UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA FACULTAD DE CIENCIAS



MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA ÁREA DE COMPUTACIÓN

DIAGNÓSTICO ASISTIDO POR COMPUTADORA PARA LA CLASIFICACIÓN DE LESIONES DE LA PIEL EN IMÁGENES DERMATOSCÓPICAS

TESIS

que para cubrir los requisitos necesarios para obtener el grado de

DOCTOR EN CIENCIAS

Presenta:

EDGAR OMAR MOLINA MOLINA

Ensenada, Baja California, México, Marzo de 2021

TESIS DEFENDIDA POR

Edgar Omar Molina Molina

Y APROBADA POR EL SIGUIENTE COMITÉ

Dra. Selene Solorza Calderón Director del Comité

Cov. Yre K.

Dr. Carlos Yee Romero Miembro del Comité

Dra. Diana Tentori Santa Cruz Miembro del Comité

Dra. Maria Victoria Meza Kubo Miembro del Comité

Josué Álvarez Borrego Dŗ.

Miembro del Comité

erm

Dr. Jesús Ramón Lerma Aragón Coordinador del Programa MYDCI

Marzo de 2021

RESUMEN de la tesis de **EDGAR OMAR MOLINA**, presentada como requisito para la obtención del grado de DOCTOR EN CIENCIAS. Ensenada, Baja California, Marzo de 2021.

DIAGNÓSTICO ASISTIDO POR COMPUTADORA PARA LA CLASIFICACIÓN DE LESIONES DE LA PIEL EN IMÁGENES DERMATOSCÓPICAS

Resumen aprobado por:

Dra. Seleñe Solorza Calderón Director de Tesis

En este trabajo de tesis se presenta una metodología híbrida que combina nueve características extraídas con descriptores fractales de pirámides de base cuadrada con las mil características obtenidas de la red neuronal convolucional DenseNet-201. Esto se hace para mantener las métricas de efectividad ya que gracias al vertiginoso avance de la tecnología y la efectividad demostrada, en los últimos años, las redes neuronales convolucionales han tenido un repunte en diferentes campos de aplicación. Así que, el objetivo de la metodología que se presenta en esta tesis es clasificar imágenes dermatoscópicas a color de lesiones de la piel. Un factor muy importante a considerar al trabajar con este tipo de bases de datos, es el desbalance de elementos por clase (el número de imágenes por clases es distinto). Para esto, se utilizó un ensamblado de tres clasificadores: KNN, SVM lineal y SVM gaussiano. El ensamblado se realizó usando la técnica de votación lineal o por mayoría. Varios de los sistemas desarrollados en esta área son clasificadores binarios: melanoma contra no melanoma, por ser esta una enfermedad cancerígena mortal. Además el método es multiclases, clasifica simultáneamente ocho clases de lesiones de la piel: queratosis actínica, carcinoma de células basales, carcinoma de células escamosas, gueratosis benigna, melanoma, nevus, dermatofibroma y lesión vascular. El sistema se implementó en el lenguaje de programación MatLab y se validó utilizando las imágenes dermatoscópicas a color de la base de datos International Skin Image Collaboration 2019 (ISIC 2019). El diagnóstico asistido por computadora presentó una exactitud de 97.35%, una precisión de 91.61%, una sensitividad de 66.45% y una especificidad de 97.85%.

Palabras Clave: Análisis fractal, CNNs, DenseNet-201, SVM, KNN, Clasificación, ISIC 2019.

ABSTRACT of the thesis presented by **EDGAR OMAR MOLINA** MOLINA, in fulfillment of the requirements of the degree of DOCTOR OF SCIENCE. Ensenada, Baja California, March 2021.

COMPUTER AIDED DIAGNOSIS FOR CLASSIFICATION OF SKIN LESIONS ON DERMOSCOPIC IMAGING

Approved by:

Dra. Selene Solorza Calderón Thesis Advisor

In this thesis work, a hybrid methodology is presented that combines nine characteristics extracted with fractal descriptors of square-based pyramids with the thousand features obtained from the convolutional neural network DenseNet-201. This is done to maintain the effectiveness metrics since thanks to the vertiginous advancement of Technology and the demonstrated effectiveness, convolutional neural networks have had a rebound in different fields of application in recent years. So, the objective of the methodology presented in this thesis is to classify dermoscopic color images of skin lesions. An essential factor to consider when working with this type of database is the classes unbalance (the number of images per class is different). For this, an assembly of three classifiers was used: KNN, linear SVM, and Gaussian SVM. The ensemble of classifiers was carried out using the linear or majority voting technique. Some of the systems developed in this area are binary classifiers: melanoma versus non-melanoma, a deadly cancer disease. Besides, the method is multiclass; it simultaneously classifies eight classes of skin lesions: actinic keratosis, basal cell carcinoma, squamous cell carcinoma, benign keratosis, melanoma, nevus, dermatofibroma, and vascular lesion. The system was implemented in the MatLab programming language and validated using the dermoscopy color images from the International Skin Image Collaboration 2019 database (ISIC 2019). The computer-aided diagnostic system presented an accuracy of 97.35%, a precision of 91.61%, a 66.45% sensitivity, and a 97.85% specificity.

Keywords: Fractal analysis, CNNs, DenseNet-201, SVM, KNN, classification, ISIC-2019.

A mi esposa Marisela Gaxiola Flores y mi hijo Edgar Omar Molina Gaxiola. A mis abuelos Ana María Proaño Holguín e Hilario Molina Acosta. A mi mamá Inés Molina Proaño.

Agradecimientos

A Dios por prestarme vida, salud, fuerza y empeño y sobre todo por nunca soltarme de su mano.

A mi esposa Marisela Gaxiola Flores y a mi hijo Edgar Omar Molina Gaxiola por todo su apoyo y amor incondicional al estar siempre conmigo motivándome a seguir adelante a pesar de las adversidades.

A mi abuela Ana María Proaño Holguín y mi abuelo Hilario Molina Acosta por todo su apoyo y amor, que aunque no pudieron culminar esta etapa de mi vida conmigo, se que están en un mejor lugar y estarán orgullos de mí.

