



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN**

**Clasificación de bacterias en imágenes de microscopía  
mediante aprendizaje profundo.**

**T E S I S**

QUE PARA OPTAR POR EL GRADO DE  
**Maestro en Ciencia e Ingeniería de la Computación**

Presenta

Ing. Alberto Daniel Fuentes Villegas

Asesor

Dr. Boris Escalante Ramírez

Facultad de Ingeniería

Ciudad Universitaria, Ciudad de México. Noviembre, 2024



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



**PROTESTA UNIVERSITARIA DE INTEGRIDAD Y  
HONESTIDAD ACADÉMICA Y PROFESIONAL**

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado "Clasificación de bacterias en imágenes de microscopía mediante aprendizaje profundo", que presenté para obtener el grado de Maestro en Ciencias e Ingeniería de la Computación, es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Programa de Posgrado, citando las fuentes, ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad e los actos de carácter académico administrativo del proceso de titulación/graduación

**Atentamente**

*A. Fuentes*  
Alberto Daniel Fuentes Villegas

Número de cuenta: 523013882

## Abstract

Most bacterial classifiers created by neural networks, machine learning, and/or image processing methods do not generalize to databases other than those they were trained on, even if the sample preparation is similar. This is inherent to differences in the images, such as the acquisition being carried out with different microscopes, with different cameras, the variation in lighting, etc.

In this work, we present an ensemble of deep neural networks designed for the classification of bacteria in a more general context. The data set we used consists of bacteria of different genera such as *Actinomyces*, *Escherichia*, *Staphylococcus*, *Lactobacillus* and *Micrococcus*; all samples prepared by Gram staining, acquired by bright field microscopy and coming from various sources. To normalize the diversity of image features, we apply domain adaptation and generalization techniques as well as data augmentation techniques. Subsequently, we used phenotypic characteristics such as color reaction to Gram stain, morphology and cell agroupment to classify bacteria.

Among the results, we obtain high classification rates through the proposed ensemble of four models, this in comparison to traditional training where we use all available images to train a single network.

## Resumen

La mayoría de los clasificadores de bacterias creados por redes neuronales, aprendizaje automático y/o métodos de procesamiento de imágenes no logran generalizar cuando se prueban en bases de datos diferentes a aquellas con las que fueron entrenadas, incluso si la preparación de la muestra en el laboratorio es similar. Esto es inherente a diferencias en las imágenes, como que la adquisición se realizó con distintos microscopios, distintas cámaras, a la variación en la iluminación, etc.

En este trabajo, presentamos un ensamble de redes neuronales profundas diseñadas para la clasificación de bacterias en un contexto más general. El conjunto de datos que utilizamos se compone de fotos de bacterias de distintos géneros como *Actinomyces*, *Escherichia*, *Staphylococcus*, *Lactobacillus* y *Micrococcus*; todas las muestras preparadas por medio de tinción de Gram, adquiridas mediante microscopía de campo claro y provenientes de diversas fuentes. Para normalizar la diversidad de características de la imagen, aplicamos técnicas de adaptación y generalización de dominio, así como técnicas de aumento de datos. Posteriormente, utilizamos características fenotípicas, como la reacción de color a la tinción de Gram y la morfología para clasificar las bacterias.

Entre los resultados obtenemos altas tasas de clasificación por medio del ensamble propuesto, esto en comparación a un entrenamiento "tradicional" donde utilizamos las imágenes disponibles de distintas fuentes para entrenar un sólo modelo, en lugar de varios (el ensamble).

*“Las campanas del monasterio de Gion reverberan advirtiendo la impermanencia de las cosas*

*En sus salones, los frutos marchitos nos recuerdan que todo lo que florece debe perecer*

*Los orgullosos se desvanecen como un sueño al despertar*

*E incluso los poderosos terminan como polvo frente al viento*

*Pero claro...*

*Nosotros somos la excepción”*

*-Heike Monogatari/Gege Akutami*

*Para Esther y Edmundo...*

# Agradecimientos

Los agradecimientos más importantes son hacia a mi familia: Mónica, Ángel, Ailed y Linda, por su apoyo, comprensión y hasta ladridos en todos estos años. A mis amigos, por los consejos, las recomendaciones, las oportunidades, las ayudas y los momentos que hemos pasado y de los que he aprendido mucho. A Leo por la sugerencia de asistir a aquella escuela de invierno. Y a Lisa, por la inspiración para ser mejor.

También le agradezco al Conahcyt por la beca recibida en estos dos años de maestría con número de CVU: 12256728. Y a los proyectos PAPIIT UNAM IN108624 y IT101624



# Índice General

<b>1. Introducción</b>	<b>1</b>
<b>1.1. Antecedentes</b>	<b>2</b>
<b>1.2. Planteamiento del problema</b>	<b>6</b>
<b>1.3. Justificación</b>	<b>7</b>
<b>1.4. Hipótesis</b>	<b>9</b>
<b>1.5. Objetivos</b>	<b>9</b>
a. General	9
b. Específicos	9
<b>2. Marco Teórico</b>	<b>10</b>
<b>2.1. Identificación tradicional y morfología de la bacteria</b>	<b>10</b>
2.1.1. Cocos, bacilos y otras morfologías	11
2.1.2. Tinción de Gram	13
2.1.3. Sobre la especie y género: taxonomía	13
<b>2.2. Procesamiento digital de imágenes</b>	<b>15</b>
2.2.1. Histogramas de color RGB	15
2.2.2. Especificación de histograma	15
<b>2.3. Algoritmo para clasificación</b>	<b>16</b>
2.3.1. Redes Neuronales Convolucionales	17
2.3.2. Arquitectura ResNet	17
2.3.3. Ensamblés	19
2.3.4. Módulo de normalización SNR	19
<b>2.4. Pruebas y validación de clasificación</b>	<b>20</b>
2.4.1. Matrices de confusión	20
2.4.2. Validación cruzada	22
<b>3. Metodología</b>	<b>23</b>
<b>3.1. Bases de datos</b>	<b>24</b>
<b>3.2. Procesamiento de imágenes</b>	<b>24</b>

<b>3.3. Algoritmo de identificación de patrones</b>	<b>27</b>
<b>3.4. Aprendizaje y pruebas</b>	<b>28</b>
<b>4. Resultados</b>	<b>32</b>
<b>4.1. Baseline</b>	<b>32</b>
<b>4.2. Ensamble de redes</b>	<b>34</b>
4.2.1. Por tinción	34
4.2.2. Por morfología	36
4.2.3. Por género	37
4.2.4. Por especie	39
4.2.5. Unión del ensamble	41
<b>5. Conclusiones</b>	<b>44</b>
<b>5.1. Baseline</b>	<b>44</b>
<b>5.2. Ensamble</b>	<b>44</b>
<b>5.3. Globales</b>	<b>45</b>
<b>5.4. Perspectivas</b>	<b>46</b>
<b>6. Bibliografía</b>	<b>48</b>



# 1. Introducción

Las bacterias se encuentran entre los organismos más sencillos y abundantes que hay en el planeta. Sus tamaños varían en el rango de los micrómetros, por lo que se utilizan microscopios para estudiarlas. Su identificación es primordial en áreas como patología y microbiología, ya que buscan tener un mejor entendimiento de las enfermedades que causan y sus posibles tratamientos [1].

Existen diversas técnicas para identificar y clasificar las bacterias, entre los ejemplos de estos métodos destacan: la caracterización fenotípica, la identificación genotípica o molecular e incluso los métodos proteómicos. Una de las técnicas más utilizadas, y la que proponemos utilizar en este trabajo, es la identificación por fenotipo (basada en las características observables) debido a su accesibilidad y bajo costo, estas características a diferencia de otros métodos que requieren tener acceso a tecnologías más avanzadas, como aquellas que permitan la secuenciación de ADN/ARN o la espectrometría de masas [2] (aplicados en genotípica y proteómica, respectivamente).

Al ser la identificación fenotípica la técnica más accesible, esta es, naturalmente, dependiente de ciertos tipos de microscopía ya que, como se ha mencionado, no siempre se tiene acceso a equipos muy sofisticados para realizar la identificación. Por esta razón, en el presente trabajo se propone basarse únicamente en imágenes obtenidas mediante microscopía óptica con campo claro, dado que este tipo de microscopía es la más conocida por su simplicidad y accesibilidad, así como su amplia aplicación en la técnica de identificación fenotípica [1].

Esta investigación incluye un conjunto de datos sólido que abarca una variedad de bacterias, entre ellas *Actinomyces*, *Escherichia*, *Staphylococcus*, *Lactobacillus*, *Micrococcus*, *Neisseria*, *Veillonella*, *Bacteroides* y *Clostridium*. Cada tipo de bacteria se caracteriza por distintas propiedades de tinción de Gram<sup>1</sup> y morfológicas. Las imágenes, adquiridas a través de microscopía de campo claro y provenientes de múltiples fuentes, plantearon desafíos únicos debido a la inherente diversidad en las características de las imágenes. Para tratar esto, se implementaron técnicas de adaptación de dominio, así como técnicas de aumento de datos, que nos permitieron abordar estas variaciones de manera efectiva.

Consideramos que las características fenotípicas, como las reacciones de color a la tinción de Gram y los atributos morfológicos, son fundamentales para el proceso de clasificación. Al integrar estas características en una cascada de redes neuronales profundas surgió un marco capaz de discernir diferencias sutiles entre especies bacterianas. Nuestro enfoque propone una alternativa en torno a las limitaciones de los clasificadores tradicionales aprovechando el poder del aprendizaje profundo para tareas específicas.

---

<sup>1</sup> La tinción de Gram es un método que consiste en literalmente “pintar” (aplicando tintes) a las bacterias para mejorar su observación bajo el microscopio. Este concepto se explica un poco más en la sección de Marco Teórico (2.1.2) del presente trabajo.

## 1.1. Antecedentes

La literatura muestra que existen trabajos previos que utilizan técnicas computacionales para la identificación de bacterias basadas en su fenotipo. Por ejemplo, el estudio de N. Mohamad, *et. al.* [3], donde mencionan distintos métodos comúnmente utilizados para la identificación de bacterias a partir de su morfología. Sin embargo, en lugar de desarrollar un método específico, presentan un abanico de opciones para la identificación mediante aprendizaje máquina. Los autores describen algunos métodos de:

- Segmentación de imágenes: Por medio de métodos de clusterización, de umbralización y, en particular, los métodos de detección de bordes.
- Extracción de características: De acuerdo al estudio, las características geométricas más relevantes para la clasificación de bacterias son: perímetro, área, radio, circularidad, compacidad, excentricidad y tortuosidad.
- Algoritmos de clasificación: Aquí se menciona que se utilizan clasificadores bayesianos, árboles de decisión y lo más extenso es sobre algoritmos  $K$ -NN<sup>2</sup> y redes neuronales multicapa.

En el trabajo de E. P. Sánchez Femat [4], identifican cuatro bacterias diferentes: *Escherichia coli*, *Bacillus cereus*, *Salmonella typhimurium* y *Staphylococcus aureus*, basándose en sus características fenotípicas macroscópicas en cultivos. Para esta identificación, se aplicaron algoritmos de aprendizaje automático, específicamente técnicas cruzadas de algoritmos de clasificación de *Naive Bayes* y  $K$ -NN. Además realizaron un preprocesamiento de las imágenes, dividiendo el proceso en cuatro etapas: recorte, filtrado, mejoramiento del histograma (expansión) y segmentación. La cantidad de imágenes utilizadas fue de 100, y estas se dividieron en conjuntos de entrenamiento y prueba.

Los resultados obtenidos con el clasificador *Naive Bayes*, particularmente para la especie *Staphylococcus aureus*, indican que alcanza una clasificación de instancias correctas (exactitud) del 96.66%, y un 92% para las cuatro clases en general. Mientras que en el caso del clasificador  $K$ -NN obtuvo un porcentaje de acierto del 100% para el *Staphylococcus aureus* (la mejor especie clasificada), con un 98% en general para las cuatro bacterias.

Entre las limitaciones descritas en las conclusiones se encuentra la necesidad de igualar las condiciones de luz al adquirir las imágenes y que la segmentación por colores se podría mejorar dependiendo de la cepa del medio de cultivo. Los autores concluyen que el método de aprendizaje puede mejorarse, ya que las principales características que se utilizan son: medio de cultivo y color de la colonia, donde el usuario que quiere identificar la bacteria debe especificar el medio de cultivo al modelo. Este trabajo se enfoca exclusivamente en aspectos macroscópicos y no aborda las especificaciones morfológicas de cada bacteria.

---

<sup>2</sup> K-NN por sus siglas en inglés *K-Nearest Neighbours*, o en español K-ésimo vecino cercano.

Por otro lado, existen trabajos que ofrecen una perspectiva interesante orientada a la morfología de células individuales de bacterias. Como el trabajo de Ruusuvoori, *et. al.* [5], dónde clasifican bacterias utilizando algoritmos como *K-NN*. En el estudio se identificaron las siguientes especies de bacterias: *Clostridium butyricum*, *Escherichia coli* y *Staphylococcus spp.*<sup>3</sup>, clasificadas en categorías basadas en formas paramétricas, tales como alargada, esférica, espiral, elipse y semicircular.

El tratamiento que se le dio a las imágenes para separar a los objetos de interés, es decir las células del fondo, fue segmentarlas, seguido de transformar la imagen a escala de grises y suavizarla mediante un filtro Gaussiano. Posteriormente, la imagen se binarizó, se aplicó un nuevo filtro, y del resultado de este filtro se extrajeron las formas de los objetos. Los resultados indican que en las 5 formas que clasifica, lo hace con una precisión del 82.9%, aunque no especifica el porcentaje por clase, indica que los mejores resultados se dan en los *Staphylococcus*. Los autores concluyen que el traslape de los objetos es una amenaza severa para el reconocimiento de bacterias a través de sus morfologías.

Otro ejemplo se encuentra en el trabajo de P. Hiremath, *et. al.* [6], donde se propone un algoritmo de reconocimiento de bacterias utilizando aprendizaje automático, con el objetivo de reconocer entre tres formas geométricas presentes en la mayoría de las especies: circulares en cocos, ovaladas en bacilos, y espirales en bacterias helicoidales. Para ello, se emplean técnicas de segmentación, para separar las formas de interés del fondo y eliminando las muestras que presentan traslapes. El vector de características se conforma de características geométricas específicas del objeto de interés, como circularidad, área, compacidad, eje mayor y menor, todos estos descriptores son calculados después de la segmentación. Luego, estos resultados se ingresan a un clasificador no especificado. Los resultados finales indican que la clasificación es correcta en un 94% de las instancias.

Además de estas técnicas de aprendizaje máquina para clasificación, existen trabajos utilizando redes neuronales convolucionales (*CNN*<sup>4</sup>). Los autores T. Treebupachatsakul y S. Poomrittigul [7], [8] presentan dos trabajos dónde hacen la clasificación de 3 especies de bacterias: el *Staphylococcus aureus*, *Micrococcus spp.*, y el *Lactobacillus delbrueckii*. En dichos trabajos se utilizaron dos conjuntos de datos distintos: el primero de un cultivo de laboratorio realizado por los mismos autores, y el segundo obtenido de una base de datos de acceso público. La finalidad del estudio fue utilizar una red neuronal con arquitectura LeNet para el reconocimiento de imágenes de bacterias con dos distintas resoluciones: estándar (1.3 MP) y alta (3.3 MP), logrando una tasa de exactitud general en validación del 89% considerando ambas resoluciones. Se entrenaron modelos por separado para cada resolución.

Por otra parte, el trabajo de K. Smith, *et. al.* [9], se centra en el análisis de muestras de sangre preparadas por el método de tinción de Gram para determinar la existencia de infecciones por cocos. Este análisis se realiza por medio de microscopía óptica y aprendizaje profundo, la clasificación utilizada fue en tres grupos: 1. Cocos gram positivos en cadenas y pares, 2. Cocos gram positivos en racimos, y 3. Bacilos gram

---

<sup>3</sup> Spp. en inglés *Species* o *Species plural*. Significa que se incluye más de una especie no especificada.

<sup>4</sup> CNN por sus siglas en inglés *Convolutional Neural Network*.

negativos. Utilizan una arquitectura *CNN* llamada *Inceptionv3* para la clasificación de cada imagen, utilizaron 500 instancias para el entrenamiento. El artículo hace énfasis en los problemas relacionados con la red neuronal y su implementación. Los resultados de la matriz de confusión con 189 instancias de prueba fueron:

- Tasa de verdaderos positivos por grupo:
  - “Bacilos” del 98.1%
  - “Cocos en cadenas y pares” del 75%
  - “Cocos en racimos” del 97.2%.
  
- Especificidad por grupo:
  - “Bacilos” de 96.3%
  - Cocos en cadenas y pares” del 98.4%
  - Cocos en racimos” del 93.2%.

En general, lograron una exactitud del 92.5% para todas las clases del modelo. Entre las limitaciones y aspectos mejorables se menciona la cantidad de morfologías bacterianas incluidas, la preparación del método de tinción, la variabilidad en la localización de las células dentro del área de la imagen capturada, la intensidad y las diferencias entre el grosor de las paredes celulares al realizar la tinción. Los autores proponen estandarizar estas variables para poder automatizar la detección y la clasificación de los organismos presentes bajo el microscopio.

Por otro lado, B. Zieliński, *et. al.* [10], aplican una combinación entre métodos de aprendizaje máquina más tradicional y aprendizaje profundo para la clasificación de distintas especies de bacterias. Para realizar este estudio crearon una base de datos llamada DiBaS, que contiene un total de 660 imágenes de 33 especies de bacterias con aproximadamente 20 imágenes por cada especie.

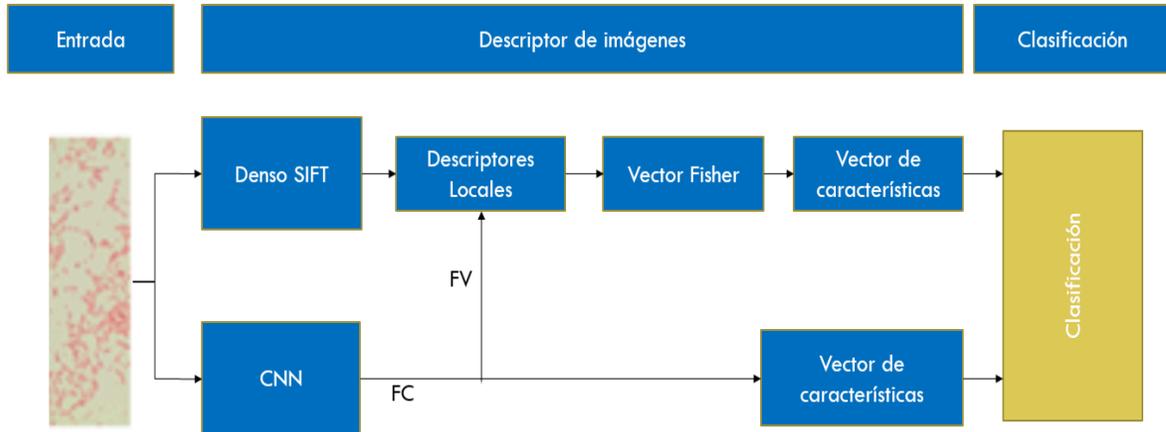
La metodología propuesta para la clasificación en este trabajo es: extraer descriptores de la imagen (mediante *CNN* o algoritmos tipo *SIFT*<sup>5</sup>, ver Figura 1), producir un vector de características a partir de los descriptores (pueden aplicar reducción de dimensionalidad mediante una técnica conocida como *Vector Fisher*), y finalmente identificar las bacterias con un modelo de clasificación como *SVM*<sup>6</sup> o Bosques Aleatorios. En general, combinan distintos métodos de extracción y clasificación de las imágenes.

