen al alcance de su poder de decisión ellos tienen que consultarme a mí, mi trabajo diario es estar ahí para resolverle los problemas a ellos, cosa de que ellos puedan seguir avanzando. Es una pirámide invertida aquí, normalmente lo que estoy haciendo es enviando correos para destrabar alguna cosa que esté trabada, que no avanza en el equipo, en otro departamento, en otro servicio de salud para promover lo que hacemos, difundir nuestros proyectos, los exitosos y los no exitosos también, llevar al equipo directivo cuales son nuestros planes de trabajo, negociarlos con ellos, normalmente todo el día estoy negociando con mis jefes de unidades y mi jefatura: plazos, calidad, costos. Entonces es un tema operativo en ese nivel.

Normalmente también intervengo en conflictos, problemas que podamos tener (internos o externos) y tratar de seguir una ruta a largo plazo. También estar atento a lo que el ministerio promueve o quiere hacer, lo que no hace y nosotros podemos hacer, analizar y medir los riesgos, gestionar los contratos con proveedores externos, recibir y analizar soluciones informáticas cuando vienen de empresas externas u otros proveedores para evaluar si realizar o no algún convenio. Un 70 u 80% de mis actividades las realizo a través de correo y reuniones de coordinación.

**3. ¿Qué desafíos presenta integrar nuevas tecnologías dentro del servicio de salud?**

Dentro de los principales desafíos es que tengan una relación costo/eficiencia que nosotros podamos abordar, tenemos que intentar que la tecnología resuelva el problema a un costo menor posible y que obedezca algún estándar de la industria, cosa de que no terminemos como una isla haciendo cosas que nadie más está haciendo y no tengamos con quien conectarnos. Trabajamos mucho con protocolos, nos preocupamos del [hl7](#), de

protocolos de comunicación XML para transporte, en interoperabilidad estamos trabajando con \_\_\_\_ y con otras arquitecturas. Si nes necesario desarrollar algo lo desarrollamos in-house, tratamos de no comprar API's o aplicativos intermedios para no depender de un proveedor en nuestra situación.

**4. ¿Cuál es el protocolo que se sigue para realizar proyectos con estudiantes tesisas?**

Se está procesando en el área jurídica una propuesta de protocolo desarrollado en conjunto con la Universidad Austral. Se espera que este año se pronuncie algo al respecto.

Por ahora el protocolo es recibir al tesista, examinar el alcance y si hay intervención a nivel de nuestras bases de datos, sistemas o información tenemos que resguardarnos con algunos contratos de confidencialidad, pero la idea es que el nuevo protocolo o convenio salga.

**5. ¿Ha participado previamente en el desarrollo de soluciones relacionados con el uso de inteligencia artificial? ¿Cuales?**

Sí, Mi primer proyecto fue para un aserradero, desarrollamos un modelo a través de una red neuronal en el año 1995-1996 aprox en el lenguaje C para poder generar cortes óptimos a un árbol a través de una fotografía. Se tomaba la fotografía, se analizaba por este sistema y determinaba cuáles eran los mejores cortes para hacer un uso óptimo del tronco. Después con inteligencia de negocios a través de modelamientos de predictivos, mezclándolos con datos de plataformas power BI, Pentaho para predecir el comportamiento de los pacientes de salud en temas de enfermedades, epidemiología y otras cosas. También desarrollé un modelo con redes bayesianas para poder modelar migración de peces para una pesquera, con lógica proposicional trabajé con un algoritmo con un árbol de decisión para poder implementar un call center, se hacía toda la entrevista a través de un modelo de lógica proposicional. Y ahora estamos con el tema de los lunares, apoyando desde afuera y estamos levantando una solución con realidad virtual con los lentes Samsung para levantar algunos aplicativos para nuestros pacientes crónicos, que tengan problemas de movilidad, tratamientos psicológicos y otros padecimientos.