A mi mamá Inés Molina Proaño por darme la vida y creer siempre en mí, apoyándome y motivándome a ser mejor cada día de mi vida.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme la beca (No. 481069) para realizar mis estudios de Doctorado.

A la Universidad Autónoma de Baja California por haberme aceptado para realizar mis estudios de posgrado.

A mi directora de tesis la Dra. Selene Solorza Calderón por su dirección y guía para el desarrollo de este trabajo de tesis.

A los miembros de mi comité de tesis el Dr. Carlos Yee Romero, la Dra. María Victoria Meza Kubo, la Dra. Diana Tentori Santa Cruz y el Dr. Josué Álvarez Borrego por sus consejos y comentarios para mejorar este trabajo.

Esta tesis fue parcialmente apoyada por el proyecto interno UABC 400/2368.

Quiero agradecer especialmente a todas aquellas personas que no mencioné, cuya influencia negativa al dudar y cuestionar mi capacidad, hicieron que yo mismo me exigiera más y más para lograr todo aquello que me propongo.

Tabla de contenido

Resu	men er	n español	i
Resu	men er	n inglés	ii
Dedic	catoria		iii
Agra	decimi	entos	iv
Lista	de fiau	Iras	vii
Lista	de tab	las	viii
I.	Introd 1.1. 1.2. 1.3. 1.4. 1.5. 1.6. 1.7.	ucción Antecedentes Justificación Planteamiento del problema Preguntas de investigación Hipótesis de investigación Objetivos I.6.1. Objetivos específicos Respuesta a las preguntas de investigación y objetivos específicos	1 3 4 5 6 6 7
11.	Carac	t erísticas fractales Antecedentes	18 18 18
III.	Carac III.1.	terísticas CNNArquitectura de redes CNNIII.1.1. Capa de convoluciónIII.1.2. Capa de agrupamientoIII.1.3. Función de activaciónIII.1.4. Capa completamente conectadaDenseNet-201	29 29 30 31 32 34
IV.	Espac IV.1. IV.2.	cios de clasificación K-vecinos cercanos	37 37 38
V.	Diagn V.1. V.2. V.3.	 óstico asistido por computadora Base de datos	40 40 42 43 44

Tabla de contenido (continuación)

VI.	Resultados VI.1. Experimentos	46 46
VII.	Comparativa con otras metodologías	53
VIII.	Discusiones	57
IX.	Conclusiones	59
Х.	Trabajo a futuro	61
REFE	ERENCIAS	62
Α.	Valores de las métricas para el Experimento 3 (Exp-3)	70
В.	Valores de las métricas para el Experimento 4 (Exp-4)	74

Lista de figuras

Figura

1.	Esquema de la obtención de los vértices de la pirámide de base cuadrada (a) Imagen $I(x, y)$. (b) Pirámide de base cuadrada.	20
2.	Esquema de la obtención de vecindarios de tamaño impar para la construcción de las bases de las pirámides. (a) $\varepsilon = 2^1$. (b) $\varepsilon = 2^2$. (c) $\varepsilon = 2^8$	24
3.	Imágenes dermatoscópicas. (a) Queratosis actínica. (b) Carcinoma de célu- las basales. (c) Melanoma.	25
4.	Ejemplo de descriptores fractales. (a) Descriptores fractales de la imagen de la Figura 3(a). (b) Amplificación del descriptor de la Figura 4(a). (c) Descriptores fractales de la imagen de la Figura 3(b). (d) Amplificación del descriptor de la Figura 4(c). (e) Descriptores fractales de la imagen de la Figura 3(c). (f) Amplificación del descriptor de la Figura 4(e)	28
5.	Ejemplo de convolución con un filtro de tamaño 3×3 con paso 1 sobre una imagen $I(x, y)$.	30
6.	Esquema de agrupamiento utilizando la función máx en vecindarios de 2×2 con paso 2	31
7.	Esquema de la función de activación ReLU utilizando una imagen $I(x, y)$ de 4×4 pixeles.	32
8.	Esquema de una capa completamente conectada	33
9.	Esquema de una red CNN.	33
10.	Arquitectura DenseNet-201. (a) Diagrama de bloques de la operación de concatenación, la C dentro del círculo amarillo indica la operación de con- catenación. (b) Diagrama de bloques de la propagación hacia adelante. (c) Diagrama de bloques de la composición de la capa.	36
11.	Esquema del espacio clasificador <i>K</i> -NN usando 5 vecinos cercanos. La estrella roja representa el punto de prueba. Las estrellas negras, círculos y pentágonos son los puntos de entrenamiento	37
12.	Esquema de la máquina de soporte vectorial. Las estrellas negras y el círculo negro representan los vectores de soporte	39
13.	Metodología propuesta. (a) Imagen RGB de entrada. (b) Extracción de características fractales y de DenseNet-201. (c) Descriptor 1D. (d) Clasifica- dores SVM-L, SVM-G y KNN-5. (e) Clasificador de ensamble. (f) Predicción final.	44

Lista de tablas

Tabla

Ι.	Imágenes por clase en ISIC 2019	42
II.	Matriz de confusión del Exp-1 con $\alpha = -2.0$. Utilizando las 25331 imágenes.	47
III.	Rendimiento del Exp-1 obtenido de la matriz de la Tabla II	48
IV.	Métricas del rendimiento de la Tabla III para el Exp-1 con $\alpha=-2.0.$.	48
V.	Métricas de rendimiento obtenidas para el Exp-2 con $\alpha = -1.9$. La base de datos es de 24839 imágenes.	49
VI.	Media \pm DE de cinco repeticiones del Exp-3 con $\alpha = 0.4$. Utilizando datos de entrenamiento de 17732 imágenes	49
VII.	Media \pm DE de cinco repeticiones del Exp-3 para los datos de prueba con $\alpha = 0.4$. Los datos de prueba contienen 7599 imágenes	50
VIII.	Media \pm DE de cinco repeticiones del Exp-4 para los datos de entre- namiento con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.	51
IX.	Media \pm DE de cinco repeticiones del Exp-4 para los datos de prueba con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes	51
Х.	Primer muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entrena- miento contiene 17732 imágenes.	70
XI.	Primer muestra del Exp-3 con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.	70
XII.	Segunda muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entre- namiento contiene 17732 imágenes.	71
XIII.	Segunda muestra del Exp-3 con $\alpha = 0.4$. con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.	71
XIV.	Tercer muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entrena- miento contiene 17732 imágenes	71
XV.	Tercer muestra del Exp-3 con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.	72
XVI.	Cuarta muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entrena- miento contiene 17732 imágenes	72
XVII.	Cuarta muestra del Exp-3 con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.	72