Los resultados indican que la combinación óptima es usar un vector de características extraído con *CNN* (por medio de una arquitectura *VGG-M*) en combinación con un clasificador *SVM* lineal, logrando así, una exactitud del 97.24% para todas las clases.

---

<sup>5</sup> Scale Invariant Feature Transform, es un algoritmo para extraer características invariantes a la escala.

<sup>6</sup> Support Vector Machine o en español máquina de soporte vectorial. Algoritmo para clasificar.



**Figura 1.** Mapa de bloques con las metodologías propuestas en [10] para la clasificación de las imágenes. FV significa Fisher Vector o Vector de Fisher, que se refiere a un método para reducir la dimensionalidad de los datos, mientras que FC viene de *Fully Connected Layer* o Capa Totalmente Conectada proveniente de la CNN.

En el trabajo presentado por Wahid, *et. al.* [11], reportan un enfoque para automatizar el proceso de reconocimiento y clasificación de bacterias con el uso de CNN. Utilizan el método de aprendizaje por transferencia para reentrenar un modelo *ResNet 50* con un conjunto de datos de aproximadamente 500 imágenes de microscopía de cinco bacterias diferentes: *Clostridium botulinum*, *Vibrio cholerae*, *Neisseria gonorrhoeae*, *Borrelia burgdorferi*, *Mycobacterium tuberculosis*. El 20% de las imágenes, elegidas y separadas aleatoriamente, se utilizaron para probar la exactitud de clasificación de la red. El modelo reentrenado fue capaz de reconocer y clasificar las cinco especies diferentes de bacterias, reportando que los resultados experimentales de predicción lograron una precisión promedio del 93% y una exactitud del 81.8%.

Por último, en el trabajo de Spahn, *et. al.* [12], presentan *ZeroCostDL4Mic*, que es un compendio de *notebooks*<sup>7</sup> creados en *Google Colab* [13], el cual incluye modelos preentrenados de redes convolucionales conocidas, entre las cuales se encuentran: *U-Net*, *CARE*, *StarDist*, *SplineDist*, *pix2pix*, *Noise2Void*, *YOLOv2* y *fnet*. Su objetivo es proporcionar a los microbiólogos herramientas para distintas tareas como segmentación, detección, eliminación de ruido y superresolución. Además proporcionan conjuntos de datos adquiridos con diferentes tipos de microscopía (campo claro, fluorescencia) e imágenes en diferentes etapas de crecimiento para tres especies de bacterias (*Staphylococcus aureus*, *Escherichia coli* y *Bacillus subtilis*). Reportan varios resultados, un ejemplo de ellos es que las redes entrenadas con tipos de imágenes combinadas (de distintos tipos de microscopía) obtienen una precisión del 88% en segmentación usando la red *StarDist*. Sin embargo, no se menciona ninguna red entrenada o preparada para hacer clasificación de bacterias.

<sup>7</sup> Un notebook de *Python* es un cuaderno interactivo que permite redactar texto, escribir/compilar código, y crear visualizaciones de resultados. Son comúnmente usados para compartir resultados en el análisis de datos, el aprendizaje automático y la programación experimental. Un ejemplo de estos *notebooks* es *Google Colab*, página web en la que no es necesario instalar un ambiente local para comenzar a programar.

## 1.2. Planteamiento del problema

Las razones más importantes para estudiar a los microorganismos es la de conocer las enfermedades que provocan e investigar cómo identificarlas y la manera de controlarlas. Es frecuente que los microorganismos originen la aparición de numerosas manifestaciones de alguna enfermedad, y síntomas similares pueden ser causados por distintos microorganismos que requieren tratamientos diferentes. Es por ello la importancia de tener una identificación certera, rápida y lo más veraz posible del origen de estos. Por ejemplo, el *Staphylococcus aureus* es un agente causante de endocarditis, neumonía, infecciones en heridas e intoxicaciones alimentarias [14]; afecciones que también podrían ser confundidas con la presencia de otros microorganismos, como *Haemophilus influenzae* (neumonía), *Streptococcus* (endocarditis) o especies de *Cándida* (hongos que causan infecciones en heridas).

Identificar un microorganismo es una tarea exigente que requiere de inversión en tiempo de expertos, estos pueden pasar varios minutos analizando una muestra, además aún hay que incluir el tiempo de preparación y de cultivo. Así como inversión económica, debido a que métodos como los genotípicos o proteómicos necesitan herramientas especializadas que son costosas [15].

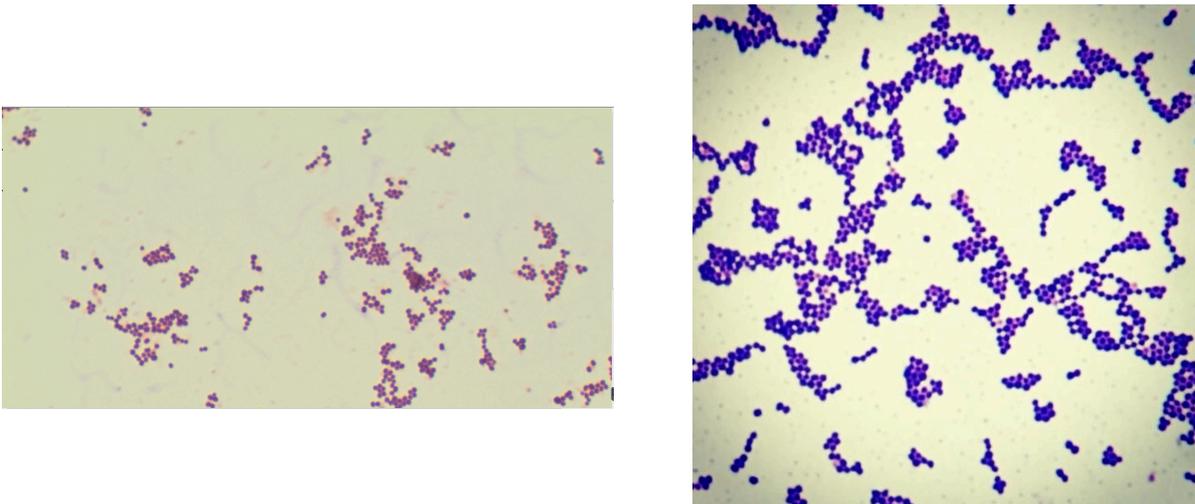
Además, desafortunadamente es una práctica común en el mundo recetar y utilizar antibióticos de alto espectro ante la mínima sospecha de alguna infección bacteriana, debido al desconocimiento del agente causante y los inconvenientes de realizar una identificación minuciosa en casos no graves. Sin embargo, esta práctica es incorrecta ya que puede causar resistencia a los antibióticos en las bacterias y esto podría desencadenar un problema de salud mundial si continúa el ritmo de uso actual [16].

Por lo tanto, la aplicación de técnicas de reconocimiento de patrones en imágenes de microscopía podría automatizar el proceso de identificación, haciéndolo más eficiente y preciso, con menor gasto de recursos tanto temporales como monetarios, además de que podría proveer una potencial solución al uso indiscriminado de antibióticos.

### 1.3. Justificación

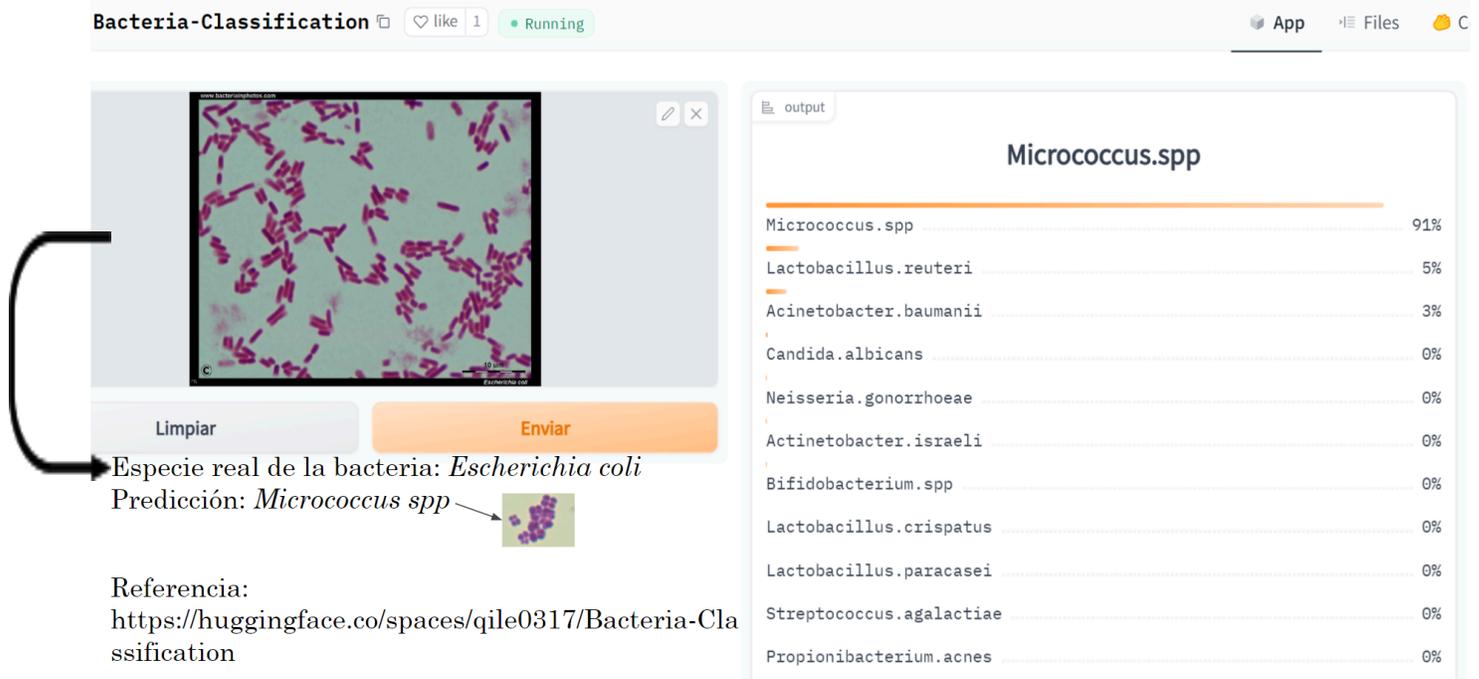
Como ya se ha expuesto en el planteamiento del problema, es importante identificar a los microorganismos para tratar la sintomatología que causan los microorganismos, y de esta manera brindar una atención inmediata y de calidad al paciente, sin poner en riesgo la seguridad de los demás.

En los antecedentes se presentaron ejemplos sobre la aplicación de técnicas de aprendizaje de máquinas y profundo para la clasificación de bacterias. Sin embargo, cada trabajo utiliza imágenes adquiridas en las mismas condiciones, lo que ocasiona que cada clasificador falle al utilizar bases de datos diferentes a la que usaron en el entrenamiento. Esto es debido a que dos imágenes de bacterias de la misma especie pueden variar, incluso si se ha seguido el mismo protocolo de tinción y preparación, si es que estas fueron adquiridas con distintos microscopios en distintos lugares, debido a factores como la iluminación, la cámara, etc. (ver Figura 2).



**Figura 2.** Ejemplos de diferencias entre dos imágenes de la misma especie de bacteria (*Staphylococcus aureus*), tomadas en diferentes condiciones. Imágenes tomadas de [10, 49].

Para ilustrar esta situación, tomemos una implementación pública basada en el banco de datos DiBaS presentado en [10]. Este modelo, implementado en una página de *Hugging Face* [28], consiste en cargar una imagen (o utilizar las que tienen disponibles de muestra), presionar enviar y esperar a que se realice una predicción sobre la especie de bacteria. El modelo, que es una red neuronal convolucional profunda, funciona con altos niveles de confiabilidad para imágenes provenientes de la base de datos DiBaS (con la que fue entrenada), pero falla cuando se presentan imágenes provenientes de otros lugares (Figura 2.1).



**Figura 2.1.** Ejemplo de la implementación de un clasificador de bacterias, disponible públicamente y entrenado con la base de datos de DiBaS [10]. Se muestra el resultado al probarlo con una imagen proveniente de [44]. **A la izquierda: Superior,** se presenta la imagen alimentada al clasificador, en este caso de la especie *Escherichia coli*. **Inferior,** se agregó la especie real de la imagen y una imagen ejemplo de la especie que predice el modelo. **A la derecha,** se presenta el resultado de la predicción, se interpreta que clasifica con un 91% de probabilidad de que la bacteria es *Micrococcus spp.*, un 5% *Lactobacillus reuteri*, un 3% *Acinetobacter baumannii* y los demás 0%.

Por ejemplos como este y sabiendo que utilizar métodos de procesamiento de imágenes o de aprendizaje automático tradicionales puede resultar en fallos debido a la elección de umbrales o la gran variabilidad entre las imágenes como se menciona en [4] y en [9], se justifica la relevancia y contribución de este trabajo para la computación, al proponer un ensamble basado en redes neuronales profundas para clasificar bacterias en imágenes adquiridas con distintos microscopios. Buscando aprovechar la capacidad de generalización de las redes neuronales profundas, así como la flexibilidad que ofrece un ensamble de modelos para que cada una de las partes se enfoquen en características definidas, reduciendo así la posibilidad de predicciones con diferencias tan grandes a las clasificaciones reales, y permitiendo agregar una mayor cantidad de especies al ir reduciendo el espacio de búsqueda con mayor certeza.

## 1.4. Hipótesis

Se proponen dos hipótesis a estudiar:

1. Un ensamble de redes neuronales convolucionales profundas, entrenado y probado en imágenes adquiridas con microscopía óptica de campo claro y provenientes de distintas fuentes, realizará con exactitud alta ( $> 90\%$ ) una clasificación multiclase de distintas especies de bacterias.
2. Es posible identificar entre especies de bacterias del mismo género basándose únicamente en características morfológicas presentes en imágenes tomadas por medio de microscopía de campo claro de muestras identificadas y preparadas previamente (este problema se explica más a detalle en la sección 2.1.3).

## 1.5. Objetivos

### a. General

Diseñar un sistema de clasificación multiclase que permita el reconocimiento de bacterias en imágenes provenientes de distintas fuentes a partir de sus características fenotípicas, utilizando redes neuronales convolucionales.

### b. Específicos

- Definir las características más relevantes para la clasificación de bacterias, así como analizar en el estado del arte cuáles son los métodos que se han implementado para esta tarea.
- Recopilar de distintas fuentes la base de imágenes que se utilizará. Prepararla definiendo las etiquetas, segmentaciones, tamaño de las instancias, etc.
- Probar distintos métodos para segmentar las imágenes de las bases de datos encontradas; estos métodos pueden ser tales como procesamiento de imágenes, K-means, redes neuronales convolucionales, entre otros.
- Definir la arquitectura del modelo que compondrá el ensamble de redes neuronales.
- Entrenar el ensamble de redes neuronales para la clasificación de bacterias, dividiendo las imágenes para entrenamiento, validación y prueba.
- Calcular métricas de desempeño de cada modelo dentro del ensamble.

## 2. Marco Teórico

### 2.1. Identificación tradicional y morfología de la bacteria

Existen diferentes maneras para identificar bacterias, una de las estrategias más utilizadas por su facilidad y bajo costo es la identificación por características fenotípicas. Esta metodología se fundamenta en identificar las características morfológicas, y en general, observables de las bacterias. Por lo que es muy común hacer identificación bacteriana desde el punto de vista microscópico, aclarando esto porque es importante mencionar que también es posible identificar de una manera macroscópica, por ejemplo, en función al análisis, clasificación y crecimiento del medio de cultivo [2].

Entre las principales características fenotípicas utilizadas para la caracterización e identificación de bacterias, podemos encontrar las siguientes [1, 2, 17]:

- **Forma:** La forma o morfología de las bacterias es una característica fundamental en su identificación. Puede existir una gran variedad de formas entre las bacterias (ver sección 2.1.1).
- **Tamaño:** El tamaño de las bacterias varía considerablemente, desde menos de un micrómetro hasta varios micrómetros de longitud, este tamaño puede influir en su observación y manejo en el laboratorio.
- **Textura:** La textura de las colonias bacterianas al crecer en medios de cultivo puede ser rugosa, lisa, viscosa, seca, entre otros. Esta característica se observa macroscópicamente y proporciona información sobre la composición de la pared celular y otros componentes extracelulares.
- **Flagelos (si tienen):** Los flagelos son estructuras filamentosas que permiten la movilidad de las células, estos pueden manifestarse en diferentes configuraciones. Su presencia, ausencia y disposición es importante para la identificación ya que están directamente relacionados con la capacidad de la bacteria para moverse en su entorno.
- **Si son encapsuladas o no:** Algunas bacterias producen una capa protectora de polisacáridos rodeando la célula, esta es conocida como cápsula. La presencia de una cápsula puede ser visualizada con tinciones especiales y es una característica distintiva de algunas especies.
- **Agrupaciones de las células:** Las células pueden acomodarse de distintas maneras, como en pares, en fila, etc. (ver sección 2.1.1)
- **Medio de cultivo:** El tipo de medio de cultivo en el que una bacteria crece puede proporcionar información valiosa. Algunos medios son selectivos, lo que significa

que permiten el crecimiento de ciertas bacterias mientras inhiben otras, o permiten la distinción entre especies basándose en reacciones bioquímicas.