**6. ¿Cómo afecta el uso de datos de pacientes en el desarrollo de nuevas soluciones? ¿Cómo se protegen los datos de los pacientes?**

Existe toda una normativa actual con respecto a datos sensibles de pacientes, acceso, almacenamiento y transporte de esos datos. Se ha definido este año un oficial de seguridad del SSV y de ciberseguridad, en un principio era yo pero el cargo ya lo traspasamos a otra persona. Esto con el motivo de implementar la normativa que está basada en la ISO 27001 de seguridad informática y protección de datos y tenemos una normativa súper clara al respecto: Eso lo aplicamos desde nuestras plataformas (datacenters, servidores, accesos a las salas de servidores, encriptación de las bases de datos, de los protocolos de comunicación, de acceso a las páginas web y trabajamos con un estándar de creación de claves (parecido a lo que usa Google), entonces las contraseñas no las tiene nadie, ni siquiera la gente de operaciones, de nada, ni de correo electrónico ni de arquitectura, es como la que tienen los bancos, fuerza a cambiar la clave cada cierto tiempo y los datos viajan encriptados, la data no se le entrega a nadie salvo que venga una solicitud formal de alguien autorizado (juzgado, jueces, o médicos

clínicos debidamente autorizados). Cuando nos piden una extracción de datos que no se enmarque en las líneas anteriores pasa por un comité de ética que valida si se puede entregar la data solicitada. Normalmente la data cuando es posible se entrega sin identificación de los pacientes, incluso al ministerio se envía con los ruts encriptados, esto evita identificar a los pacientes.

**7. ¿Cree que el desarrollo de una solución basada en Deep Learning podría disminuir la tasa de error de la derivación con especialistas?**

Creo firmemente que sí, pero depende mucho del modelo que se cree, de la casuística con la que se pruebe, depende de que además tenga una variable de algún periodo de pruebas con clínicos, validación, que se pueda contrastar con data de diagnósticos que ya se tengan versus data que ya esté entregando el sistema y se pueda llegar a un porcentaje de aproximación de un margen aceptable. En Chile esto no está normado, lo único que existe en Chile es que un sistema informático no pueda realizar diagnósticos, pre-diagnósticos ni cosas de ese estilo, por lo tanto lo veo más que nada como un apoyo al personal clínico que está tratando al paciente y le permita realizar un diagnóstico más certero. Sobre todo alertar por falsos positivos o falsos negativos y que al personal clínico le haga replantear el diagnóstico y hacer una reevaluación: pedir más exámenes, más estudios, enfrentar el problema de otra manera por que existe un sistema que le está generando una alerta y en ese sentido sería una ayuda muy buena

**8. ¿Con qué soluciones/servicios ya existentes en el SSV podría integrarse la solución propuesta por el proyecto?**

Nosotros tenemos una plataforma de telemedicina que tiene 5 o 6 años en explotación y en esa plataforma existe un módulo llamado “tele-dermato” en donde los médicos especialistas y médicos generales hacen derivaciones y contra derivación, referencia contra referencia de los pacientes, es decir, un médico general vé un paciente y si tiene la sensación de que no puede resolverlo ya sea por que es un tema dermatológico complicado, a través de esta plataforma el ingresa una fotografía, información del paciente y hay un médico especialista que en el hospital va revisando esa lista de trabajo y va realizando sus observaciones, pidiendo más exámenes y eventualmente la misma plataforma les permite hacer una especie de chat en el que queda todo registrado y es parte de una ficha clínica electrónica. Todo los eventos se registran con tiempos, logs, de quién hizo algo qué día, qué hora, y al final puede existir una resolución del caso o una derivación del paciente al especialista presencial al hospital base por ejemplo. A esta plataforma se pueden adjuntar imágenes, archivos, etcétera. Sería ideal que esta plataforma pudiese contar con una especie de asistente para el médico general para detectar un posible diagnóstico que no esté bien.

**9. ¿Cómo cree que podrían reaccionar los especialistas (dermatólogos) y médicos generales frente a este proyecto?**

Hubo una experiencia el año pasado con oftalmología, en donde una empresa desarrolló un algoritmo donde tras tomarse una fotografía de un ojo, el algoritmo es capaz de detectar glaucoma o algunas otras patologías en los ojos según la empresa con un alto porcentaje de efectividad. En la práctica nunca se usó porque ellos lo promocionaron como que iba a reemplazar al prediagnóstico, procesando alta cantidad de pacientes y derivando sólo a los que se les detectara una patología. Entonces podía pasar que el sistema se equivoque con los que haya rechazado y esos pacientes podrían demandar al sistema al ser mal diagnosticados y por ese motivo se finiquitó el proyecto. Yo creo que