Lista de tablas (continuación)

Tabla

XVIII.	Quinta muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entrena- miento contiene 17732 imágenes.	73
XIX.	Quinta muestra del Exp-3 con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.	73
XX.	Primer muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17,387 imágenes.	74
XXI.	Primer muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.	74
XXII.	Segunda muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.	75
XXIII.	Segunda muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.	75
XXIV.	Tercer muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.	75
XXV.	Tercer muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.	76
XXVI.	Cuarta muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.	76
XXVII.	Cuarta muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes	76
XXVIII.	Quinta muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.	77
XXIX.	Quinta muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes	77

Capítulo I. Introducción

En las últimas décadas, el procesamiento de imágenes ha sido aplicado en diferentes áreas, permitiendo mejorar la información de una imagen para su interpretación, representación, descripción y procesado. El cerebro humano realiza procesamiento de imágenes consciente e inconscientemente, hace transformaciones y distorsiones geométricas para identificar objetos en escenarios compuestos, sin embargo, automatizar dichos procesos en una imagen digital no es una tarea trivial. Con procesamiento digital de imágenes nos referimos a la manipulación de imágenes digitales por medio de computadoras y comprende las áreas de investigación: localización, restauración, reconstrucción, seguimiento, segmentación y reconocimiento de patrones (Gonzalez, 2002). En este trabajo de tesis, me enfocaré al área de reconocimiento de patrones.

De acuerdo con Theodoris (Theodoris, 2008), el reconocimiento de patrones es una disciplina científica cuyo objetivo es la clasificación de objetos en muchas categorías o clases. Tou y González (Tou y Gonzalez, 1977) consideran que un sistema de reconocimiento de patrones puede dividirse esencialmente en tres etapas básicas: la primera es la adquisición de imágenes a clasificar; la segunda consiste en la extracción de características relevantes y reducir la dimensionalidad de los datos, así como reducir el tiempo de cómputo requerido; y en la última fase, mediante una serie de funciones de decisión, se lleva a cabo la clasificación de las imágenes.

En la naturaleza, no siempre es posible encontrar objetos con morfologías bien definidas y bordes regulares, esto implica que automatizar el reconocimiento de patrones de ese tipo de objetos es uno de los grandes retos de esta área de estudio. Dichas metodologías de reconocimiento de patrones pueden aplicarse, por ejemplo, al análisis de derrames de petróleo, uso de suelos, detección de deforestaciones, clasificación de galaxias, clasificación de frentes de temperatura de superficies marinas, clasificación de frentes de clorofila en superficies marinas. En este trabajo de tesis nos enfocamos al desarrollo de una metodología para clasificar lesiones en la piel en imágenes dermatoscópicas a color.

Debido a los cambios en las condiciones ambientales en las que vivimos actualmente, observamos que más personas padecen algún tipo de cáncer de piel (Fitzpatrick, 1975). Por ejemplo, la radiación ultravioleta (UV) es uno de los principales factores de riesgo. En 1975, Fitzpatrick propuso una escala de I a VI de cómo reacciona el tipo de piel a los rayos UV (Fitzpatrick, 1975). El tipo I incluye las pieles muy claras (más susceptibles a desarrollar algún tipo de cáncer), y el tipo VI comprende pieles de color marrón oscuro intensamente pigmentadas (menos afectadas). Por lo tanto, este tipo de enfermedades son más frecuentes en países con una población predominantemente de piel clara. En Irlanda, el diagnóstico de cáncer de piel supera los 11000 casos al año. La mayoría de las personas que viven ahí tienen la piel clara, lo que significa piel tipo I o II en la escala de Fitzpatrick. El gobierno de Irlanda, preocupado por la creciente incidencia de cáncer de piel, prioriza la necesidad de implementar un plan nacional de prevención del cáncer de piel en su Estrategia Nacional 2019-2022 contra el Cáncer (of Health, 2020). Anualmente en los Estados Unidos, el diagnóstico de cáncer de piel supera los cinco millones de casos (Society, 2019). En 2010, las enfermedades de la piel fueron la cuarta causa a nivel mundial de enfermedades no mortales que generaron pérdidas económicas debido a incapacidades (Hay et al., 2014).

Una de las primeras señales de advertencia del cáncer de piel es que las lesiones cutáneas, como lunares nuevos, protuberancias, llagas, escamas o manchas oscuras, crecen y no desaparecen. Principalmente, el cáncer de piel se divide en dos categorías, melanoma y no melanoma (Martínez y Otley, 2001). Por lo regular, los dermatólogos emplean los criterios ABCDE en el diagnóstico: asimetría (A), borde (B), color (C), estructura diferencial (D), y evolución (E) (Nachbar *et al.*, 1994). En la asimetría (A), el dermatólogo busca si la lesión es irregular y mide la irregularidad de los bordes (B). Con el análisis de

color (C) determina si la mancha tiene una coloración inusual. Con la medida del diámetro (D), determina si el punto es más extenso que un cuarto de pulgada. Las características de evolución (E) indican si la lesión está cambiando de tamaño, color o forma. Por lo tanto, el dermatólogo analiza la forma, apariencia y color del lunar o mancha. Hoy en día, la identificación del tipo de lesión en la piel es una de las tareas prioritarias a nivel mundial. Por ende, caracterizar el papel que juegan las texturas en las enfermedades de la piel nos ayudará a desarrollar métodos de diagnóstico asistido por computadora más robustos.