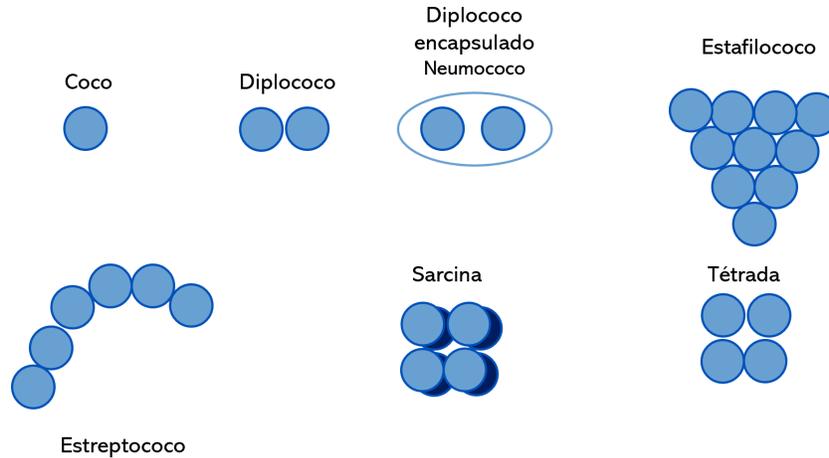
- Tipo y reacciones a tinciones: Existen distintos tipos de tinciones bacterianas, algunos ejemplos de ello son la tinción de Gram y la tinción ácido-alcohol resistente, estas tinciones son utilizadas para diferenciar bacterias basándose en las características de su pared celular y otras estructuras características.
- Margen de crecimiento a distintas variables del medio de cultivo: Las bacterias tienen diferentes tolerancias a factores ambientales como la temperatura, el pH, la concentración de oxígeno, y la disponibilidad de nutrientes. Estos márgenes de crecimiento pueden ser observados y medidos en experimentos controlados.

Sin embargo, aún conociendo estas características es importante recalcar que no siempre pueden ser medidas al mismo tiempo o con gran detalle, por lo que en muchos casos no son suficientes de forma individual para distinguir especies bacterianas.

### **2.1.1. Cocos, bacilos y otras morfologías**

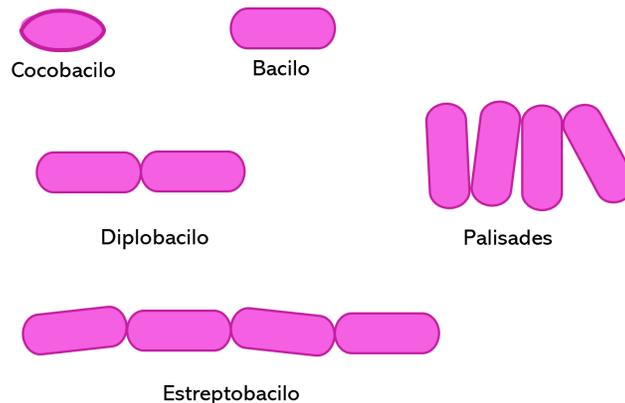
Existe una amplia diversidad de formas en las bacterias, presentando una gran variación en la apariencia, con especies que pueden verse como rectángulos, cuadrados, estrellas, triangulares, filamentosas o alargadas, esféricas, o incluso elipsoidales. En el caso de este trabajo, nos enfocaremos en identificar cocos y bacilos (ver Figura 3 y Figura 4), debido a que existe mayor información disponible de estas morfologías.

Los cocos son bacterias de forma esférica cuyo tamaño varía entre 0.5 y 1.5 micrómetros. En la literatura, sus subtipos están relacionados con la forma en que se agrupan. Aquellos que permanecen individuales se llaman cocos; si se agrupan en pares, se conocen como diplococos; cuando forman cadenas, se denominan estreptococos y los que forman estructuras en forma de racimos, estafilococos. Estas características morfológicas pueden ser usadas para su identificación/clasificación, ya que dependiendo de su agrupación pueden derivar en distintas patologías. Entre las especies de cocos conocidas podemos encontrar: *Micrococcus luteus*, *Staphylococcus aureus*, *Streptococcus pneumoniae*.



**Figura 3.** Formas que pueden tomar las bacterias de tipo cocos. Imagen basada en [18].

Por otro lado, los bacilos son bacterias con forma elipsoidal o alargada, comúnmente descritas como bacterias en forma de bastón. Su tamaño varía más que el de los cocos, con algunas especies que pueden llegar a medir hasta los 8 o 10 micrómetros, sin embargo, también se encuentran en tamaños más pequeños, como de 0.5 a 2 micrómetros. Entre las especies conocidas de bacilos más importantes para los humanos se encuentran: *Salmonella*, *Escherichia Coli* y *Klebsiella pneumoniae*, por mencionar algunos ejemplos [1]. De manera similar a los cocos, también se pueden agrupar de muchas maneras y, dependiendo de esta agrupación tener diferentes denominaciones y distintas características.



**Figura 4.** Algunas agrupaciones que pueden tomar las bacterias de tipo bacilos. Imagen basada en [18].

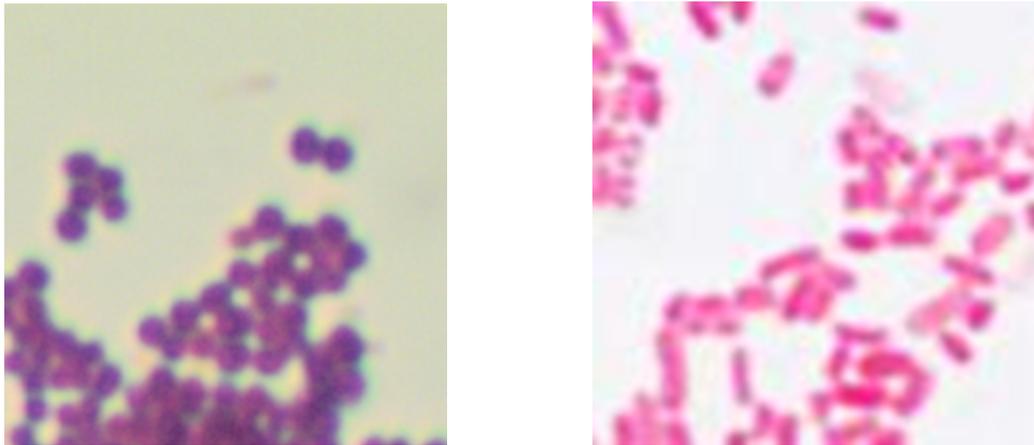
### 2.1.2. Tinción de Gram

De manera general, es importante remarcar que las bacterias son organismos transparentes lo cual dificulta su observación a través de los microscopios, esto ha llevado al desarrollo de técnicas de tinción que permitan su observación y diferenciación del fondo.

El método tradicional que se utiliza en la identificación de bacterias es la tinción de Gram, la cual se basa en el uso de una mezcla de colorantes que se aplica sobre bacterias fijadas para poder observarlas mejor bajo el microscopio [19].

Según las características de su pared celular, las bacterias Gram positivas se tiñen de un color violeta o azul, mientras que las Gram negativas adoptan un color rosa o rojo (ver Figura 5).

Históricamente, este tipo de tinción ofrece una de las primeras taxonomías de bacterias que fueron conocidas, aunque en la actualidad ésta no se considera una clasificación natural.



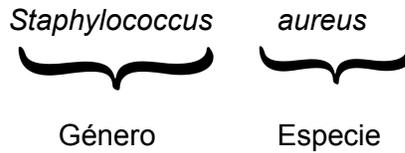
**Figura 5.** Izquierda. Ejemplo de tinción de Gram positiva (color púrpura o violeta) [10]. Derecha. Ejemplo de tinción de Gram negativa (color rosado) [40]. Las imágenes, adquiridas en microscopio de campo claro, pertenecen a distintas especies.

### 2.1.3. Sobre la especie y género: taxonomía

Como se mencionó en secciones anteriores, a partir de la tinción de Gram y de las características morfológicas, es posible comenzar a identificar y agrupar entre los distintos tipos de bacterias que existen por medio del color y la forma. De manera general, al agrupamiento entre características similares se le conoce como taxonomía [17]. Y, aunque en este caso no estemos lidiando con una taxonomía como tal (ya que no es una clasificación que asegure que tengamos ancestros en común) podemos utilizarla para inferir cosas sobre las especies de bacterias.

Así, a cada grupo de bacterias (con determinadas características en común) se les asigna un nombre. Este nombre es compuesto por dos palabras que indican el género y

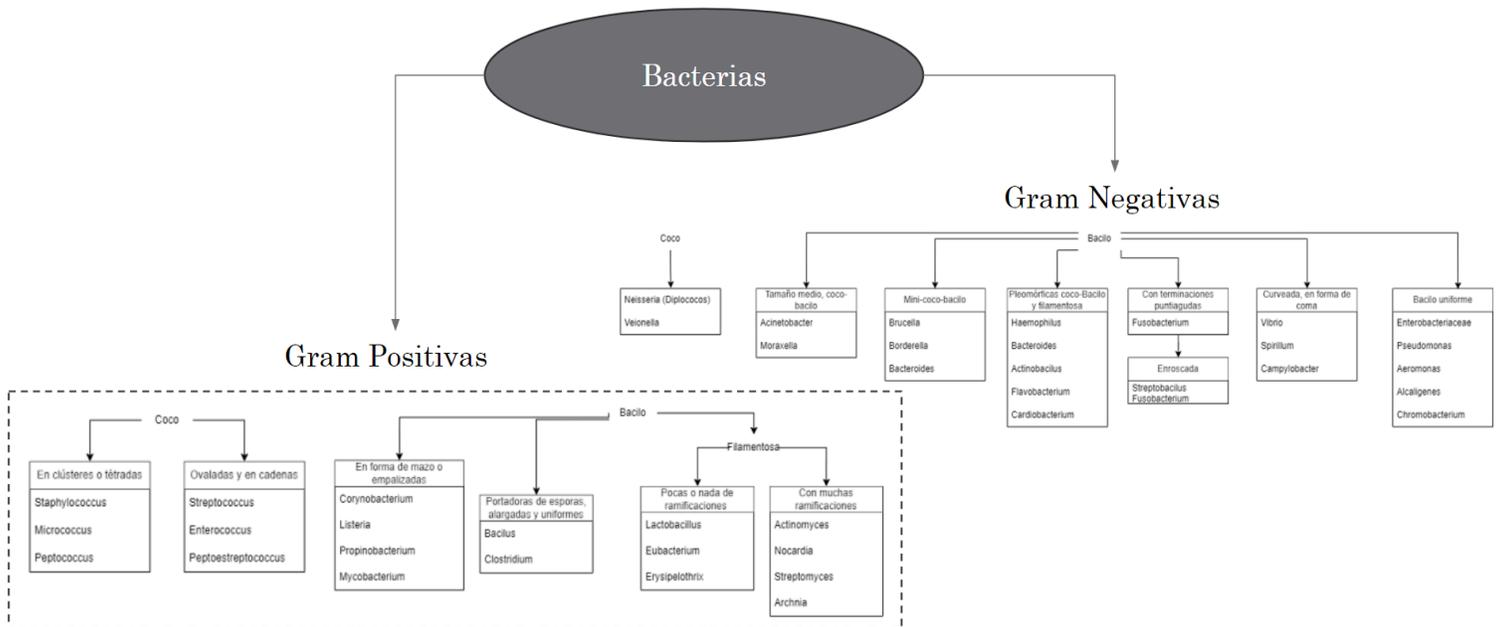
la especie, en ese orden y de manera jerárquica. Empezando con el género en mayúscula y la especie en minúscula, por ejemplo para el *Staphylococcus aureus*:



Naturalmente, pueden existir muchas especies pertenecientes al mismo género, estas suelen ser fenotípicamente muy parecidas entre sí, de ahí la dificultad de intentar identificarlas exclusivamente con características fenotípicas.

Por ejemplo, para el género *Staphylococcus* existen otras especies además de la mencionada *aureus*, como lo son *epidermis* y *saprophyticus*, por hablar de algunas otras utilizadas en este trabajo (ver inciso 5 de la sección 3.4).

En la Figura 6 se muestra el esquema de un agrupamiento (lo tomaremos a símil con la taxonomía) que describe una forma de identificación fenotípica de las bacterias, acotado a dos morfologías. Además, este esquema ilustra el amplio espacio de búsqueda que conlleva la tarea de identificar distintas especies de bacterias.



**Figura 6.** Esquema de agrupamiento de especies de acuerdo a su forma y tinción de gram. El esquema está limitado a solo dos tipos de morfologías, cocos y bacilos. Es importante remarcar que se trata de una clasificación no natural. Basado en [20].

## **2.2. Procesamiento digital de imágenes**

El procesamiento digital de imágenes es un conjunto de técnicas que se aplican a las imágenes con la finalidad de mejorar la calidad de la presentación o poder facilitar la extracción de la información de ellas [21, 22].

Las técnicas más aplicadas en el procesamiento digital para el reconocimiento o la clasificación de imágenes son:

- Transformación a escala de grises
- Filtrado de la imagen
- Binarización
- Detección de bordes

### **2.2.1. Histogramas de color RGB**

En las imágenes digitales, un histograma de color representa la cantidad de píxeles que tienen cierta intensidad en cada uno de los canales de colores fijos dentro del espacio de color en turno, regularmente este espacio es el conocido como RGB aunque existen otros espacios que pueden ser usados [23, 24].

El espacio RGB (Rojo, Verde, Azul en español) es una representación de cómo se forman los colores a partir de las combinaciones e intensidades de estos tres. En este modelo cada imagen está representada por tres histogramas, cada uno perteneciente a un canal del espacio, que en conjunto conforman la imagen con un amplio abanico de colores. Cada canal se representa en una escala de intensidades (de manera similar a la escala de grises) por lo que los valores se dividen en un rango. Cuando los valores de estos histogramas son los más bajos implica la ausencia del color dentro del rango por lo que la imagen se presentará en un color negro, si por el contrario, se toman los valores más altos del rango entonces la imagen se presentará en un color blanco.

### **2.2.2. Especificación de histograma**

Dentro del procesamiento de imágenes existen técnicas que nos permiten modificar el contenido de los histogramas de la imagen que tenemos, tal vez con la intención de mejorar el contraste aplicando la ecualización del histograma<sup>8</sup>, o posiblemente queramos darle al histograma en una forma específica (la forma de otro histograma) que se adapte mejor a nuestras necesidades. Esta última transformación se conoce como especificación de histograma.

Para este procedimiento necesitamos dos imágenes, una con el contenido que queremos preservar (imagen original o de contenido), y otra, que contiene el histograma que queremos igualar (imagen de referencia).

---

<sup>8</sup> Ecualización del histograma: Es una técnica de procesamiento de imágenes que consiste en hacer que en un histograma se redistribuyan los niveles de intensidad de una imagen, de manera que el histograma se aplane y se extienda.

De manera concisa, el proceso para realizarlo es el siguiente [25]:

- 1) Calcular los histogramas de la imagen original y de la imagen de referencia.
- 2) Calcular la distribución de densidad de las intensidades en cada histograma a partir de la cantidad de píxeles por intensidad dentro de toda la imagen, dicho de otra manera, dividir la cantidad de píxeles de una intensidad  $j$  entre la cantidad de píxeles totales de la imagen. Esto se debe hacer por cada intensidad (e histograma) presente dentro de las imágenes. De este punto obtenemos una tabla que llamaremos **Pr** con entradas  $r_j$ , donde  $r_j$  es la densidad de la intensidad en turno.
- 3) Mapear la distribución de intensidades calculadas mediante una transformación dada por la ecuación:

$$T(r) = (L - 1) \sum_{j=0}^k Pr(r_j), \quad k = 0, 1, 2 \dots L - 1$$

donde  $L$  es la cantidad de niveles deseados para la representación de valores (para una imagen de 8 bits,  $L = 256$ ), **Pr** es la tabla de probabilidad calculada en el punto 2 con cada entrada  $r_j$ , y  $j$  es el valor de intensidad sobre el que se está calculando la transformación. Se deben redondear los valores resultantes a un número entero.

- 4) Una vez mapeados los histogramas de ambas imágenes (de referencia y contenido), el paso siguiente es sencillamente encontrar coincidencias o similitudes dentro del mapeo transformado entre los resultados del histograma de la imagen original y el de referencia, es posible que se encuentre más de una coincidencia por lo que siempre se debe tomar la de menor valor numérico.
- 5) Formar la imagen especificada, asignando a cada píxel de la imagen de contenido el valor de representación del histograma de referencia según las coincidencias encontradas en el resultado del mapeo previo.

Todo el proceso se aplica en el histograma de cada canal RGB de las imágenes, por supuesto manteniendo concordancia rojo con rojo, verde con verde y azul con azul.

### 2.3. Algoritmo para clasificación

Dentro del aprendizaje supervisado, un algoritmo de clasificación se puede definir como una serie de pasos definidos con el objetivo de predecir etiquetas de clases discretas de instancias no conocidas previamente, basándose en información conocida y etiquetada en el pasado. Entre estos algoritmos de clasificación encontramos las redes neuronales convolucionales.

### 2.3.1. Redes Neuronales Convolucionales

Las redes neuronales convolucionales [21] son un tipo especializado de arquitectura de red neuronal artificial diseñada para analizar información bidimensional. A diferencia de las redes neuronales tradicionales, que se basan únicamente en multiplicaciones de matrices, las *CNN*<sup>9</sup> utilizan la operación de convolución, que implica deslizar un filtro (o *kernel*) a través de los datos de entrada, como una imagen, para capturar patrones o características locales específicas.

Por lo tanto, el componente central de las *CNN* son las capas convolucionales, que extraen características aplicando filtros que detectan bordes, formas y otros elementos visuales. Además, las capas de agrupamiento (*pooling*<sup>9</sup>) reducen la muestra de los datos para disminuir la dimensionalidad y el costo computacional mientras se enfocan en las características esenciales, esto es lo que le permite a la red aprender patrones complejos y representaciones jerárquicas de los datos.

Algunas aplicaciones principales de las *CNN* en el campo de visión por computadora, son [21]:

- Clasificación de imágenes, donde pueden reconocer, categorizar figuras y patrones dentro de una imagen
- Detección de objetos, que implica identificar y localizar objetos dentro de una imagen
- Segmentación de imágenes, que consiste en asignar categorías a cada píxel de la imagen.

Las *CNN* ofrecen ventajas significativas, como aprender de manera eficiente características espaciales sin la necesidad de diseñar manualmente estas características. En general, las redes neuronales convolucionales son una herramienta poderosa para analizar datos visuales, es por ello que han revolucionado el campo de la visión por computadora.

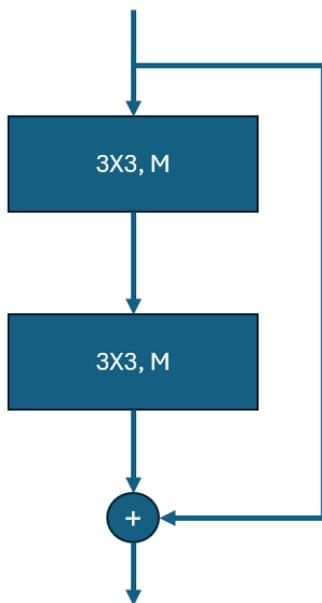
### 2.3.2. Arquitectura ResNet

*ResNet*, abreviatura de *Residual Net*, es una red neuronal convolucional utilizada para tareas de clasificación de imágenes. Fue diseñada principalmente para resolver el problema del desvanecimiento del gradiente en las redes neuronales profundas. Se presentó por primera vez en el artículo *Deep Residual Learning for Image Recognition* [26].