fué mal planteado, cabe destacar que el porcentaje de efectividad asegurado no era tan alto, ya que tampoco fué probado con población chilena (venía con valores basados en datos extranjeros), los algoritmos fueron desarrollados afuera y no estaba bien probado, no tenía mecanismos de entrenamiento, suponiendo que estaba basado en un algoritmo de red neuronal y no hubo un proceso de validación o auditoría a través del ISP (Instituto de Salud Pública) o a través del mismo Ministerio de Salud. También más que un diagnóstico o pre-diagnóstico debió presentarse como un apoyo al personal clínico. Eso atemorizó y provocó inquietud en algunos funcionarios clínicos ya que daba la impresión de que podría reemplazarlos.

Lo bueno de nuestra región es que en varias especialidades los funcionarios clínicos ya están utilizando nuestra plataforma desde hace muchos años, están muy acostumbrados a que la tecnología los asista, a registrar información, para ellos es una herramienta en lugar de una competencia y están felices ya que pueden obtener información para sus casos y guarda el historial de la historia clínica lo que implica que ya no utilizan fichas de papel, pueden pedir exámenes de laboratorio, de rayos X, y lo ven en el sistema, entonces ellos ya tienen mucha afinidad con la tecnología. El sistema sólo lo estamos usando en nuestra región, pero estamos realizando un convenio con Araucanía Sur, ya que ellos quieren integrarse con nosotros y justamente estas semanas también con Punta Arenas y Magallanes porque también quieren utilizar nuestra plataforma. Entonces, dependiendo de cómo se les presente esta solución al personal clínico y su alcance yo creo que el terreno está fértil para lograr esta solución.

**10. ¿Qué se espera de este proyecto? ¿Cuáles son las expectativas?**

Lo que espero es que podamos integrarlo en nuestra plataforma en Valdivia, en donde yo pueda llamar a algo con ciertos parámetros de entrada y este me entregue resultados que a mí me sirvan para poder presentarlos en la plataforma y guardarlos en la ficha clínica electrónica. Con eso nosotros estamos súper bien.

Si va a quedar alojado en nuestros data centers no es problema, si va a quedar afuera, en un tercero, me gustaría que fuera un tercero a través de un convenio que no nos cobre. Es necesario que un servicio que está en producción no dependa de alguien externo, o que el proveedor del servicio comience a cobrar. Lo importante es que el producto ofrecido quede muy bien documentado, para que el SSV pueda darle continuidad operacional y no tenga que llamar al creador. En resumen: no sólo se pide un buen producto, sino una buena documentación que nos permita darle continuidad operacional.

## ANEXO B: Estadísticas sobre el cáncer (Global Cancer Observatory)

Tabla 14: Número de casos de cáncer (no melanoma) por cada 100.000 habitantes durante el año 2020

	Cáncer de mama	Cáncer de próstata	Cáncer a la tiroides	Cáncer testicular	Cáncer al hígado	Cáncer a los ovarios
<b>Mundo</b>	2.261.419	1.414.259	586.202	74.458	905.677	313.959
<b>Chile</b>	5.331	8.157	1.164	824	1.560	837
<b>Argentina</b>	22.024	11.686	4.106	2.047	2.437	2.199
<b>EE.UU</b>	253.465	209.512	52.912	9.407	42.284	23.820

Tabla 15: Tasa de mortalidad de otros tipos de cáncer (no melanoma) por cada 100.000 habitantes durante el año 2020

	Cáncer de mama	Cáncer de próstata	Cáncer a la tiroides	Cáncer testicular	Cáncer al hígado	Cáncer a los ovarios
<b>Mundo</b>	58,5	36,0	7,5	1,9	11,6	8,1
<b>Chile</b>	55,0	86,5	6,1	8,7	8,2	8,6
<b>Argentina</b>	95,1	53,0	9,1	9,3	5,4	9,5
<b>EE.UU</b>	151,6	127,9	16,0	5,7	6,9	14,2

Tabla 16: Total de casos y tasa de mortalidad de melanoma por cada 100.000 habitantes durante el año 2020

	Total casos	Tasa de mortalidad
<b>Mundo</b>	324.635	4,2
<b>Chile</b>	857	4,5
<b>Argentina</b>	1.731	3,8
<b>EE.UU</b>	201.617	28,8