I.1. Antecedentes

La textura es una parte esencial del aspecto de la superficie de la piel. Por esta razón, se han desarrollado técnicas que emplean el uso de las texturas. Estos métodos se basan en análisis estadístico, análisis espectral o dimensión fractal, entre otros. El enfoque estadístico utiliza distribuciones de píxeles (Haralick *et al.*, 1973; Galloway, 1975; Haralick, 1979; Sun y Wee, 1983; Amadasun y King, 1989; Chu *et al.*, 1990; Dasarathy y Holder, 1991; Ojala *et al.*, 1996; Tang, 1998; Guo *et al.*, 2016; Liu *et al.*, 2016; Dash y Senapati, 2018; Al-antari *et al.*, 2018; Al-masni *et al.*, 2018; Al-antari *et al.*, 2020). En los métodos espectrales, los patrones se analizan en el dominio de la frecuencia (Arivazhagan *et al.*, 2006; Regniers *et al.*, 2016; Guerra-Rosas y Álvarez-Borrego, 2015; Depeursinge *et al.*, 2017; Guerra-Rosas *et al.*, 2017; Garza-Flores *et al.*, 2019; Pardo *et al.*, 2018). El marco teórico de la dimensión fractal permite evaluar la complejidad de la imagen (Emerson *et al.*, 1999; Lam *et al.*, 2002).

Las técnicas de extracción basada en texturas se han aplicado a innumerables problemas, logrando excelentes resultados en clasificación multiclase (Chen *et al.*, 2006; Ashour *et al.*, 2016). Sin embargo, la tarea de clasificación de lesiones cutáneas es tan compleja que la mayoría de las propuestas se basan en el problema de clasificación binaria (Garnavi *et al.*, 2012; Barata *et al.*, 2014; Amelard *et al.*, 2015; Spyridonos *et al.*, 2017; Khan et al., 2019; Albahar, 2019; Marka et al., 2019). Debido a que el melanoma es un cáncer mortal siempre y cuando no se detecte a tiempo, la mayoría de esos trabajos se centran en identificarlo. Como la queratosis actínica es una lesión catalogada como pre-cancerígena (Dodson *et al.*, 1991), los trabajos relacionados a su clasificación automatizada son escasos. Entre estos se encuentra el desarrollado por Spyridonos y colaboradores (Spyridonos et al., 2017) que proponen la detección de queratosis actínica en los espacios de color RGB, YIQ y CIE, emplearon LBP (Local Binary Patterns, por su siglas en inglés) para la extracción de características y el espacio clasificador SVM (Support Vector Machine, por sus siglas en inglés) con kernel gaussiano, reportando una exactitud del 91.7% en su mejor configuración utilizando el espacio de color YIQ. Por otra parte, dada la complejidad del problema, pocos trabajos presentan sistemas de clasificación multiclases (Shimizu et al., 2015; Wahba et al., 2018; Garza-Flores et al., 2019; Gessert et al., 2020; Bajwa et al., 2020). Uno de estos es el de Wu y colaboradores (Wu et al., 2019), quienes utilizaron cinco redes neuronales convolucionales para clasificar las lesiones cutáneas de la base de datos Xiangva-Derm. Seleccionaron 2656 imágenes faciales de queratosis seborreica, queratosis actínica, rosácea, lupus eritematoso, carcinoma de células basales y carcinoma de células escamosas. Los mejores resultados los obtuvieron con la red neuronal Inception-ResNet-v2 logrando una sensibilidad del 67.2% y una precisión del 63.7%.

I.2. Justificación

La mayoría de los diagnósticos asistidos por computadora intentan reproducir lo que hace un dermatólogo. El dermatólogo otorga una puntuación a las características ABCDE de la lesión de la piel (Nachbar *et al.*, 1994). El resultado final es la suma de los puntajes dados y de acuerdo al umbral predeterminado, la lesión se determina o no como una enfermedad.

Comúnmente, las metodologías de diagnóstico asistido por computadora comprenden

cuatro pasos: segmentación, extracción de características, puntuación de características y clasificación (Celebi *et al.*, 2007). Una segmentación adecuada debe separar completamente la región sana de la piel enferma, por lo regular esto no es posible. Este tipo de imágenes tienen bordes poco definidos y en la mayoría de los casos, presentan bajo contraste entre las regiones de la piel sana y la enferma, eso complica el proceso de segmentación y provoca una segmentación incorrecta de la lesión. En los trabajos de Silveira y colaboradores (Silveira *et al.*, 2009) y, Celebi y colaboradores (Celebi *et al.*, 2015), se analizaron diferentes algoritmos de segmentación de última generación, que en muchos de los casos no lograron discriminar correctamente la piel sana de la enferma, principalmente en las imágenes de queratosis actínica y carcinoma de células basales. Eso ocasiona que no se consideren secciones significativas de las lesiones.

I.3. Planteamiento del problema

Dada la complejidad en la discriminación entre diferentes lesiones cutaneás de la piel, es necesario descubrir qué estrategias de análisis de textura y color, así como diferentes atributos, son necesarios para lograr una clasificación entre diferentes clases de lesiones de la piel.

I.4. Preguntas de investigación

Con base a la problemática de la presente propuesta de tesis, se planteó la siguiente pregunta de investigación:

¿Qué metodología debe llevarse a cabo para construir un modelo para el reconocimiento de patrones en imágenes digitales para la clasificación de distintas clases de lesiones de la piel con un nivel alto de confiabilidad?

De esta pregunta general se desprende lo siguiente:

- 1. ¿Cuáles son las bases de datos que son referente en este campo de investigación?
- ¿Cuáles son las características más representativas de las manchas de piel para llevar a cabo el reconocimiento automatizado de patrones?
- ¿Cuáles técnicas de extracción de características emplear para la construcción de un modelo para la clasificación de imágenes de lesiones de la piel?
- 4. ¿Cuál espacio de clasificación o grupo de espacios de clasificación son los más eficientes para lograr una correcta discriminación entre diferentes clases de lesiones de la piel?
- 5. ¿Cuáles métricas son las más adecuadas para llevar a cabo la evaluación de la clasificación?

I.5. Hipótesis de investigación

La extracción de características de textura y color, así como de características de nivel alto, son necesarias para una correcta clasificación entre distintas clases de lesiones cutáneas de la piel. Con estas características se podrá obtener un sistema de clasificación con un nivel alto de exactitud.