La arquitectura de estas redes consiste en el agregado de bloques residuales entre las capas de la red, estos bloques son llamados saltos de conexión o “*skip connections*”, tienen la intención de que la red aprenda la diferencia (el residual) entre la entrada y la salida de cada bloque (Figura 7).

---

<sup>9</sup> Las capas de agrupamiento o, en inglés, *pooling* se utilizan para reducir las dimensiones espaciales (ancho y alto) de los mapas de características. Consiste en promediar, o tomar el valor máximo o mínimo, de secciones determinadas de la imagen, esto ayuda a hacer la red más eficiente computacionalmente y reduce las posibilidades de sobreajuste.



**Figura 7.** Bloque Residual (BR) propuesto en [26] de la red ResNet para arquitecturas de 18 y 34 capas. Cada rectángulo es una capa convolucional, la multiplicación 3x3 indica el tamaño del filtro, mientras que la letra M indica la cantidad de mapas de salida de cada convolución (el número de filtros), esta es variable dependiendo de la posición en la red. El salto de conexión se presenta en el operador de suma entre la entrada y la salida del bloque.

Los bloques residuales se definen de distintas maneras dependiendo de la profundidad de la red que se quiera entrenar. Por ejemplo, en la Tabla 1 [26] se muestra el tamaño de las capas para las arquitecturas utilizadas en este trabajo, ResNet 18 y 50.

Nombre de la capa	18 Capas	50 capas
Layer 1	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Layer 2	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Layer 3	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Layer 4	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$

**Tabla 1.** Descripción de la arquitectura dependiendo de la profundidad de la red, para 18 y 50 capas. Cada corchete indica un bloque residual como se muestra en la Figura 7, estos bloques se repiten las veces indicadas en la multiplicación a la derecha. Se omiten las capas de entrada y salida que complementan la profundidad de la red. Tabla modificada, tomada de [26].

### 2.3.3. Ensamblajes

Un ensamble, en el contexto de redes neuronales y aprendizaje automático, se refiere a una técnica en la que múltiples modelos se combinan para mejorar el rendimiento general del sistema en alguna tarea. La idea principal detrás del ensamble es que diferentes modelos o redes pueden captar diferentes aspectos de los datos, y así, al combinarlos se puede obtener una predicción más robusta y precisa [29].

### 2.3.4. Módulo de normalización SNR

La normalización y restitución de estilo (*Style Normalization and Restitution*, SNR) [30], es una técnica para reducir las variaciones de estilo irrelevantes y resaltar las características importantes en las imágenes.

Consiste en un módulo aplicable a la salida de cada capa convolucional de cualquier arquitectura *CNN*, que reduce la cantidad de distribuciones de los mapas de características en el conjunto de entrenamiento mediante la normalización de los datos de entrada en cada capa. Ya que el proceso de normalización puede llegar a eliminar rasgos relevantes de la imagen, en el módulo se propone decidir cuáles son las características importantes que han sido eliminadas después de la normalización, y que estas se sumen para continuar en la siguiente capa (ver Figura 8).

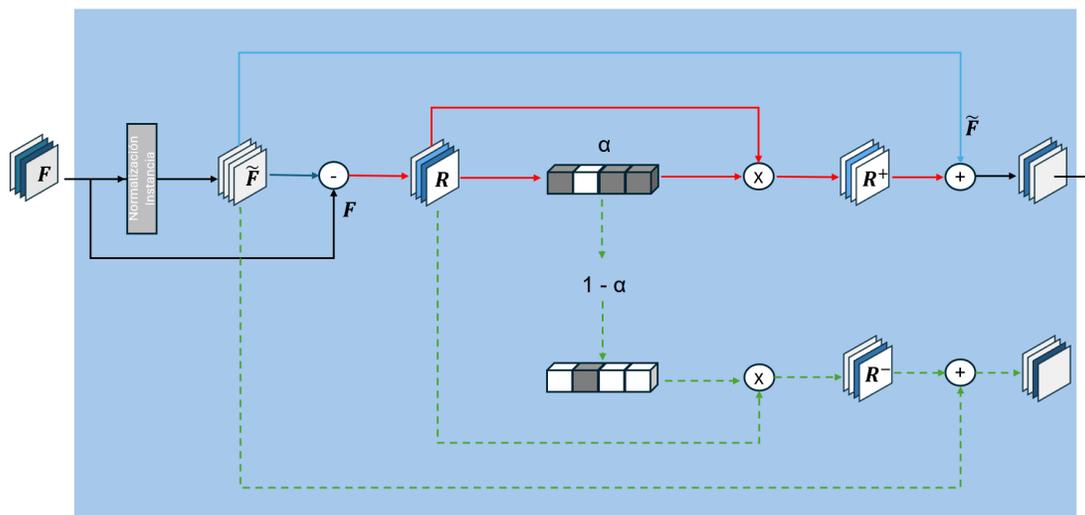


Figura 8. Arquitectura del módulo SNR. Basada en [30].

El módulo funciona de la siguiente manera: Al recibir los mapas de características  $F$ , provenientes de una capa convolucional previa, se les aplica normalización de instancia (IN), después se resta al contenido original,  $F$ , el resultado de la normalización,  $\tilde{F}$ , para conseguir la información que se perdió al realizar la normalización, así obteniendo  $R$ . Posteriormente,  $R$  entra a una red  $\alpha$ , compuesta por dos capas convolucionales con una función de activación ReLU y otra sigmoide, respectivamente; que aprenden a discriminar la información relevante  $R^+$  (flechas rojas en la Figura 8) de la no relevante  $R^-$  (flechas verdes punteadas en la Figura 8). Por último, se suma el resultado de la

normalización a la información relevante encontrada por la red ( $F^{\sim} + R^+$ ) y se continúa a la siguiente capa convolucional de la arquitectura.

Se reporta que los modelos entrenados con SNR son menos susceptibles a errores causados por cambios en la iluminación, el fondo o las características de las imágenes.

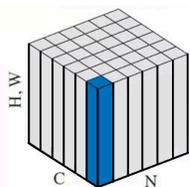
Ya que la normalización es un concepto relevante en este módulo y justificando su importancia, podemos decir que un desafío común en el entrenamiento de redes neuronales profundas es la variación en la distribución de los datos a medida que avanzan por la red, estos cambios pueden dificultar el aprendizaje adecuado en las redes. Por lo que la normalización es una técnica que ayuda a estabilizar esta distribución, mejorando la velocidad y estabilidad del entrenamiento. Existen diferentes tipos, como la normalización por lotes, por capas y por instancia, cada una calculando la media y desviación estándar a diferentes niveles de los datos para normalizar la información.

En este trabajo aplicamos dos tipos de normalización:

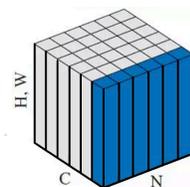
- Por instancia (IN, por las siglas en inglés de Instance Normalization): Aplicada al principio del módulo SNR, en este caso se calcula la media y desviación estándar de cada canal individual para un solo ejemplo.
- Por lote (BN, por las siglas en inglés de Batch Normalization): Un cambio que proponemos en el módulo en lugar de la IN, aquí se calcula la media y desviación estándar a través de un canal para todo el lote de entrenamiento.

De manera gráfica estos tipos de normalización son:

Normalización de instancia



Normalización por lote



**Figura 9.** Definición visual de la normalización por instancia y por lote, C son los canales o filtros de la capa, N la cantidad de instancias del lote y H,W son la altura y el ancho de las imágenes respectivamente, representados en el eje z del cubo. Imagen inspirada en [31].

## 2.4. Pruebas y validación de clasificación

### 2.4.1. Matrices de confusión

Una matriz de confusión [32], es una herramienta que se utiliza para evaluar el rendimiento de un modelo de clasificación. Es una matriz de tamaño  $n \times n$ , donde n es el número de clases en el problema, las filas corresponden a las clases reales, y las columnas a las clases predichas por el modelo. Sirve para mostrar cuándo una clase es confundida con otra. Por esto, permite analizar distintos patrones de errores y usar esta

información para mejorar el rendimiento del clasificador o para decidir si es aceptable o no.

Por ejemplo, en una clasificación binaria, se consideran las clases que se saben reales p (positivo) y n (negativo), y las clases predichas por el modelo P (positivo) o N (negativo).

	N (Predicción del modelo)	P (Predicción del modelo)
n (Etiqueta real)	Verdaderos Negativos	Falsos Positivos
p (Etiqueta real)	Falsos Negativos	Verdaderos Positivos

Tabla 2. Ejemplo de una matriz de confusión binaria.

A partir de la matriz de confusión, se pueden obtener distintos tipos de métricas para poder validar la capacidad de reconocimiento del algoritmo [32]. Considerando **TP**: Verdaderos Positivos (*True Positives*), **TN**: Verdaderos Negativos (*True Negatives*), **FP**: Falsos Positivos (*False Positives*), **FN**: Falsos Negativos (*False Negatives*).

Estas métricas están definidas como:

- Exactitud: La relación entre el número de predicciones correctas (ya sean positivas o negativas) y el total de predicciones posibles dentro del conjunto de prueba.

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precisión: La división entre los positivos reales encontrados por el modelo y todos los casos positivos clasificados (en otras palabras, incluyendo los falsos positivos).

$$\text{Precisión} = \frac{TP}{TP + FP}$$

- Especificidad: Son los casos negativos que el modelo clasificó de manera correcta.

$$\text{Especificidad} = \frac{TN}{TN + FP}$$

- Tasa de Falsos Positivos: La proporción de casos negativos que fueron equivocadamente clasificados como positivos.

$$\text{Tasa de Falsos Positivos} = \frac{FP}{TN + FP}$$

- Tasa de Falsos Negativos: La proporción de casos positivos equivocadamente clasificados como negativos.

$$\text{Tasa de Falsos Negativos} = \frac{FN}{TP + FN}$$

- Tasa de Verdaderos Positivos: Son los casos positivos que el modelo clasificó de manera correcta.

$$\text{Tasa de Verdaderos Positivos} = \frac{TP}{TP + FN}$$

## 2.4.2. Validación cruzada

La validación cruzada de K-iteraciones [33] es una técnica utilizada para hacer análisis estadístico, particularmente en los problemas de clasificación. Su objetivo es asegurar que el modelo no solo se ajusta bien a los datos con los que se ha entrenado, sino que también tiene una buena capacidad de generalización para datos no vistos, esto independientemente del conjunto de pruebas con el que se calcule la exactitud del modelo. Es una técnica especialmente útil para evitar problemas de sobreajuste<sup>10</sup> y subajuste<sup>11</sup>.

El proceso consiste en la separación en K grupos de la información que se usará para entrenar el modelo, K es variable y usualmente se elige de entre 5 a 10 grupos [34], cada grupo tiene aproximadamente el mismo tamaño. Después se toman K-1 grupos y se utilizan para entrenar el modelo, el grupo restante se utiliza como conjunto de prueba para calcular las métricas del modelo. Este proceso es iterativo hasta que todos los grupos hayan servido como conjunto de validación, el resultado de la validación es el promedio de la métricas de las K pruebas.

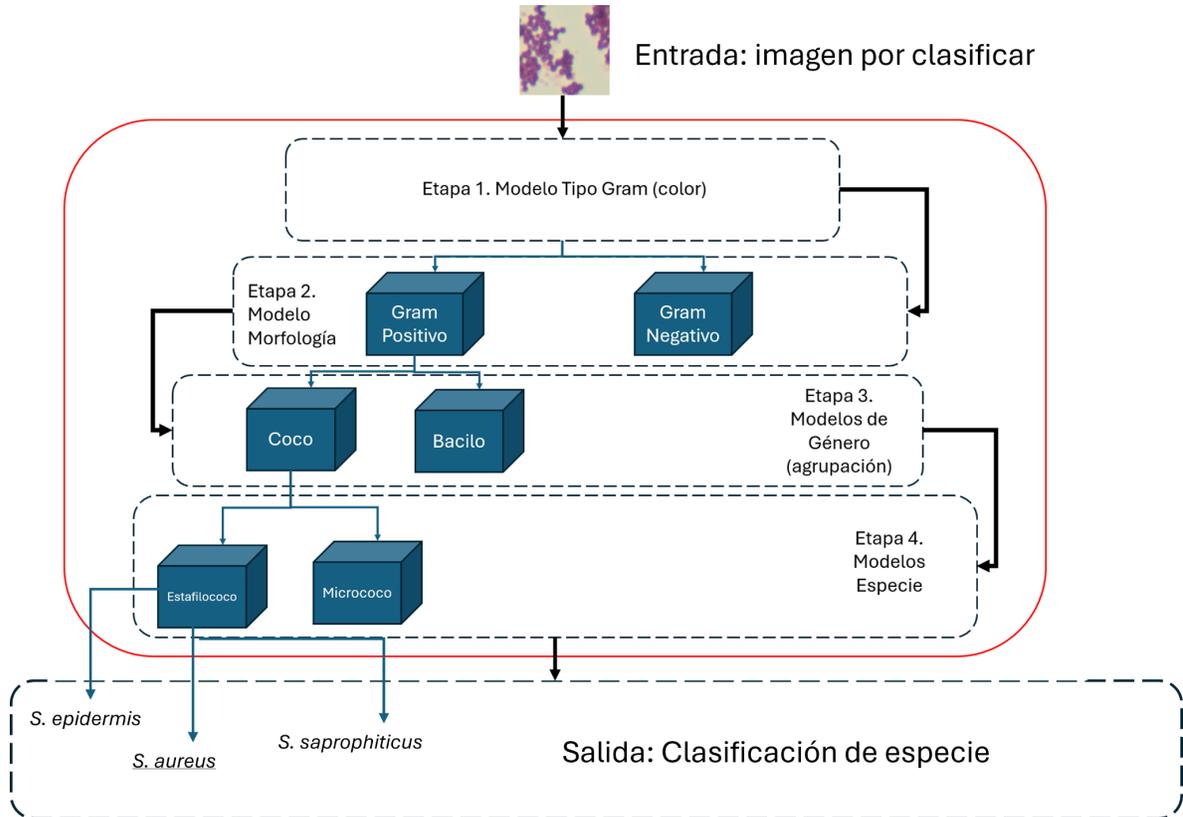
Es importante considerar que la información se debe de dividir de forma aleatoria procurando mantener equilibrada la representación de cada clase dentro de cada grupo.

---

<sup>10</sup> *Overfitting*: Es cuando el modelo se ajusta demasiado a la información de entrenamiento.

<sup>11</sup> *Underfitting*: Cuando el modelo no puede encontrar una tendencia en los datos de entrenamiento.

### 3. Metodología



**Figura 10.** Esquema del ensamblado en cascada propuesto (cuadro rojo). En los cuadros con líneas punteadas se presentan las etapas que contienen el/los modelo(s) y en los bloques azules un ejemplo del funcionamiento para identificar al *Staphylococcus aureus*, por simplicidad se obvian las ramificaciones de Gram negativo aunque también se propone que existan al clasificar otras especies.

La metodología propuesta es la creación de un ensamblado de redes neuronales convolucionales dispuesto en forma reminiscente a un árbol de decisión basado en la clasificación jerárquica como la presentada en la Figura 6.

La Figura 10, muestra el funcionamiento del ensamblado propuesto:

- La imagen entra a la primera etapa, dónde el modelo clasifica entre el tipo de tinción de Gram (positivo o negativo).
- La salida de la primera etapa se pasa al segundo modelo para clasificar la morfología, en este caso coco o bacilo.
- En la etapa tres se determina el género de la bacteria.
- Por último, en el modelo final se clasifica la especie presente en la imagen.

Cada modelo se entrenó de manera especializada en la etapa que está clasificando, considerando las características que recibe de la etapa anterior de forma jerárquica. De esta manera, para la etapa 3 y 4 tendremos distintos clasificadores dependiendo de las

combinaciones de los anteriores, en la sección 3.5 se explica a fondo la metodología de los modelos entrenados.

### 3.1. Bases de datos

La información se recopiló de distintas bases de datos, la principal fue DiBaS dataset [10] de la que se obtuvieron 32 especies, con aproximadamente 20 imágenes de cada una, para un total de 717 imágenes. Para abordar el problema de la diferencia de características entre imágenes obtenidas de distintas bases de datos, seleccionamos tres especies, *Staphylococcus aureus*, *Escherichia coli* y *Micrococcus spp.*, también presentes entre las 32 especies que se obtuvieron de la base de datos DiBaS. De estas tres especies se obtuvieron 28 imágenes de 17 lugares diferentes (ver Tabla 3), con aproximadamente 10 imágenes por especie.

Fuente	Número de Imágenes	Fuente	Número de Imágenes
Microbiology in Pictures [35]	4	Researchgate [43]	2
Weber University [36]	3	Bacteria in Photos [44]	2
Medical labs [37]	1	GettyImages [45]	1
Flickr The Rubin Lab [38]	2	Gmspop [46]	1
IStockPhoto [39]	3	College of Baltimore County [47]	1
MDS Manuals [40]	1	Quizlet [48]	1
Bioimagen [41]	1	MicroBiosh [49]	1
Wikipedia [42]	2	Medium [50] / Shutterstock [51]	1 / 1

**Tabla 3.** Fuentes de las que se obtuvieron las imágenes de *Staphylococcus aureus*, *Escherichia coli* y *Micrococcus spp.*

La base de datos final consta de 745 imágenes provenientes de distintas fuentes. Sin embargo, se puede notar que existe una proporción mayor de las imágenes de una sola base de datos, por lo que para mitigar que las redes solo aprendan y se adapten a identificar exclusivamente las imágenes de esta fuente, se aplicaron técnicas de aumento de datos<sup>12</sup> presentadas en la siguiente sección.

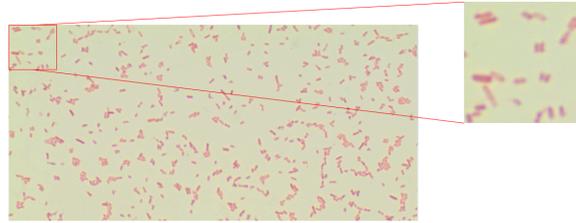
### 3.2. Procesamiento de imágenes

Para contar con una mayor cantidad de imágenes por especie que permitieran entrenar adecuadamente una red neuronal, se realizó un preprocesamiento de las imágenes aplicando técnicas de aumento de datos.

Primero, debido a la cantidad bastante limitada de imágenes por especie (~20 imágenes), se decidió recortarlas para aprovechar su gran tamaño. De esta manera, se

<sup>12</sup> El aumento de datos consiste en generar artificialmente nuevos datos de entrenamiento a partir de un conjunto de datos original. En el contexto de *CNN*, esto se aplica principalmente a imágenes.

obtuvieron imágenes de dimensiones menores, logrando aumentar el número de muestras por especie. En general, se recortaron imágenes cuyo tamaño original era de aproximadamente 1,000 x 2,000 píxeles para obtener imágenes de aprox. 200 x 200 píxeles (véase Figura 11).

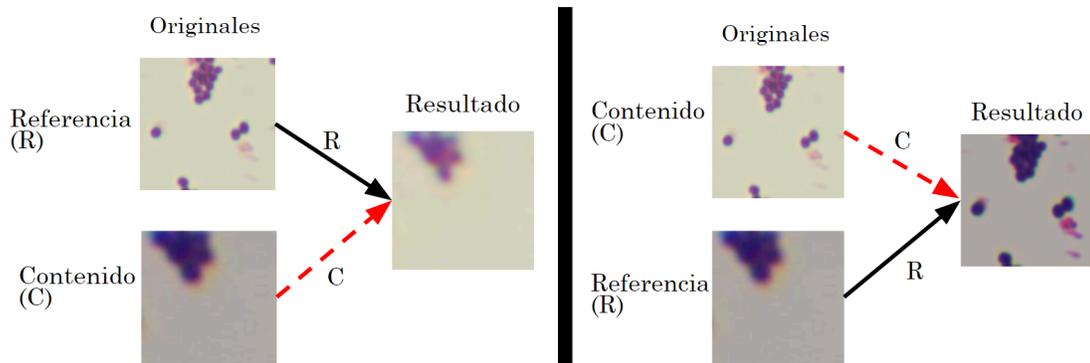


**Figura 11.** Ejemplo de recorte de imagen para aumentar la cantidad de datos. Los recortes realizados no están superpuestos.

De esta manera, aumentamos de 745 imágenes a más de 10,000 imágenes disponibles, con aproximadamente 1,000 imágenes por especie. Se descartaron aquellos recortes de imágenes que sólo contenían el fondo y no células identificables mediante inspección visual.

Posteriormente, buscando mejorar la generalización de las redes a las diferentes características visuales que pueden tomar las imágenes, se hizo aumento de datos por medio de la especificación del histograma, utilizando la implementación disponible en la función *match\_histograms* de *scikit-image* [52].

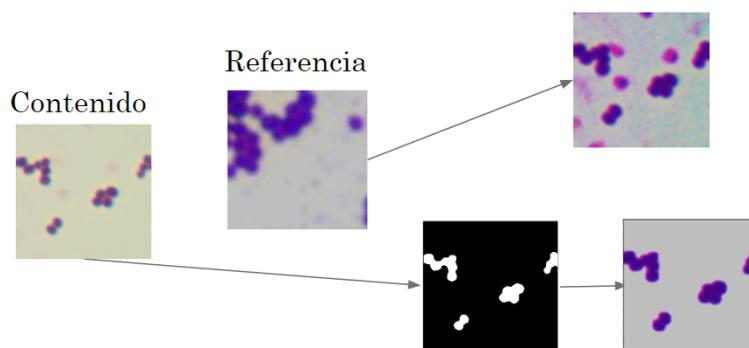
La especificación de histograma se realizó entre imágenes de distintas fuentes pero de la misma especie de bacteria, concretamente para *Staphylococcus aureus*, *Escherichia coli* y *Micrococcus spp.* Se tomaron ejemplos de las fuentes con menor representación de imágenes y se utilizaron como referencia para aplicar la especificación de histograma a otras imágenes, elegidas de manera aleatoria, de la base de datos DiBaS. Así, se generaron mil imágenes “nuevas” por especie, que toman el aspecto de aquellas fuentes con poca representación. A estas imágenes resultantes también se les aplicó rotación al azar entre 0, 90, 180 y 270 grados.



**Figura 12.** Ejemplos del resultado de la especificación del histograma entre dos imágenes provenientes de distintas bases de datos. La flecha negra representa la imagen de Referencia (R) a la que se quiere igualar el histograma, mientras que la roja representa la imagen de Contenido (C) que cambiará para especificarse a la de referencia. Se presentan ambos resultados de alternar entre imagen de Contenido y Referencia.

El resultado de la especificación no siempre alcanza la calidad presentada en la Figura 12. Algunas imágenes resultantes presentan un alto nivel de ruido y parecen extremadamente contaminadas, lo cual puede ser perjudicial para el entrenamiento de las redes; por lo que en algunos casos se mejoraron las especificaciones por medio de máscaras de segmentación de las imágenes.

Para ello, primero se obtienen las máscaras de las imágenes a través de umbralización y operaciones morfológicas (de apertura y cierre), después se hace un análisis visual de las máscaras resultantes y se corrigen a mano. Una vez obtenidas las máscaras, se aplica especificación de histograma entre las imágenes de referencia y contenido de forma normal. Posteriormente, a la máscara de segmentación de la imagen de contenido se le modifica pixel a pixel para tener una imagen con colores similares a los de la imagen de referencia (se toma el color de un pixel de célula y de fondo de referencia). A esta máscara se le aplica dos veces especificación de histograma con la imagen de referencia para obtener una imagen con mayor dilación (por el efecto de la lente del microscopio). Finalmente, el resultado se vuelve a convertir a máscara binaria con píxeles de valor 255 en las células y ceros en el fondo, y por último, esta máscara binaria se multiplica por la especificación de histograma entre las imágenes originales. Al fondo de la imagen resultante se le asignan los valores promedio del fondo de la imagen de referencia (ver Figura 13). Mediante este proceso se consiguieron 300 imágenes de tres fuentes diferentes, exclusivamente para *Staphylococcus aureus*.



**Figura 13.** Ejemplo de la especificación de histograma con filtrado espacial. En la parte superior se muestra la imagen con ruido al aplicar la especificación de forma normal. En la parte inferior se muestra el resultado de hacer la especificación mediante las máscaras de segmentación.

Realizando estos procesos, se obtiene la base de datos final como se muestra en la Tabla 4.

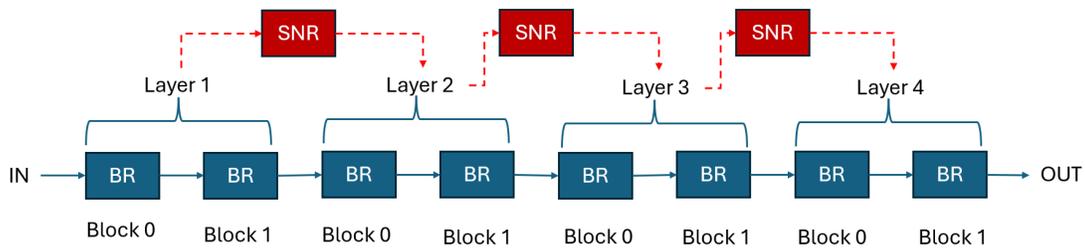
Género	Cantidad de imágenes (CI)	Aumento de Datos (AD)	CI + AD
Neisseria	1,242	No	1242
Escherichia	1,241	1,601	2,842
Lactobacillus*	600	No	600
Micrococcus	1,086	1,015	2,101
Veillonella	1,133	No	1,133
Staphylococcus**	1,361	907	2,268
Actinomyces	2,270	No	2,270
Clostridium	1,242	No	1,242
Bacteroides	1,162	No	1,162
Subtotal	11,337	3,523	-
<b>Total</b>	-	-	<b>14,860</b>

**Tabla 4.** Distribución por género de las imágenes de la base de datos final. \*Reportando sólo una especie de *Lactobacillus*. \*\*Sólo contando la especie *Staphylococcus aureus*.

### 3.3. Algoritmo de identificación de patrones

Las arquitecturas que utilizamos para la clasificación fueron ResNet 18 y 50. Como generalidades: Todas fueron programadas en *Python* [53] con la biblioteca *Pytorch* [54], se utiliza la función de pérdida de entropía cruzada y en la capa de salida se aplica una función softmax para obtener las probabilidades de pertenencia a cada clase. Los modelos se entrenaron durante 100 épocas, utilizando una partición del conjunto total de datos previamente dividido en subconjuntos que denominamos (a) “Entrenamiento” y (b) “Prueba” (detallados en las tablas correspondientes de cada modelo en la sección 3.4). En general, dentro del conjunto “Entrenamiento”, el 80% de los datos se destinó al entrenamiento de cada red y el 20% a su validación. Además, el conjunto de Prueba se mantuvo separado desde el inicio. Posteriormente, se realizaron pruebas de validación cruzada en algunos de los modelos.

La arquitectura más utilizada fue la de ResNet 18, de manera visual se explica más a detalle en la Figura 14.



**Figura 14.** Arquitectura de ResNet 18, BR significa Bloque Residual y es el presentado en la Figura 6, dos BR constituyen un bloque que llamamos *layer*. En los modelos en los que se utiliza el módulo SNR, estos se aplican entre cada bloque *layer* y están representados por los rectángulos rojos y las flechas punteadas. La capa de entrada (IN) es una capa convolucional con filtros de 7x7 y un paso (*stride*) de 2, mientras que la capa de salida (OUT) es una capa densa que recibe el resultado de aplicar *average pooling* a los resultados del bloque *layer* 4. Para ResNet 50 se modifican los bloques *layer* y BR como se especifica en la Tabla 1.

### 3.4. Aprendizaje y pruebas

Se entrenaron distintos modelos de clasificación con arquitectura ResNet, en general, la tasa de aprendizaje,  $\alpha$ , usada fue de 0.002, y lotes de tamaño 80, además de un optimizador Adam, esto es así para cada punto descrito a continuación a menos que se indique algo diferente.

Específicamente, se entrenaron modelos con 5 ideas principales:

1. Generar un “*baseline*” o modelo con el cuál compararnos. Entrenando arquitecturas ResNet 18 y 50 con las imágenes disponibles de tres especies y distintas fuentes, agregando normalización y restitución de dominio.

El *baseline* se refiere a un entrenamiento “clásico”, es decir, entrenando con las imágenes disponibles de las tres clases (especies) provenientes de distintas fuentes (Ver Tabla 5) sin ningún aumento de datos, manteniendo representación similar entre fuentes y entre clases, con el objetivo de tener un modelo con el cual comparar el ensamble propuesto. Cabe mencionar que en este caso el resultado real de la tinción de los datos de entrenamiento es positivo ( $G^+$ ) para dos especies y negativo ( $G^-$ ) para una.

Para generar un *baseline* se entrenaron cuatro modelos, cada uno con diferentes estrategias, con la intención de tener un punto de comparación y probar distintos métodos para mejorar el rendimiento de la red.

Las arquitecturas de los modelos propuestos son:

- A. La arquitectura ResNet 18.
- B. La arquitectura ResNet 18 con el módulo SNR, aplicando normalización de instancia (IN).
- C. La arquitectura ResNet 18 con el módulo SNR modificado, cambiando IN por normalización por lote (BN).
- D. La arquitectura ResNet 50.

Las clases que se utilizaron para entrenar estos modelos y su distribución son:

Especie	Cantidad de imágenes		Cantidad de fuentes diferentes
	Entrenamiento	Prueba	
Escherichia coli (Ec) $G^-$	122	66	8
Staphylococcus aureus (Sa) $G^+$	108	126	8
Micrococcus spp. (Ms) $G^+$	92	78	6

**Tabla 5.** Distribución de entrenamiento y prueba para el modelo *baseline*. Se especifica el tipo de tinción de Gram (G) como superíndice a un lado del nombre de cada uno.

## 2. Clasificar por resultado de la tinción de Gram (positiva o negativa)

Para la primera etapa del ensamble se entrenaron los modelos para clasificar dependiendo del color del que se tiñe la muestra, estas clases son Gram positivo (color violeta-azul) o Gram negativo (rosa-rojo). Para entrenar este modelo se utilizó el aumento de datos que representa la variabilidad de las fuentes, así como una variedad de especies disponibles. En este caso se entrenaron dos modelos, ResNet 18 y ResNet 50, sin ninguna modificación a su arquitectura original. Para darle mayor claridad a los resultados, las distribuciones de datos para entrenamiento y prueba se presentan en la sección 4.2.1.

Las especies utilizadas para este modelo son:

### Gram Positivo

- *Staphylococcus aureus*
- *Micrococcus spp.*
- *Cutibacterium acnes*
- *Bifidobacterium spp.*

### Gram Negativo

- *Escherichia coli*
- *Acinetobacter baumannii*
- *Veillonella*
- *Neisseria gonorrhoeae*

## 3. Clasificación por morfología (bacilo o coco)

Al igual que en el inciso anterior utilizamos aumento de datos, en este caso se entrenaron modelos con tres estrategias diferentes:

- I. Un modelo por tipo de tinción, utilizando imágenes a color de las bacterias. De manera explícita, esta estrategia resulta en dos modelos con dos clases cada uno (siendo estas clases coco o bacilo), en este caso dependerá del resultado del modelo de la etapa anterior para decidir si, en esta segunda etapa, se utiliza el modelo entrenado con la tinción positiva o el que fue con la negativa.
- II. Un modelo por tipo de tinción con imágenes en escala de grises. Similar al caso anterior, aquí también resultan dos modelos, cada uno con sus dos clases, pero todas las imágenes en escala de grises. De igual forma,

dependerá del tipo de resultado de la etapa uno para decidir qué modelo es el siguiente en el flujo del ensamble.

- III. Con imágenes a color y en escala de grises sin diferenciar entre los tipos de tinción, sólo separando por forma. De aquí resultan dos modelos independientes (a diferencia de los modelos en las estrategias I Y II, que se complementan); uno entrenado con imágenes a color y sin importar el tipo de tinción, y el otro, entrenado con imágenes en escala de grises, de igual manera sin considerar el tipo de tinción. Esta estrategia significa que en esta etapa sólo se utilizará un modelo sin importar el resultado de la etapa uno, con la única diferencia entre ellos en la representación de la entrada (a color o en grises).

Las imágenes que se utilizaron para entrenar estos modelos son heterogéneas, sin embargo, no se utilizan todas las especies disponibles para evitar que la red disminuya su rendimiento al enfocarse mayormente en una base de datos específica. En esta etapa se utilizó una tasa de aprendizaje de 0.0001.

#### 4. Clasificación por género

En este modelo se clasifican las bacterias por el tipo de organización celular que tienen, en este caso lo hacemos para dos especies de bacterias sin aplicar aumento de datos.

Se entrenan cuatro modelos, similares al punto uno de esta sección (*baseline*), con el propósito de comparar los resultados manteniendo las mismas estrategias, hiper-parámetros e instancias de prueba y entrenamiento. Con la única diferencia de que redujimos las clases a dos especies en lugar de tres; ya que para el ensamble de redes propuesto, al llegar a esta etapa ya se habrán filtrado aquellas especies que pertenezcan a un tipo Gram diferente, en este caso el *Staphylococcus aureus* y el *Micrococcus spp.* pertenecen al tipo Gram positivo, mientras que *Escherichia coli* es Gram negativo.

Especie	Cantidad de imágenes		Cantidad de fuentes diferentes
	Entrenamiento	Prueba	
<i>Staphylococcus aureus</i> (Sa) $G^+$	108	126	8
<i>Micrococcus spp.</i> (Ms) $G^+$	92	78	6

**Tabla 6.** Distribución de entrenamiento y prueba para el modelo que clasifica género dentro del ensamble.

#### 5. Clasificación por especie

Por último, realizamos una clasificación por especie entre imágenes de bacterias que pertenecen al mismo género. En este caso nos enfocamos en el género *Staphylococcus*, con tres especies diferentes: *aureus*, *epidermis* y *saprophyticus*.

Es importante mencionar que para los microbiólogos identificar entre especies del mismo género solo con las imágenes no es una tarea sencilla, ya que las imágenes suelen ser muy similares entre sí. Se entrenaron los mismos modelos que en el punto uno, con la distribución mostrada en la Tabla 7. Aquí se agregan dos especies no consideradas en la Tabla 4, con mil instancias por cada una. No se utiliza aumento de datos y se acota a una sola base de datos (DiBaS).

Especie	Cantidad de imágenes	
	Entrenamiento	Prueba
Staphylococcus aureus (Sa)	501	499
Staphylococcus epidermis (Se)	526	524
Staphylococcus saprophyticus (Ss)	534	524

**Tabla 7.** Distribución de entrenamiento y prueba para el modelo que clasifica algunas especies dentro del género *Staphylococcus*.

## 4. Resultados

### 4.1. Baseline

Recordando que se entrenaron los siguientes modelos:

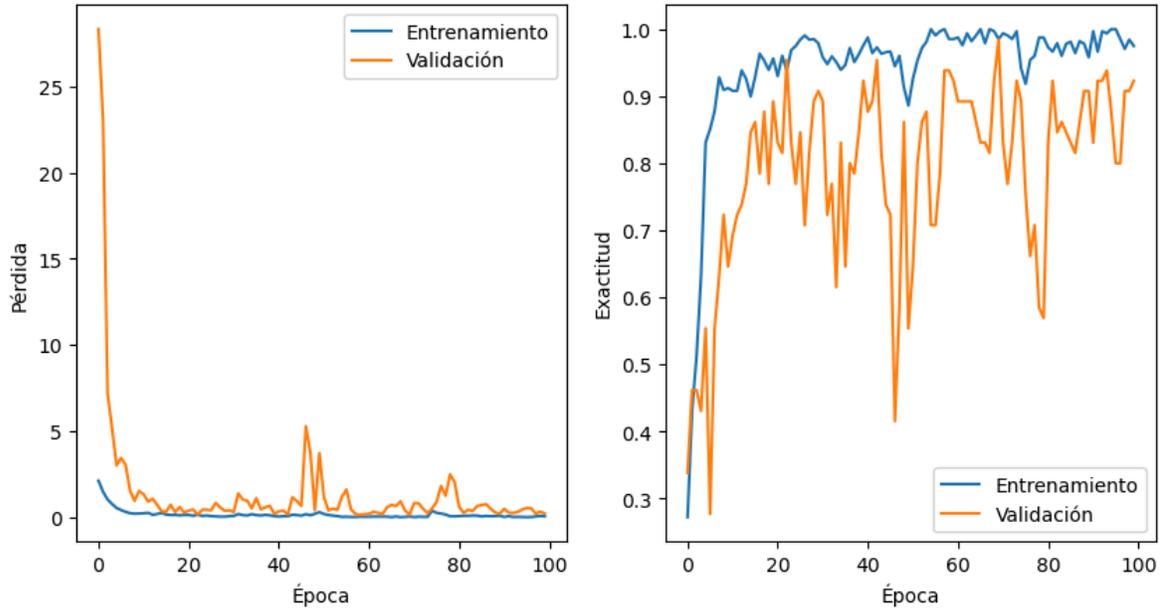
- A. Solo la arquitectura ResNet 18.
- B. ResNet 18 con módulo SNR aplicando normalización de instancia (IN).
- C. ResNet 18 modificando SNR, cambiando IN por normalización por lote (BN).
- D. ResNet 50.