I.6. Objetivos

Diseñar una metodología de reconocimiento de patrones en imágenes digitales a color para la clasificación de distintas clases de lesiones de la piel.

I.6.1. Objetivos específicos

- 1. Seleccionar la base de datos con la cual se validará la metodología.
- 2. Identificar los atributos más representativos de las imágenes de lesiones.
- Determinar cuales técnicas de extracción de características son las adecuadas para caracterizar los atributos representativos de las imágenes.

- Determinar los espacios de clasificación o la estrategia de clasificación más apropiada para el diagnóstico asistido por computadora.
- 5. Determinar cuales son las métricas más apropiadas para la evaluación de la efectividad en la clasificación.

I.7. Respuesta a las preguntas de investigación y objetivos específicos

El objetivo principal del anteproyecto, consistió en diseñar una metodología de reconocimiento de patrones en imágenes dermatoscópicas para la clasificación de distintas clases de lesiones de la piel. Para lograr dicho objetivo primero fue necesario responder las preguntas de investigación y alcanzar el correspondiente objetivo específico planteado para cada pregunta.

1. ¿Cuáles son las bases de datos que son referente en este campo de investigación?

En la literatura es posible encontrar diferentes tipos de bases de datos de imágenes de lesiones de la piel. Básicamente estas bases de datos, se pueden dividir en dos tipos, las que incluyen imágenes digitales y las de imágenes dermatoscópicas. Las imágenes digitales de lesiones de la piel son capturadas con cámaras fotográficas convencionales o de alta resolución. Por su parte las imágenes dermatoscópicas son adquiridas con dermatoscopios, los cuales son dispositivos electrónicos que iluminan y permiten ampliar la zona de la piel para obtener imágenes más nítidas. Por lo que las imágenes dermatoscópicas tienen una mayor calidad y amplificación de la lesión (Celebi *et al.*, 2007), que hace que se aprecien con mayor claridad los detalles finos de las lesiones, como los bordes, los colores y el contraste entre la piel sana y enferma.

En la literatura de esta área de estudio básicamente se han empleado tres bases de

datos: Dermofit Image Library (Ballerini *et al.*, 2013), Xiangya-Derm (Xie *et al.*, 2019), ISIC (Gutman *et al.*, 2016). La Dermofit Image Library está compuesta por imágenes digitales avaladas por dermatólogos y dermatopatólogos de la Universidad de Edimburgo. Contiene 1300 imágenes digitales de la cara, dividida en 10 clases: queratosis actínica (45 imágenes), carcinoma de células basales (239 imágenes), nevus (331 imágenes), queratosis seborreica (257 imágenes), carcinoma de células escamosas (88 imágenes), carcinoma intraepitelial (78 imágenes), granuloma piogénico (24 imágenes), hemangioma (97 imágenes), dermatofibroma (65 imágenes) y melanoma (76 imágenes). La licencia tiene un precio de 75 euros y está disponible únicamente para uso académico. Esta base de datos contiene muy pocas imágenes por clase, insuficientes para entrenar una red CNN, ya que de acuerdo con Soekhoe (Soekhoe *et al.*, 2016) son necesarias al menos mil imágenes por clase. Por otra parte, una base de datos nos debe permitir reproducir o generar el espacio de clasificación para identificar cada una de las clases, lo que no podremos lograr con la Dermofit Image Library a menos que incrementemos el número de imágenes usando transformaciones afines o distorsiones geométricas.

Otra base de datos de imágenes digitales es la Xiangya-Derm (Xie *et al.*, 2019) que contiene 107,565 imágenes digitales de diferentes partes del cuerpo, dividida en 541 clases. Cabe hacer notar que los autores no mencionan la cantidad de imágenes por clase. La base de datos es propiedad del Departamento de Dermatología del Hospital Xiangya de la Universidad Central Sur, no está disponible para el uso del público en general, aunque los autores mencionan que se espera que pronto esté disponible.

The International Skin Image Collaboration (ISIC) es una base de datos de imágenes dermatoscópicas de diferentes partes del cuerpo. Es resultado de la colaboración de universidades, academia e industria, como el Centro del Cáncer de la Universidad de Arizona, la Universidad de Queensland, la Universidad Médica de Viena y el Centro de investigación Watson IBM T.J. en Yorktown Heights, New York, entre otros. Esta base de datos es completamente gratuita y de libre acceso. Adicionalmente crece año tras año e incluye nuevas versiones con más clases y mayor cantidad de imágenes.

La primera versión es la ISIC 2016 (Gutman *et al.*, 2016), la cual cuenta con 900 imágenes dermatoscópicas, de las cuales 727 imágenes son benignas (no se especifica de qué tipo de lesiones) y 173 de melanoma. La segunda versión es la ISIC 2017 (Codella *et al.*, 2017), que incrementó a 2000 imágenes dermatoscópicas divididas en tres clases: queratosis actínica con 254 imágenes, nevus con 1372 y melanoma con 374 imágenes. La ISIC 2018 (Codella *et al.*, 2018b; Tschandl *et al.*, 2018b) es la tercera versión de la base de datos, consistente de 23,665 imágenes dermatoscópicas, separadas en siete clases de lesiones. Estas clases incluyen 334 imágenes de queratosis actínica, 583 de carcinoma de células basales, 18618 de nevus, 1674 de queratosis benigna, 157 de lesión vascular, 122 de dermatofibroma, y 2177 de melanoma. Por último, la ISIC 2019 contiene 25,331 imágenes distribuidas en ocho clases de lesiones y está compuesta por 867 imágenes de queratosis actínica, 3323 de carcinoma de células basales, 628 de carcinoma de células escamosas, 2624 de queratosis benigna, 4522 de melanoma, 12875 de nevus, 239 de dermatofibroma y 253 de lesión vascular.

Objetivo específico 1. Seleccionar la base de datos con la cuál se validará la metodología.