Los resultados por modelo se presentan en la siguiente tabla:

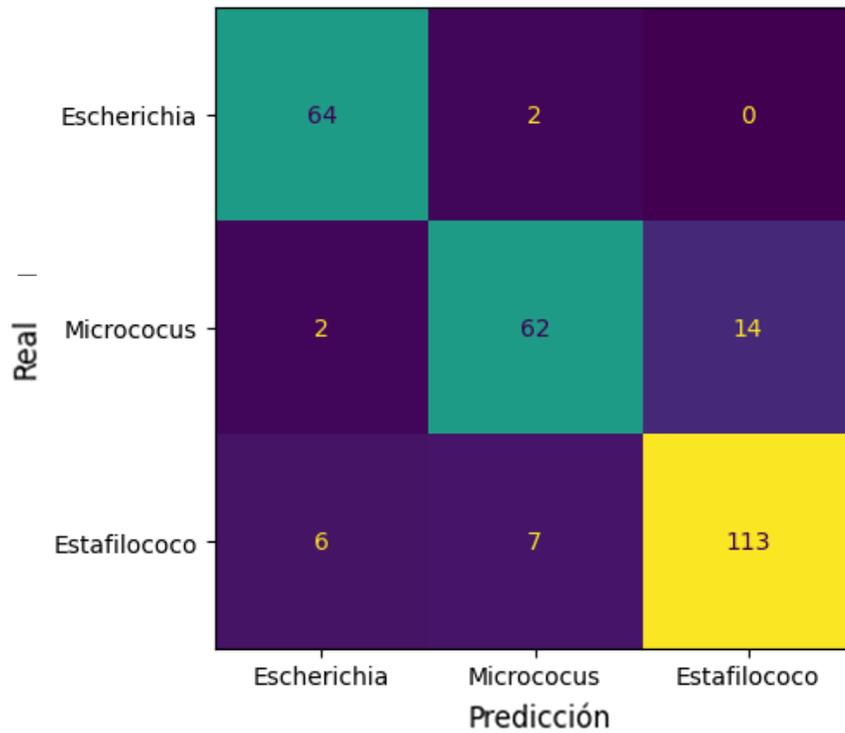
Modelo	Exactitud	Exactitud por clase
A	47 %	Ec: 27.3 %, Sa: 10.3 % Ms: 82.5 %
<b>B</b>	<b>88.5 %</b>	<b>Ec: 96.96 %, Sa: 92.1 % Ms: 79.48 %</b>
C	77 %	Ec: 66.7 %, Sa: 100 % Ms: 70.6 %
D	70 %	Ec: 98.5 %, Sa: 71.8 % Ms: 55.6 %
Promedio	70.62 %	Ec: 72.36 %, Sa: 68.55 % Ms: 72 %

**Tabla 8.** Resultados de exactitud por modelo y por clase para *baseline*. En negritas se muestra el modelo que obtuvo el mejor resultado.

A partir de los resultados presentados en la Tabla 8 se tomó el mejor modelo, este es el ResNet 18 con módulo SNR-IN, en la Figura 15 se presentan las gráficas de exactitud y pérdida en entrenamiento/validación, además de la matriz de confusión obtenida al hacer predicción en el conjunto de prueba.



**Figura 15.** Gráfica de pérdida (del lado izquierdo) y gráfica de exactitud (lado derecho) para los conjuntos de entrenamiento y validación del modelo *baseline* ResNet 18 SNR-IN.



**Figura 16.** Matriz de confusión de datos de prueba para el modelo *baseline* ResNet 18 SNR-IN

Después se hicieron experimentos de validación cruzada de 10, 5 y 2 grupos con los datos de entrenamiento, los resultados son los siguientes:

Modelos	Validación cruzada (Exactitud promedio)		
	10 iteraciones	5 iteraciones	2 iteraciones
ResNet 18 SNR-IN	89.99 % ± 5.84%	94.83 % ± 3.66%	90.84 % ± 8.71%
ResNet 18	87.26 % ± 18.08%	84.86 % ± 11.45%	83.16 % ± 13.52%

**Tabla 9.** Resultados de validación cruzada de 10, 5 y 2 iteraciones para el modelo *baseline* ResNet 18 SNR-IN. Se muestra el promedio de las iteraciones con su desviación estándar. Como medio de comparación también se muestran los resultados de validación cruzada de la arquitectura ResNet 18 sin ninguna modificación.

Como análisis relevante de estos resultados podemos enfatizar tres cosas. La primera, que el entrenamiento es ruidoso al momento de calcular la exactitud por época, con fluctuaciones abruptas y cambios repentinos en los valores. La segunda, de acuerdo con la matriz de confusión, las clases con mayor tasa de error son estafilococos y micrococos, siendo ambos cocos Gram positivos y que solo varían en la agrupación de las células. La tercera, los modelos varían mucho en su exactitud al momento de hacer validación cruzada de varias iteraciones, esta variación se reduce cuando se trata de arquitecturas que utilizan el módulo de SNR en comparación con las que no.

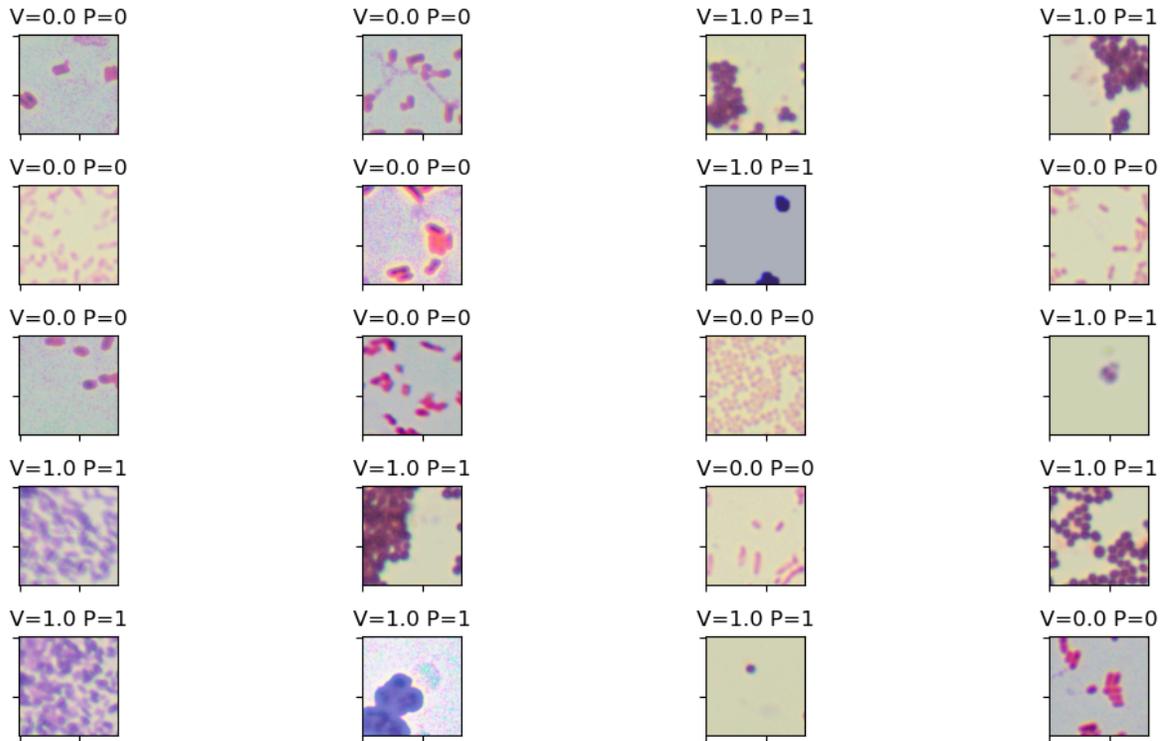
## 4.2. Ensamble de redes

A continuación se presentan los resultados por modelo que en conjunto conforman el ensamble en cascada propuesto.

### 4.2.1. Por tinción

Para el primer modelo del ensamble, que clasifica por resultado de tinción, se utilizaron imágenes provenientes de varias fuentes y con aumento de datos.

La Tabla 10 muestra que la arquitectura ResNet 50 fue el mejor modelo con apenas cuatro errores de 4,059 instancias de prueba, mientras que ResNet 18 tuvo 14 errores con la misma cantidad de instancias. En la Figura 17 se muestran ejemplos de las predicciones del mejor modelo en prueba, ResNet 50.



**Figura 17.** Ejemplo de predicciones del modelo ResNet 50 en el conjunto de prueba, también se muestra la heterogeneidad de las muestras a clasificar. Las etiquetas 0 indican Gram negativo, las etiquetas 1 Gram positivo. V significa Verdadero, P Predicción.

Tipo Gram (Clases)	Cantidad de imágenes		Especies / Fuentes	Resultados de exactitud en prueba	
	Entrenamiento	Prueba		ResNet 18	<b>ResNet 50</b>
Gram Negativo	2,383	2,200	4/8	99.7 %	<b>99.9 %</b>
Gram Positivo	2,238	2,059	4/8	99.9 %	<b>99.9 %</b>

**Tabla 10.** Resultados del modelo para clasificar la respuesta a la tinción. Se especifica la cantidad y distribución de imágenes de cada clase para el entrenamiento y la prueba, además se describen cuántas especies y cuántas fuentes se utilizaron. En negritas se resalta el mejor modelo.

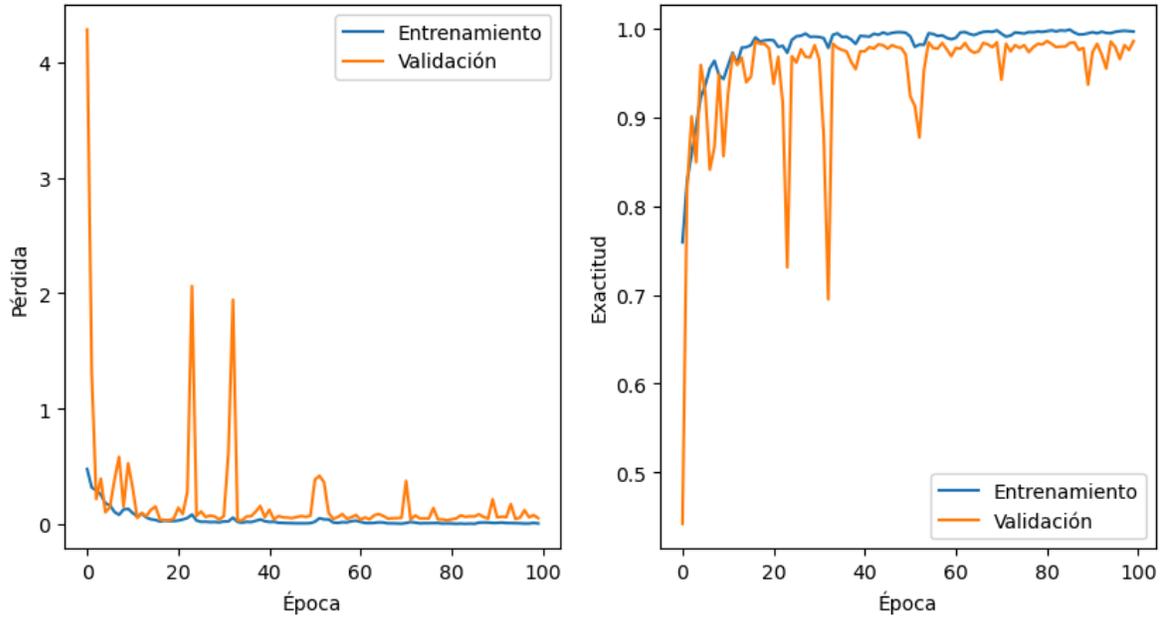
### 4.2.2. Por morfología

Para este modelo se utilizaron tres estrategias de entrenamiento para poder clasificar la presencia de cocos o bacilos en la imagen, en total se entrenaron 6 modelos por arquitectura (ResNet 18 y ResNet 50).

Tipo de imagen de entrenamiento (por Gram)	Cantidad de imágenes		Especies / Fuentes	Resultados de exactitud en prueba	
	Entrenamiento	Prueba		ResNet 18	ResNet 50
Gram Negativo Color	2,937	2,070	5 / 6	69 %	86 %
Gram Positivo Color	2,963	2,294	6 / 5	98 %	99 %
Gram Negativo Grises	2,937	2,070	5 / 6	49 %	83 %
Gram Positivo Grises	2,963	2,294	6 / 5	97 %	97 %
<b>Ambos Color (<math>G^+</math> y <math>G^-</math>)</b>	<b>6,225</b>	<b>4,365</b>	<b>10 / 11</b>	<b>97 %</b>	<b>97 %</b>
Ambos Grises ( $G^+$ y $G^-$ )	6,225	4,365	10 / 11	93 %	93 %

**Tabla 11.** Resultados de la prueba de clasificación por morfología. En distintos colores se resaltan las estrategias que se probaron para seleccionar el modelo más óptimo para esta etapa. En negritas se destaca el mejor modelo.

En la Tabla 11, se muestra que los mejores resultados se obtuvieron con la estrategia de entrenar el modelo sin separar por tipo Gram y hacerlo con las imágenes a color. En comparación entre arquitecturas, tanto ResNet 18 como 50 obtuvieron el mismo desempeño. En la Figura 18 se presentan los resultados de entrenamiento, validación y prueba del modelo ResNet 50.



**Figura 18.** Gráfica de pérdida (del lado izquierdo) y gráfica de exactitud (lado derecho) para los conjuntos de entrenamiento y validación del modelo para clasificar morfología con arquitectura ResNet 50 y la estrategia de utilizar imágenes de ambos tipos de tinción a color.

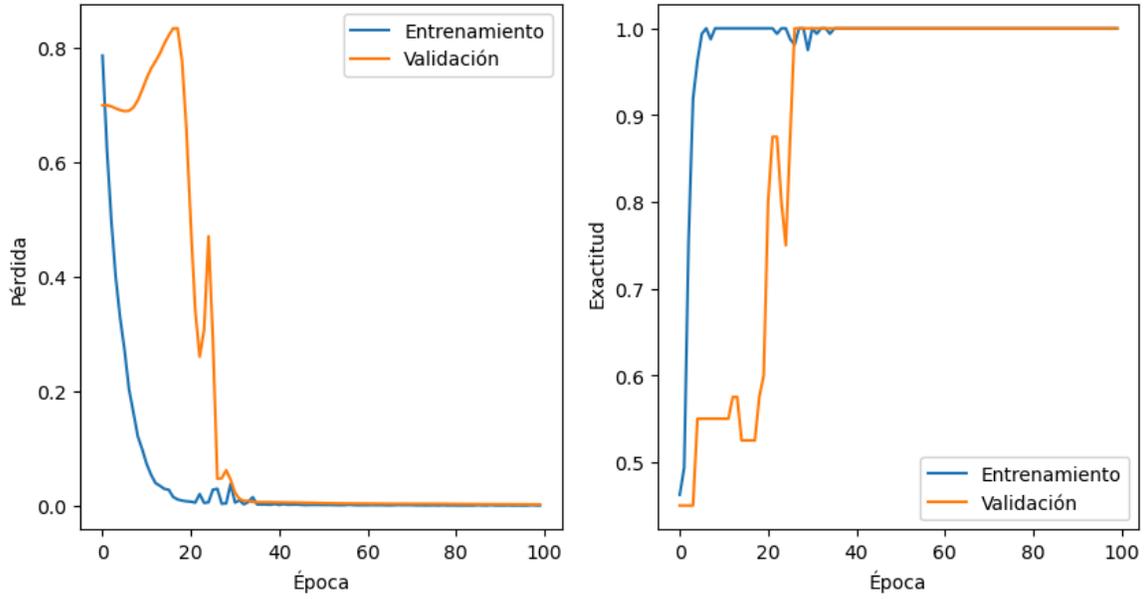
### 4.2.3. Por género

El modelo de clasificación por género fue entrenado utilizando las mismas estrategias y condiciones que el presentado en *baseline*, pero únicamente con las clases restantes al seguir el flujo del ensamble. Los resultados obtenidos por arquitectura son:

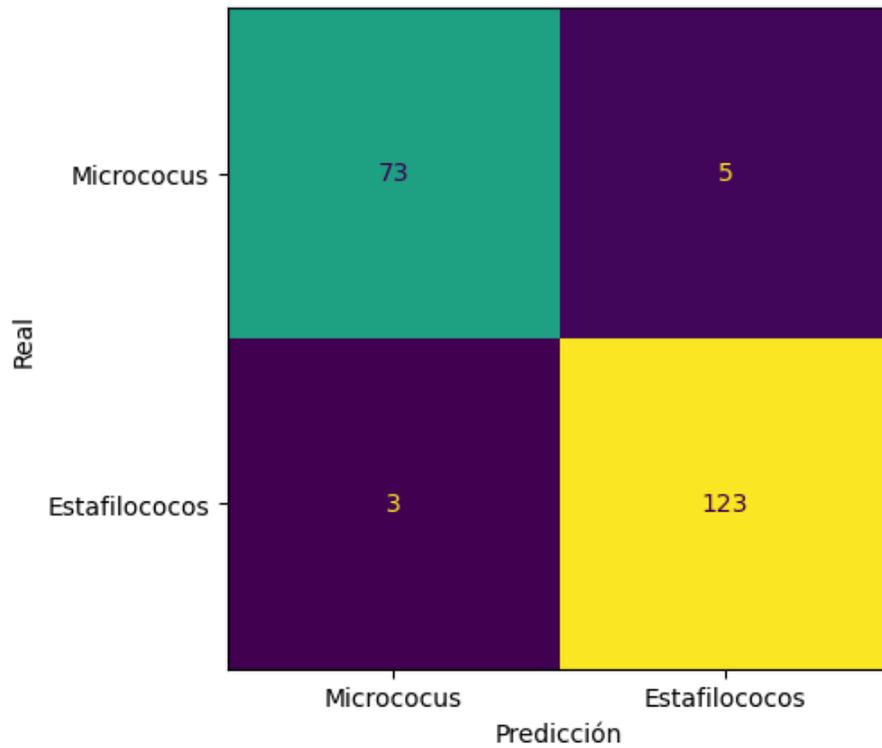
Modelo	Exactitud	Exactitud por clase
A	96 %	Sa: 97.6 % Ms: 93.6 %
B	94 %	Sa: 97.6 % Ms: 88.5 %
<b>C</b>	<b>96.08 %</b>	<b>Sa: 97.62 % Ms: 93.6 %</b>
D	92 %	Sa: 84.6 % Ms: 96.8 %
Promedio	95 %	Sa: 94 % Ms: 94.72 %

**Tabla 12.** Resultados de exactitud por modelo y por clase para el modelo de género. En negritas se destaca el mejor modelo.