En este trabajo de tesis se va a utilizar la base de datos ISIC 2019, ya que es la base referente en la literatura y contiene una cantidad considerable de imágenes por clase, lo que la hace viable para poder entrenar espacios clasificadores sin la necesidad de tener que incrementar el número de elementos mediante imágenes generadas por computadora (transformaciones afines o distorsiones geométricas).

2. ¿Cuáles son las características más representativas de las manchas de piel para llevar a cabo el reconocimiento automatizado de patrones?

Para poder diferenciar entre cada una de las clases de las imágenes, es necesario identificar los atributos principales de éstas para poder extraer características apropiadas para la clasificación. Un atributo es una característica distintiva de una imagen, a partir de la cual es posible diferenciarla de otras imágenes (Tou y Gonzalez, 1977). En la práctica diaria, los dermatólogos emplean el criterio ABCDE: asimetría (A), borde (B), color (C), estructura diferencial (D), y evolución (E) de una lesión para determinar si es maligna o benigna (Nachbar *et al.*, 1994).

Objetivo específico 2. Identificar los atributos más representativos de las imágenes de lesiones.

Los atributos más representativos de una lesión de la piel son la asimetría, los bordes, el color, el diámetro y la evolución.

3. ¿Cuáles técnicas de extracción de características emplear para la construcción de un modelo para la clasificación de imágenes de lesiones de la piel?

De acuerdo con Kabbai y colaboradores (Kabbai *et al.*, 2019) la extracción de características se puede dividir en tres niveles. El primero de ellos corresponde a las técnicas de bajo nivel, las cuáles describen a las imágenes con vectores de características que se extraen de las características locales (características extraídas de regiones o vecindarios de pixeles). Los métodos de nivel medio crean una representación global o de la imagen completa a través de la extracción de las estadísticas de los atributos locales (Yan *et al.*, 2017). Por último, los métodos de alto nivel son los basados en las redes neuronales convolucionales (LeCun *et al.*, 1998). Entre las técnicas de bajo nivel se encuentran las metodologías SIFT (Lowe, 2004), SURF (Bay *et al.*, 2008) y los patrones binarios locales (LBP, por sus siglas en inglés) (Ojala *et al.*, 1996). Las técnicas de bajo nivel como SURF y SIFT no fueron consideradas en esta metodología ya que se basan en la localización de bordes para determinar los puntos claves y definir una región de interés (Barajas-Garcia et al., 2019). Como las imágenes de lesiones de la piel presentan en muchos de los casos bajo contraste entre la piel sana y la lesión, SURF y SIFT no detectan una cantidad representativa de puntos claves para poder trabajar con dichas metodologías en este tipo de problemas (Barajas-Garcia et al., 2019). Entre las técnicas de nivel medio se encuentran las bolsas de palabras (BoW, por sus siglas en inglés) (Csurka et al., 2004), las cuales cuentan la ocurrencia de las características locales sin considerar la relación espacial existente entre ellas (Yan et al., 2017). Las técnicas basadas en BoW no fueron consideradas en esta metodología, ya que al estar basadas en el algoritmo SIFT (Csurka et al., 2004), heredan sus deficencias en la búsqueda de puntos clave en imágenes con bordes poco definidos. Por otra parte, las CNN son estrategias de alto nivel que emplean múltiples filtros en cada capa para localizar desde líneas o bordes en sus capas iniciales, hasta encontrar patrones más complejos como sombras en sus capas superiores (Huang et al., 2017). Debido a estas características, las CNN son apropiadas para extraer características tanto simples como complejas, sin embargo, al estar basadas en la función de convolución, no presentan la propiedad de invariancia a escala, traslación o rotación, por lo que se requiere una base de datos muy grande.

Los dermatólogos emplean el criterio ABCDE para clasificar las lesiones de la piel, para esto se debe de contar con una técnica de segmentación automatizada. Uno de los algoritmos ampliamente usados para esta tarea es el propuesto por Otsu (Otsu, 1979), el cual busca separar el objeto de interés mediante la búsqueda del umbral óptimo que divida en dos partes al histograma de frecuencias de los valores de intensidad de la imagen. Este algoritmo tiene un funcionamiento muy eficiente cuando hay un marcado contraste entre el objeto de interés y el fondo, cosa que no sucede en las imágenes de lesiones de la piel. Zortea y colaboradores (Zortea *et al.*, 2017) proponen una variante del algoritmo de Otsu enfocada a lesiones de la piel, sin embargo la metodología sigue presentando un funcionamiento deficiente en la segmentación de lesiones con bordes difusos. Recientemente Al-masni y colaboradores (Al-masni *et al.*, 2018) proponen una metodología de segmentación basada en redes CNN, la cual necesita una gran cantidad de imágenes para entrenar el modelo. Para ello combinan las bases de datos ISIC 2017 y la PH2 las cuales juntas contienen 4000 imágenes, generando a partir de ellas un total 12,000 imágenes por computadora usando la transformación de rotación con ángulos de 90,180 y 270 grados. Hasta el momento no es posible realizar una correcta segmentación automatizada de la lesión de la piel, en consecuencia el análisis de la asimetría (A) como el de los bordes (B) no se puede llevar a cabo tal como lo haría un dermatólogo.

Como las imagenes dermatoscópicas de la piel presentan formas irregulares, con la teoría fractal se puede caracterizar la asimetría y la irregularidad de los bordes (Mandelbrot, 1977). Hasta el momento, el análisis fractal se ha empleado principalmente para automatizar la clasificación de plantas (Backes et al., 2009) e imágenes médicas de quistes (Florindo et al., 2017). De acuerdo con Backes y colaboradores (Backes et al., 2009) la textura es un atributo visual muy importante empleado principalmente para describir la organización de los pixeles en una imagen. En la literatura es posible encontrar diferentes técnicas de análisis de textura. Las matrices de co-ocurrencia de niveles de gris (GLCM, por sus siglas en inglés) son una de las primeras técnicas desarrolladas para automatizar los patrones en las texturas (Haralick et al., 1973). Esta es una técnica de nivel medio empleada en una imagen en escala de grises para contabilizar la frecuencia en la que los dos niveles de gris L_1 y L_2 , son vecinos en una dirección específica θ , por lo regular $\theta = 0,45,90$ y 135 grados. Las matrices de longitudes de cadenas de niveles de gris (GLRLM, por sus siglas en inglés) corresponden a otra de las técnicas muy utilizadas en patrones complejos de textura. Las GLRLM cuentan la frecuencia con que la cadena de longitud n = 1, 2, 3, ..., N, cuyos elementos son el nivel de gris L = 0, 1, ..., 255 aparece

en la imagen en una dirección específica. Por lo regular se calculan las matrices GLRLM de las direcciones 0, 45, 90 y 135 grados (Galloway, 1975).