El mejor desempeño lo presenta la arquitectura ResNet 18 con módulo SNR-BN, obteniendo un 97% de exactitud. En la Figura 19 se muestra el resultado de las gráficas de entrenamiento por época y en la Figura 20 la matriz de confusión, utilizando la misma partición de imágenes (de entrenamiento y prueba) que en el modelo *baseline*.



**Figura 19.** Gráfica de pérdida (del lado izquierdo) y gráfica de exactitud (lado derecho) para conjuntos de entrenamiento y validación del modelo del ensamble por género ResNet 18 SNR-BN.



**Figura 20.** Matriz de confusión de datos de prueba para el modelo del ensamble por género ResNet 18 SNR-BN.

De igual manera que en el punto 9.1 de los resultados (el modelo *baseline*), se realizaron experimentos de validación cruzada de 10, 5 y 2 vías con los datos de entrenamiento (ver Tabla 13).

Modelos	Validación cruzada (Exactitud promedio)		
	10 iteraciones	5 iteraciones	2 iteraciones
ResNet 18 SNR-BN	98.48 % ± 1.30%	97.97 % ± 1.69%	97.97 % ± 0.007%
ResNet 18	98.48 % ± 1.75%	98.48 % ± 1.05%	97.47 % ± 2.85%

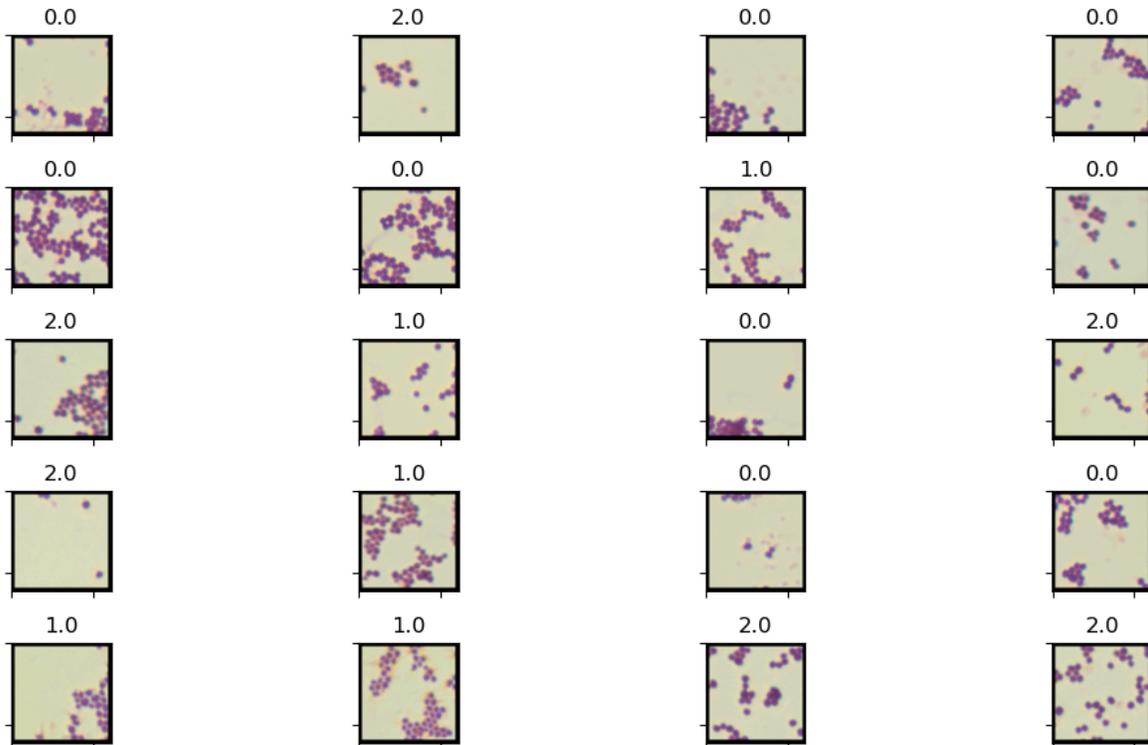
**Tabla 13.** Resultados de validación cruzada de 10, 5 y 2 iteraciones para el modelo de la etapa 3 del ensamble, clasificación por género, con arquitectura ResNet 18 SNR-BN. Se muestra el promedio de las iteraciones con su desviación estándar. Nuevamente, como medio comparativo también se muestran los resultados de validación cruzada de la arquitectura ResNet 18 sin ninguna modificación.

De estos resultados podemos denotar que:

- Todos los modelos alcanzan tasas de exactitud mayor al 90%.
- Las gráficas de entrenamiento por época tienen transiciones sin tantas fluctuaciones.
- La desviación estándar en los modelos de validación cruzada es baja, esto es verdad tanto para los que aplican módulo SNR como los que no.

#### 4.2.4. Por especie

Para mostrar la funcionalidad total del ensamble se entrenó el último modelo con la intención de poder clasificar imágenes del mismo género pero de diferente especie. En la Figura 21 se muestran ejemplos de las tres especies del género *Staphylococcus* que se utilizaron para este entrenamiento.



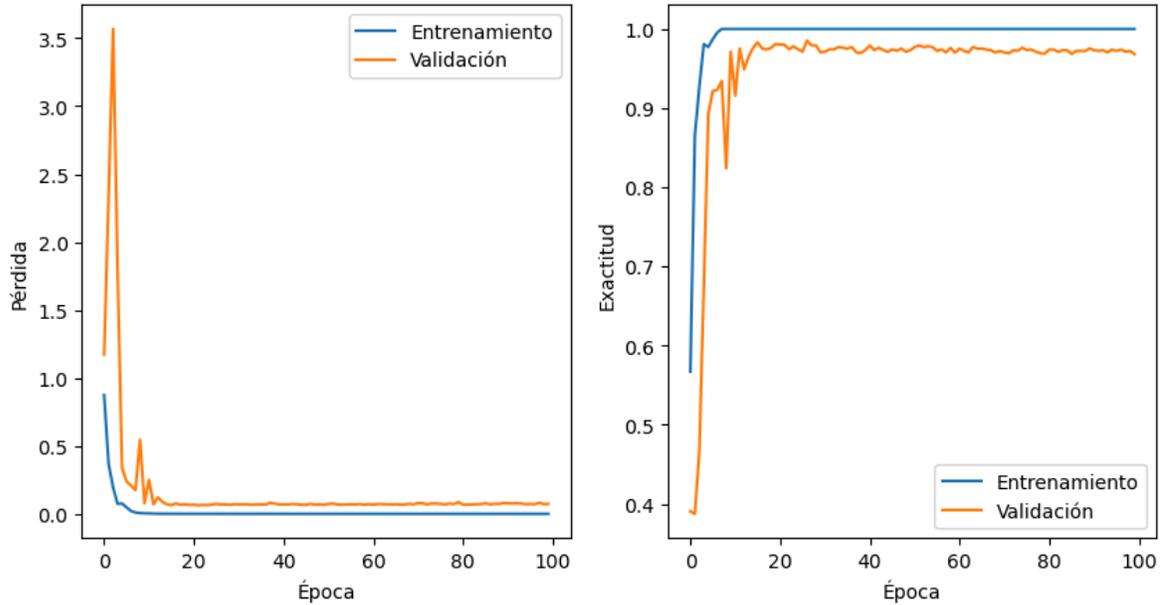
**Figura 21.** Ejemplos de instancias para el último modelo del ensamble (clasificación por especie); son imágenes representativas de tres especies del género *Staphylococcus*, el significado de las etiquetas se presenta en el siguiente orden: 0-*aureus*, 1-*epidermis*, 2-*saprophyticus*. Todas estas imágenes se obtuvieron de la base de datos DiBaS por lo que visualmente son muy similares entre sí.

Los resultados por modelo:

Modelo	Exactitud	Exactitud por clase
A	91 %	Se: 84.7 %, Sa: 89.2 % Ss: 99.6 %
<b>B</b>	<b>92 %</b>	<b>Se: 92 %, Sa: 83.8 % Ss: 99 %</b>
C	91 %	Se: 91.6 %, Sa: 82.4 % Ss: 99.4 %
D	87 %	Se: 85.7 %, Sa: 79.4 % Ss: 95.8 %
Promedio	90.25 %	Se: 88.5 %, Sa: 83.7 % Ss: 98.45 %

**Tabla 14.** Resultados de exactitud por modelo y por clase para el modelo de especie. En negritas se destaca el modelo con mejores resultados.

Gráficas de entrenamiento y validación del mejor modelo:



**Figura 22.** Gráfica de pérdida (del lado izquierdo) y gráfica de exactitud (lado derecho) para conjuntos de entrenamiento y validación del modelo del ensamble por especie ResNet 18 SNR-IN.

Los resultados de validación cruzada para el mejor modelo se muestran en la tabla 15.

Modelo	Validación cruzada (Exactitud promedio)		
	10 iteraciones	5 iteraciones	2 iteraciones
ResNet 18 SNR-IN	94.84 % ± 0.9%	94.21 % ± 0.76%	89.95 % ± 1.52%

**Tabla 15.** Resultados de validación cruzada de 10, 5 y 2 iteraciones para el modelo por especie ResNet 18 SNR-IN. Se muestra el promedio de las iteraciones con su desviación estándar.

#### 4.2.5. Unión del ensamble

Por último, realizamos una unión de los mejores modelos que fueron encontrados en los resultados por etapa y utilizamos el mismo conjunto de prueba que usamos en el modelo de *baseline* (explicado en la Tabla 5 de la sección 3.4). Esta unión se hace en forma de cascada, donde se toma el resultado del primer modelo para decidir el siguiente a cargar.

En este caso, presentamos la matriz de confusión total del ensamble de redes (ver Figura 23), clasificando los géneros de *Micrococcus* y *Staphylococcus*, así como agregando una clase llamada Otros para aquellas instancias cuyo resultado no pertenezca a alguno de los dos géneros mencionados. En esta ocasión, las imágenes de la bacteria *Escherichia coli* serán categorizadas en esta última clase. Sin embargo, es importante reconocer que la combinación de los resultados previos de cada modelo son relevantes, por lo que se agregan las matrices de confusión de los modelos de tinción y de morfología con el mismo conjunto de prueba (Figura 24 y 25).

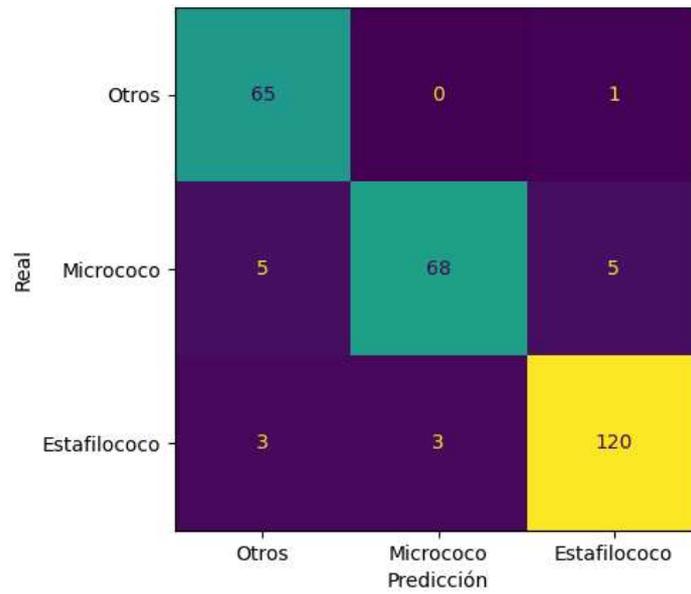


Figura 23. Matriz de confusión de los datos de prueba para todo el ensamble en conjunto.

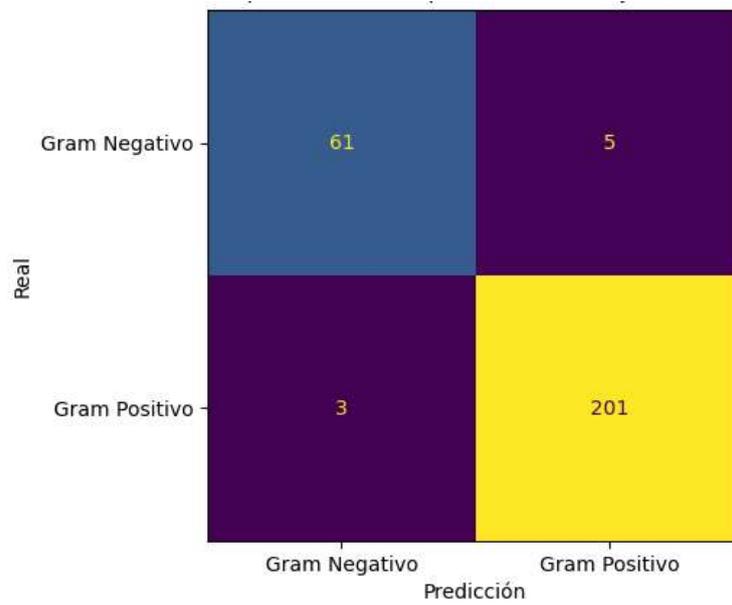
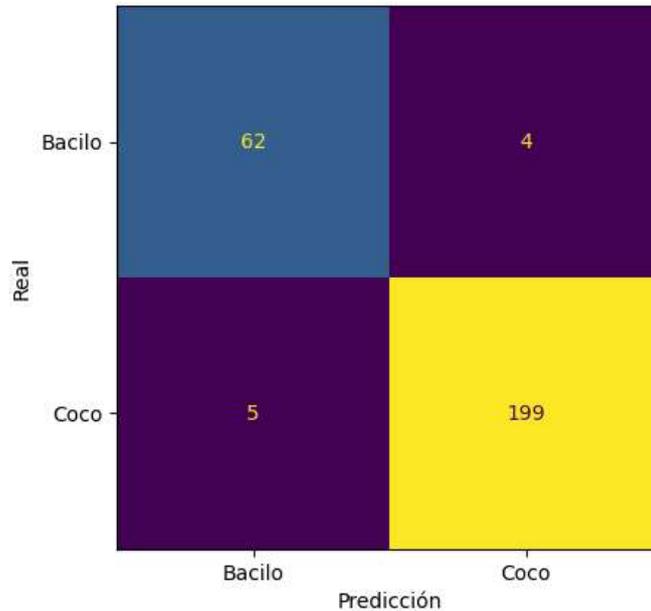


Figura 24. Matriz de confusión de datos de prueba para el primer modelo del ensamble (clasificación por tinción).



**Figura 25.** Matriz de confusión de datos de prueba para el segundo modelo del ensamble (clasificación por morfología).

Los resultados muestran que al clasificar los géneros de bacterias que tenemos en nuestro conjunto de prueba obtenemos una exactitud del 94% de 270 muestras, mostrando una mejora considerable contra el *baseline*, donde un sólo modelo fue entrenado para la clasificación y obtuvo una exactitud del 88.5%. Es importante notar que en este caso, sustituimos la clase de *Escherichia coli* por otra denominada Otros, dado que no tenemos modelos que lleguen a clasificar hasta especies de *Escherichia*, además existen otras combinaciones de las cuáles no tenemos modelo para continuar clasificando.

Por otro lado, al enfocarnos en la rama para clasificar géneros de cocos Gram positivos (la que tenemos más completa) obtenemos muy buenos resultados, sin embargo, debemos mencionar que existen varios fallos en la clasificación tanto de morfología (ver Figura 25) como de tinción (ver Figura 26), lo que nos llevaría a que si tuviéramos los modelos pertinentes no podríamos clasificar correctamente una cantidad considerable de bacterias de *Escherichia coli* (que son bacilos Gram negativos). Esto es explicable por la calidad de las imágenes aumentadas para bacterias bacilos que se utilizaron al entrenar el modelo de morfología, ya que no se les dió el mismo tratamiento que aquellas a las que se aplicó filtrado espacial al momento de hacer especificación de histograma, como en las bacterias de *Staphylococcus*, resultando en pérdida de información morfológica de los bacilos después de la especificación. Por lo que para mejorar este modelo resulta importante mejorar estas imágenes.

## 5. Conclusiones

### 5.1. Baseline

El mejor modelo encontrado en esta sección es aquel con la adición del módulo de normalización y restitución de estilo con normalización de instancia. Al compararlo con los resultados de la arquitectura sin aplicarlo, se puede apreciar que mejora la generalización de las redes y, estadísticamente como se muestra en la Tabla 9, reduce la variabilidad al entrenar muchos modelos mediante validación cruzada. Por lo que se concluye que aplicar esta clase de estrategias puede resultar de ayuda para mejorar los resultados cuando se tienen diversas fuentes y pocas instancias de cada una, sin embargo, no resuelve el problema de poder generalizar completamente a bases de datos que el modelo no conoce o de las que ha visto pocas instancias.

En la Figura 15 observamos que los resultados de exactitud para el conjunto de validación muestran que es un entrenamiento con mucho ruido, al tener fluctuaciones abruptas en los valores de exactitud y de validación entre épocas. Esto era esperado por la variación entre las imágenes, y la razón por la que los modelos no pueden clasificar correctamente a datos desconocidos cuando la cantidad de imágenes de entrenamiento es limitada.

A partir de la Figura 16 podemos concluir que al entrenar con imágenes de varias especies y de distintas fuentes, las clases entre las que ocurre mayor confusión en el modelo son aquellas cuyas características morfológicas son más similares entre sí.

### 5.2. Ensamble

Al entrenar con datos aumentados por especificación de histograma podemos alcanzar una generalización del problema en las redes convolucionales, es importante recalcar que no siempre se aplicó el aumento de datos, se decidió no hacerlo entre especies diferentes porque los datos podrían perder detalles o características reales relevantes, tampoco se utilizó para entrenar los modelos de todas las etapas, específicamente no se usó para género ni especie, porque se prefirió conservar las distribuciones de los datos reales para mantener detalles inherentes y así obtener resultados más confiables sobre el funcionamiento del ensamble.

La tarea de clasificar las imágenes por tipo de tinción resulta sencilla, y es uno de los beneficios de utilizar aprendizaje profundo. Ya que hacerlo con métodos más tradicionales, como el procesamiento de imágenes u otros algoritmos de aprendizaje automático, puede ser mucho más complicado, esto se debe a la sensibilidad para detectar cambios de color, especialmente si se aplica un umbral para determinar esos cambios o si se utilizan distintos espacios de color.

En cuanto a la tarea de clasificar morfología, el entrenamiento y el resultado se complica más utilizando imágenes en escala de grises que con imágenes a color. Debido

a que en las imágenes a escala de grises hay un ruido presente que no puede ser diferenciado de las células. Los resultados óptimos se obtuvieron por medio de las imágenes a color y sin distinguir entre los tipos Gram, sólo diferenciando entre las morfologías. Esto sugiere que las tareas de clasificar por tinción y por morfología pueden ser tareas complementarias, lo que indicaría que estos dos modelos podrían combinarse y sustituirse por uno solo.

Ahora, tomando los resultados del modelo *baseline* y comparándolos con los resultados del modelo de clasificación por género, podemos notar grandes diferencias. Empezando por la exactitud resultante de las distintas estrategias empleadas, en la etapa de género todos los modelos resultaron tener una exactitud mayor al 90%, mientras que en nuestro modelo *baseline* sólo uno pudo lograrlo, esto muestra que nuestra propuesta de entrenar redes especializadas en cada una de las etapas taxonómicas funciona e incrementa la exactitud de los resultados, reduciendo la confusión entre clases similares al hacer énfasis en detalles más sutiles. Este comportamiento también se muestra en las gráficas de entrenamiento y validación donde es notable que el entrenamiento se simplifica al comparar las gráficas con crecimientos más suaves del modelo de género con las fluctuaciones del modelo base. Otro aspecto a recalcar es la diferencia entre los resultados de desviación estándar en la validación cruzada, para los mejores modelos de *baseline* y género, estos nos permiten comprobar la capacidad para replicar los resultados independientemente de los pesos iniciales o de la elección en los datos de entrenamiento y de prueba.

Por último, en la sección 4.2.4 mostramos que es posible identificar especies dentro de un mismo género sólo por la imagen. Aunque estos resultados no son perfectos, sí se obtienen clasificaciones cercanos al 90% de exactitud de manera consistente como se observa en la validación cruzada de la Tabla 15.

### 5.3. Globales

El uso del enfoque de ensamble nos ofrece mayor flexibilidad para reentrenar o modificar modelos en cualquier etapa que sea necesaria. Esto incluye la posibilidad de agregar características multimodales cuando sea necesario, lo que podría mejorar la clasificación de tal forma que no solo dependa de las imágenes. Además, el ensamble permite aproximarnos a la resolución del problema del gran espacio de búsqueda que representa la cantidad de bacterias que existen en el mundo, de una manera más simple e intuitiva.

Un beneficio adicional es que se requiere menor optimización de hiper-parámetros al entrenar modelos seccionando por características, esto es representado por las diferencias en los resultados de las Figuras 15 y 19 de los modelos *baseline* y género. Además, los resultados nos indican que para esta tarea es más sencillo entrenar modelos con clases cuyas características sean más cercanas entre sí.

Algunas desventajas o limitaciones identificadas en esta propuesta son:

- Variabilidad de la morfología: Dependiendo del tipo de bacteria la metodología puede no ser efectiva ya que algunas especies bacterianas pueden cambiar su morfología bajo diferentes condiciones ambientales, lo que se conoce como pleomorfismo [59]. Dado que nuestra aproximación se basa principalmente en la identificación de características morfológicas específicas, el pleomorfismo puede limitar la eficacia del ensamble en ciertas circunstancias.
- Limitaciones de la tinción de Gram: Existen bacterias que no se tiñen con el método de tinción de Gram debido a que no poseen pared celular que se pueda teñir, por ejemplo especies como *Mycoplasmas*, o especies que se tiñen de manera muy tenue como *Legionella pneumophila*, *Treponema pallidum*, *Leptospira spp.* y *Bartonella spp.* [60].

## 5.4. Perspectivas

De este trabajo se desprenden varias cuestiones explorables a futuro. Una de ellas es la pregunta natural que surge de analizar la idea del ensamble de redes neuronales: ¿Cuánto espacio de memoria podrían ocupar estos modelos? En nuestro caso, los cuatro modelos principales elegidos como mejores por cada etapa ocupan un espacio en memoria de 209.4 MB, sin embargo, esto depende de la cantidad de modelos para clasificar géneros y especies que se quieran realizar, por supuesto primero es necesario obtener los datos, pero de manera general cada modelo ResNet 18 ocupa 42 MB, mientras que los modelos de ResNet 50 ocupan 80 MB. Por lo tanto, una de las primeras cuestiones explorables a futuro es la posibilidad de probar con distintas arquitecturas como *inceptionv3* o *VGG*, en búsqueda de modelos que sean lo más livianos posible, y así eficientizar el uso de memoria necesario.

Por otro lado, al aplicar aumento de datos mediante especificación de histograma se obtuvieron buenos resultados, sin embargo, en varios casos se tuvieron que modificar los resultados mediante máscaras de segmentación que nos permitieran el reducir el ruido que se genera a partir de la especificación y así evitar que la red neuronal confunda las características que queremos que aprenda. Por lo tanto, una idea que se puede seguir explorando, es la de aplicar especificación de histograma en diferentes espacios de color, como por ejemplo HSV, cuya aplicación posiblemente pueda simplificar la aplicación de la especificación entre los histogramas.

Finalmente, el poder distinguir entre especies del mismo género mediante imágenes puede ser un reto o incluso algo no posible para los especialistas debido a la gran similitud entre ellas, a pesar de esto, las redes neuronales entrenadas demuestran que es posible obtener altas tasas de clasificación solamente guiándose por las imágenes. Entonces, una cuestión interesante por analizar es encontrar a qué características le dan importancia las redes dentro de la tarea para determinar la especie, esto puede hacerse a través de métodos como Gradcam o similares que nos permitan encontrar patrones para aprender y comprender un poco más sobre cómo la red toma estas decisiones.



## 6. Bibliografía

- [1] R. Murray, S. Rosental, A. Pfaller, "Microbiología Médica". E.U.A. Elsevier, 2017.
- [2] G. Bou, A. Fernández Olmos, C. García, J. A. Sáez Nieto, S. Valdezate, "Métodos de identificación bacteriana en el laboratorio de microbiología", Enfermedades Infecciosas y Microbiología Clínica, ELSEVIER. España. 2011.
- [3] N. Mohamad, N. Jusoh, Z. Zaw Htike, S. Lei Win, "Bacteria Identification From Microscopic Morphology: A Survey". International Journal on Soft Computing, Artificial Intelligence and Applications, 3(2), pp.1-12. 2014
- [4] E.P. Sánchez Femat, "Aplicación móvil para el conteo automático e identificación preliminar de colonias de bacterias mediante reconocimiento de patrones". Tesis para Ingeniería en Sistemas Computacionales. IPN, CDMX, México, 2018
- [5] P. Ruusuvoori, J. Seppälä, T. Erkkilä, A. Lehmuusola, J.A. Puhakka, O. Yli-Harja. "Efficient automated method for image-based classification of microbial cells". Department of Signal Processing and Department of Chemistry and Bioengineering Tampere University of Technology. Tampere, Finlandia. 2009.
- [6] P. Hiremath, P. Bannigidad, "Automatic classification of bacterial cells in digital microscopic images", International Journal of Engineering and Technology, India, 2009.
- [7] T. Treebupachatsakul, S. Poomrittigul, "Bacteria Classification using Image Processing and Deep learning" en el 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), JeJu, Corea del Sur. 2019.
- [8] T. Treebupachatsakul, S. Poomrittigul, "Microorganism Image Recognition based on Deep Learning Application", 2020 International Conference on Electronics, Information, and Communication (ICEIC). Barcelona, España. 2020.
- [9] K. Smith, A. Kang, J. Kirby, "Automated Interpretation of Blood Culture Gram Stains by Use of a Deep Convolutional Neural Network". Journal of Clinical Microbiology, 56(3). 2017
- [10] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzychczy-Włoch, D. Ochońska, "Deep learning approach to bacterial colony classification". *PLOS ONE*. 2017, 12(9), p.e0184554.
- [11] M. F. Wahid, T. Ahmed and M. A. Habib, "Classification of Microscopic Images of Bacteria Using Deep Convolutional Neural Network" 2018 10th International Conference on Electrical and Computer Engineering (ICECE). Dhaka, Bangladesh. 2018
- [12] C. Spahn, E. Gómez-de-Mariscal, R.F. Laine, et al, "DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches", *Commun Biol* 5 688, 2022.
- [13] "Introduction to Colab". Google Collaboratory. <https://colab.research.google.com/notebooks/intro.ipynb>.

- [14] L. Bush, "Infecciones por Staphylococcus aureus - Infecciones", Manual MSD versión para público general, 2020. [Online]. Disponible en: <https://www.msdmanuals.com/es/hogar/infecciones/infecciones-bacterianas-bacterias-grampositivas/infecciones-por-staphylococcus-aureus>.
- [15] H.D. Isenberg (Editor), "Clinical Microbiology Procedures Handbook", American Society for Microbiology. Washington DC: ASM Press; 2004.
- [16] Who.int. 2020. "Resistencia a los antibióticos." [online] Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/resistencia-a-los-antibi%C3%B3ticos>
- [17] M. Gobernado, J. L. López-Hontangas, "Identificación Bacteriana", Enfermedades Infecciosas y Microbiología Clínica, vol. 21, págs. 54-60, ELSEVIER, 2003
- [18] Ruiz M., LadyofHats - Trabajo propio utilizando: Adobe Illustrator, Bacilos: Bacterial morphology diagram, Wikipedia. [online] Disponible en: [https://es.wikipedia.org/wiki/Bacilo#/media/Archivo:Bacterial\\_morphology\\_diagram.svg](https://es.wikipedia.org/wiki/Bacilo#/media/Archivo:Bacterial_morphology_diagram.svg)
- [19] CanalSALUD. 2020. "¿En Qué Consiste La Tinción De Gram?" -*Canalsalud*. [online] Disponible en: <https://www.salud.mapfre.es/pruebas-diagnosticas/otras-pruebas-diagnosticas/tincion-de-gram/>
- [20] Bergey's Manual of Determinative Bacteriology, Identification flow charts. [Online] Disponible en: <https://static1.squarespace.com/static/54c7f25fe4b0447c7f8aed4b/t/5504c972e4b06ae7c7462e69/1426377074561/IDFlowcharts.pdf>
- [21] J. M. Llamas, "Reconocimiento de imágenes mediante redes neuronales convolucionales". Tesis para Ingeniería de software. U.P.M., Madrid, España.
- [22] R. Szeliski, "Computer vision, algorithms and applications". Londres, UK. Springer, 2011.
- [23] J. Igual (2015, enero). "Todo sobre el histograma en fotografía". [Online]. Disponible en: <http://fotoigual.com/tutoriales/todo-sobre-el-histograma-en-fotografia/>
- [24] W. Burger, M. J. Burge, "Digital Image Processing", 2da. Ed. Londres, UK: Springer, 2018.
- [25] R. C. González, P. Woods, "Digital Image Processing", Addison Wesley, 2002.
- [26] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database". IEEE conference on computer vision and pattern recognition (pp. 248–255). 2009
- [28] <https://huggingface.co/Bacteria-Classification/qile0317>, <https://huggingface.co/spaces/qile0317/Bacteria-Classification>.

- [29] TN University, "Diccionario sobre inteligencia artificial: 100 conceptos claves sobre sistemas inteligentes", No. 2, TN Editorial. Marzo de 2024.
- [30] X. Jin, C. Lan, W. Zeng and Z. Chen, "Style Normalization and Restitution for Domain Generalization and Adaptation" in IEEE Transactions on Multimedia, vol. 24, pp. 3636-3651, 2022, doi: 10.1109/TMM.2021.3104379.
- [31] Synced, SyncedReview [Online] Disponible en: <https://medium.com/syncedreview/facebook-ai-proposes-group-normalization-alternative-to-batch-normalization-fb0699bffa7>
- [32] Think Big. 2020. "Machine Learning A Tu Alcance: La Matriz De Confusión - Think Big Empresas." [online] Disponible en: <https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/>
- [33] J. Brownlee (2020, agosto). "Repeated k-Fold Cross-Validation for Model Evaluation in Python." [Online]. Disponible en: <https://machinelearningmastery.com/repeated-k-fold-cross-validation-with-python/>
- [34] Hastie, T., Tibshirani, R., & Friedman, J. (2009). "The Elements of Statistical Learning", vol. 2. Springer.
- [35] H. Newman, "Microbiology pictures - photo gallery of medically important bacteria. Microbiology in pictures, bacteria images and photographs with texts. On-line photo atlas of bacteria.", Microbiologyinpictures.com, 2020. [Online]. Disponible en: <https://www.microbiologyinpictures.com/index.php>
- [36] K. Beazer, Clinical Microbiology, Medical Laboratory Sciences. Weber State University. [Online] Disponible en: <https://weber.edu/kendalbeazer/GramStains.html>
- [37] Medical labs, [Online]. Disponible en: <https://www.medical-labs.net/micrococcus-luteus-930/>
- [38] J. Rubin, "The Rubin Lab" [Online]. Disponible en: <https://www.flickr.com/photos/therubinlab/>
- [39] Md. A. Islam, [Online]. Disponible en: <https://www.istockphoto.com/es/foto/escherichia-coli-tambi%C3%A9n-conocida-como-e-coli-es-una-bacteria-gramnegativa-anaer%C3%B3bica-gm1346544897-424285699>  
<https://www.istockphoto.com/es/foto/escherichia-coli-tambi%C3%A9n-conocida-como-e-coli-es-una-bacteria-gramnegativa-anaer%C3%B3bica-gm1346544897-424285699>  
<https://www.istockphoto.com/es/foto/escherichia-coli-tambi%C3%A9n-conocida-como-e-coli-es-una-bacteria-gramnegativa-anaer%C3%B3bica-gm1346544897-424285699>
- [40] M. T. Vazquez-Pertejo, "Microscopía", Wellington Regional Medical Center, Manual MSD versión para profesionales, 2022. [Online]. Disponible en: <https://www.msmanuals.com/es/professional/enfermedades-infecciosas/diagn%C3%B3stico-de-laboratorio-de-las-enfermedades-infecciosas/microscopia>

- [41] PIMCD2012 Proyecto Microbiología, Bioimágenes, Banco de imágenes de la Facultad de Ciencias Biológicas Universidad Complutense de Madrid. [Online]. Disponible en: <http://bioimagen.bioucm.es/foto/6540>
- [42] J. Reischig, "Micrococcus luteus bacteria". file licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. [Online]. Disponible en: [https://commons.wikimedia.org/wiki/File:Eubacteria\\_%28259\\_00F%29\\_Micrococcus\\_luteus\\_bacteria.jpg](https://commons.wikimedia.org/wiki/File:Eubacteria_%28259_00F%29_Micrococcus_luteus_bacteria.jpg)
- [43] N. Jumaah, J. Sanmukh, S. Doblin. (2014). Prevalence of Bacterial Contamination when using a Diversion Pouch during Blood Collection: A Single Center Study in Malaysia. The Malaysian journal of medical sciences : MJMS. 21. 47-53.
- Prevalence of Bacterial Contamination when using a Diversion Pouch during Blood Collection: A Single Center Study in Malaysia - Scientific Figure on ResearchGate. [Online]. Disponible en: [https://www.researchgate.net/figure/Staphylococcus-aureus-morphology-visualised-using-Gram-staining-10-magnification\\_fig2\\_266028804](https://www.researchgate.net/figure/Staphylococcus-aureus-morphology-visualised-using-Gram-staining-10-magnification_fig2_266028804)
- Prevalence of Bacterial Contamination when using a Diversion Pouch during Blood Collection: A Single Center Study in Malaysia - Scientific Figure on ResearchGate. [Online]. Disponible en: [https://www.researchgate.net/figure/Coagulase-negative-Staphylococcus-morphology-visualised-using-Gram-staining-10\\_fig3\\_266028804](https://www.researchgate.net/figure/Coagulase-negative-Staphylococcus-morphology-visualised-using-Gram-staining-10_fig3_266028804)
- [44] [Online]. Disponible en: [www.bacteriainphotos.com](http://www.bacteriainphotos.com)
- [45] E. Reschke, "Staphylococcus aureus, gram positive spherical bacteria, 500X at 35mm". 2012, [Online]. Disponible en: <https://www.gettyimages.com.mx/detail/foto/staphylococcus-aureus-gram-positive-spherical-imagen-libre-de-derechos/139820248>
- [46] [Online]. Disponible en: <https://www.gmpsop.com/sample/MICLAB-070-Identification-of-Microorganisms-to-Genus-and-Species-Level-sample.pdf>
- [47] G.E. Kaiswe, Microbiology, School of Mathematics and Science, Biology Department. College of Baltimore County. [Online] Disponible en: <https://cwoer.ccbcmd.edu/science/microbiology/Lab%20Manual/lab1/lab1.html>
- [48] Quizlet, [Online]. Disponible en: <https://quizlet.com/732371811/mirco-quiz-ex-78-18-19-flash-cards/>
- [49] mICROBIO, "Staphylococcus aureus. Coloración de Gram. Cocos Gram positivos". mICROBIOsh [Online]. Disponible en: <https://x.com/mICROBIOsh/status/1224693953414430726?lang=es>
- [50] A. Bharani, "Learn about streptococcus pyogenes a pathogenic Bacterium". [Online]. Disponible en: <https://arvinbharani.medium.com/learn-about-streptococcus-pyogenes-a-pathogenic-bacterium-6e613a4894f1>

- [51] Dimarion, "Bacterias de escherichia coli (E. coli) en microscopio, ampliación 1000X". [Online]. Disponible en: <https://www.shutterstock.com/es/video/clip-3949175-escherichia-coli-bacteria-e-under-microscope-magnification>
- [52] Scikit-Image documentation, disponible en: <https://www.scikit.image.org>
- [53] Python documentation, disponible en: [www.python.org](http://www.python.org)
- [54] Pytorch documentation, disponible en: [www.pytorch.org](http://www.pytorch.org)
- [55] P. Joshi, "OpenCV with Python By Example". Birmingham, UK: Ed. Packt Publishing, 2015.
- [56] C. M. Bishop, "Pattern recognition and machine learning". E.U.A. Springer, 2006.
- [57] OpenCV documentation, disponible en: <https://www.docs.opencv.org>
- [58] Scikit-learn documentation, disponible en: <https://www.scikit-learn.org>
- [59] P. Robertson, H. Abdelhady, y R. A. Garduño, "The many forms of a pleomorphic bacterial pathogen—the developmental network of Legionella pneumophila" *Frontiers in Microbiology*, vol. 5, p. 670, Dec. 2014
- [60] G. A. Jiménez Tobón, A. Vélez Hoyos, "Tinción de Gram de tejido: alcances y limitaciones" *Medicina & Laboratorio*, vol. 18, núm. 11-12, 2012.