Objetivo específico 3. Determinar cuáles técnicas de extracción de características son las adecuadas para caracterizar los atributos representativos de las imágenes.

Dado el bajo rendimiento de las técnicas de segmentación en las imágenes de lesiones de la piel, no es eficiente utilizar técnicas de bajo nivel, como SIFT y SURF que se basan en la extracción de los bordes. En esta tesis se va a trabajar con la textura de la imagen completa y con técnicas de medio y alto nivel. La primer técnica de extracción de características a emplear es una red neuronal convolucional (CNN), la cuál nos permitirá extraer bordes, líneas y características locales en las primeras capas, y características más complejas o globales, como parches de texturas en las capas más profundas.

Para que los modelos de diagnóstico asistido por computadora de redes neuronales tengan buen rendimiento se tiene que usar varias redes, como en el trabajo de Gessert y colaboradores (Gessert *et al.*, 2020) que emplearon cinco redes neuronales convolucionales entre las que se encuentran SENet154 (Hu *et al.*, 2018), ResNext (Mahajan, 2018) y EfficientNet-B0 a B7 (Tan y Le, 2019), para clasificar imágenes dermatoscópicas en la base de datos ISIC 2019, obteniendo una sensibilidad de 72.5%. Para reducir el número de operaciones que se requieren para entrenar a cinco redes neuronales, en este trabajo se emplea un modelo híbrido que combina características extraídas de una red neuronal convolucional con descriptores fractales. No se utilizan las características obtenidas con las matrices GLCM y GLRLM porque no aportaron características relevantes o significativas a aquellas obtenidas con los descriptores fractales. 4. ¿Cuál espacio de clasificación o grupo de espacios de clasificación son los más eficientes para lograr una correcta discriminación entre diferentes clases de lesiones de la piel?

Es común que las bases de datos de lesiones de la piel tengan clases desbalanceadas lo cual implica un reto para la clasificación, como es el caso de la base de datos ISIC 2019 seleccionada en este trabajo. Existen básicamente tres alternativas para solucionar el desbalance de clases: el sobremuestreo, el submuestreo y el ensamblado de espacios de clasificación. El sobremuestreo es una técnica que busca generar imágenes adicionales de las clases minoritarias hasta balancear la base de datos. En el submuestreo se eliminan imágenes de las clases mayoritarias hasta tener el balance necesario en la base de datos, sin embargo la generación de imágenes mediante la computadora no es una práctica válida para el área de la dermatología. Para emplear la estrategia de ensamblado existen diferentes espacios de clasificación que se pueden usar para la clasificación de imágenes de lesiones de la piel.

El espacio de clasificación de árbol de decisión es uno de los más sencillos de interpretar, ya que cada nodo del árbol representa una característica, cada rama una regla de decisión y las hojas corresponden al resultado de la clasificación. Otro de los espacios de clasificación existentes es el clasificador bayesiano ingenuo, el cual supone que hay independencia condicional entre las clases a clasificar, lo cual en la práctica no suele suceder. Por su parte, el espacio de clasificación *K* vecinos cercanos (KNN, por sus siglas en inglés) clasifica a cada vector de características de la clase a la que pertenecen los *K* vecinos más cercanos (Cover y Hart, 1967). La máquina de soporte vectorial (SVM, por sus siglas en inglés) encuentra el hiperplano óptimo que maximiza la separación de los dos puntos más cercanos entre cada par de clases (Cortes y Vapnik, 1995).

Objetivo 4. Determinar los espacios de clasificación o la estrategia de clasificación más apropiada para el diagnóstico asistido por computadora

Dado que el submuestreo y el sobremuestreo no son opciones apropiadas para trabajar con bases de datos de imágenes de lesiones de la piel, en este trabajo se utiliza la estrategia de ensamblado de los espacios de clasificación KNN y SVM con kernel lineal y gaussiano. Los espacios de clasificación de árboles de decisión y el clasificador bayesiano ingenuo no fueron utilizados debido a su baja efectividad en la clasificación de bases de datos desbalanceadas.

5. ¿Cuáles métricas son las más adecuadas para llevar a cabo la evaluación de la clasificación?

Uno de los aspectos cruciales en la clasificación de imágenes es poder evaluar correctamente la efectividad de los modelos de reconocimiento de patrones. Una de las estrategias para evaluar el rendimiento de los modelos de reconocimiento de patrones son los intervalos de confianza, los cuales son utilizados para cuantificar el error en la incertidumbre de la precisión de la predicción de los modelos de reconocimiento de patrones (Foody, 2009). Otra métrica frecuentemente empleada en reconocimiento de patrones es la correlación de Pearson, la cual mide la relación lineal entre dos variables aleatorias (Benesty *et al.*, 2009). Otra medida es la correlación no lineal, la cual emplea el módulo y la fase de la transformada de Fourier de una imagen objetivo y una de referencia con un factor k para medir la similaridad entre dos señales o imágenes (Kumar y Hassebrook, 1990).

Como el objetivo es desarrollar una metodología de diagnóstico asistido por computadora para clasificar imágenes de lesiones de la piel, los resultados de la clasificación deben medirse con las métricas empleadas por los dermatólogos. En los ensayos clínicos se emplea una metodología basada en las interpretaciones de las pruebas clínicas donde pueden existir cuatro resultados. Al confirmarse que un paciente presenta una enfermedad y el resultado del diagnóstico corrobora la existencia de dicha enfermedad, el resultado es verdadero positivo (TP, por sus siglas en inglés). Por otra parte, cuando al paciente no se le detecta alguna enfermedad y el diagnóstico señala que efectivamente no la presenta, el diagnóstico es verdadero negativo (TN, por sus siglas en inglés). Si al realizar una prueba de diagnóstico se encuentra que el paciente presenta una enfermedad, pero ese paciente realmente no presenta tal enfermedad, el resultado es falso positivo (FP, por sus siglas en inglés). Por último, si el resultado de una prueba de diagnóstico indica que el paciente no presenta la enfermedad pero realmente ese paciente sí está enfermo, el resultado es falso negativo (FN, por sus siglas en inglés) (Zhu y Zeng, 2010). Con los valores de FP, TP, FN y TP podemos calcular cuatro métricas ampliamente usadas en la literatura médica para evaluar el rendimiento de la metodología propuesta: exactitud, precisión, sensibilidad y especificidad (Metz, 1978). La exactitud mide la proporción de las pruebas totales diagnosticadas correctamente. La precisión mide la proporción de pruebas positivas diagnosticadas verdaderamente como positivas. La sensibilidad o exhaustividad (recall en inglés) mide la proporción de pacientes con una enfermedad que tienen una prueba positiva. La especificidad mide la proporción de pacientes sin una enfermedad que tienen una prueba negativa (Chu, 1999). Algunas otras métricas basadas en los valores TP, FN, TN, FP son: la métrica F (F-score, en inglés) la cual es la media harmónica de precisión y sensibilidad (Chinchor, 1992), la curva ROC (Receiver Operating Characteristic), la cual se calcula como la relación entre la sensibilidad o fracción de verdaderos positivos (TPF, por sus siglas en inglés) y la especificidad o fracción de verdaderos negativos (TNF, por sus siglas en inglés) al variar un umbral de discriminación a partir del cual existe la enfermedad. A partir de la curva ROC es posible calcular el AUC (Area Under Curve) la cual proporciona una medición del rendimiento de todos los umbrales de clasificación posibles (Metz, 1978).

Objetivo 5. Determinar cuáles son las métricas más apropiadas para la evaluación de la efectividad en la clasificación.

Para la evaluación del diagnóstico asistido por computadora de imágenes dermatoscópicas de lesiones de la piel, se utilizarán las métricas empleadas por los médicos en los ensayos clínicos, esto es, la exactitud, precisión, sensibilidad y especificidad.

Capítulo II. Características fractales

II.1. Antecedentes

El término fractal surgió de los trabajos de Benoit Mandelbrot cuando estudió estructuras geométricas irregulares que se repetían a diferentes escalas (Mandelbrot, 1977). Una de las características distintivas de los objetos fractales es que las medidas estándar no pueden determinar su dimensión. La dimensión fractal señala cómo el objeto fractal llena el espacio a medida que se refina la unidad de medida. Generalmente, la dimensión fractal es un número racional en lugar de un número entero, por ejemplo el punto, la línea recta, el cuadrado y el cubo tienen dimensión cero, uno, dos y tres respectivamente.

Por su naturaleza, las lesiones de la piel presentan bordes irregulares, por lo que es natural medir su tamaño con la dimensión fractal, por lo que Wahba y colaboradores (Wahba *et al.*, 2018) recomendaron incluir la dimensión fractal en la regla ABCDE. Recientemente, Backes y colaboradores (Backes *et al.*, 2009) propusieron un descriptor de textura basado en la dimensión volumétrica fractal (descriptor de Bouligand - Minkowski (Schroeder, 1991)) para identificar plantas. Por otro lado, Florindo y colaboradores (Florindo *et al.*, 2017) utilizaron los descriptores de Bouligand-Minkowski para identificar imágenes histológicas de queratoquiste odontogénico, un tipo de quiste mandibular.

II.1.1. Descriptores fractales

Florindo y Bruno (Florindo y Bruno, 2019) proponen el uso de descriptores fractales basados en pirámides de base cuadrada. En su trabajo, parten de la necesidad e importancia de analizar todos los pixeles de las imágenes a diferentes escalas. Ese análisis pixel a pixel con pirámides de base cuadrada se realiza para lograr una interpretación exhaustiva de la textura de las imágenes de objetos naturales. Comúnmente, las texturas no pueden ser analizadas con la geometría euclidiana. Un ejemplo típico de este tipo de texturas es la presente en las imágenes dermatoscópicas de lesiones cutáneas. Dada una imagen de tamaño $M \times N$ en escala de grises y denotada por I(x, y), a partir de ella se construye una pirámide de base cuadrada (Figura 1(b)) con una base cuadrada de lado $\varepsilon + 1$ y de vértices

$$a = \left(x, y, I(x, y)\right),\tag{1}$$

$$b = \left(x + \varepsilon, y, I(x + \varepsilon, y)\right),\tag{2}$$

$$c = \left(x + \varepsilon, y + \varepsilon, I(x + \varepsilon, y + \varepsilon)\right),\tag{3}$$

$$d = \left(x, y + \varepsilon, I(x, y + \varepsilon)\right),\tag{4}$$

$$e = \left(x + \frac{\varepsilon}{2}, y + \frac{\varepsilon}{2}, z\right),\tag{5}$$

donde z esta dada por

$$z = \frac{I(x,y) + I(x+\varepsilon,y) + I(x,y+\varepsilon) + I(x+\varepsilon,y+\varepsilon)}{4}.$$
(6)

En la Figura 1(b) se muestra un bosquejo del proceso para obtener los vértices y las aristas de la pirámide de base cuadrada. A partir de esos vértices y aristas se calculan las longitudes de las aristas, mediante





$$s = \sqrt{(I(x+\varepsilon, y) - I(x, y))^2 + \varepsilon^2},$$
(7)

$$t = \sqrt{(I(x+\varepsilon, y+\varepsilon) - I(x+\varepsilon, y))^2 + \varepsilon^2},$$
(8)

$$u = \sqrt{(I(x, y + \varepsilon) - I(x + \varepsilon, y + \varepsilon))^2 + \varepsilon^2},$$
(9)

$$v = \sqrt{(I(x,y) - I(x,y+\varepsilon))^2 + \varepsilon^2},$$
(10)

$$o = \sqrt{(I(x,y) - z)^2 + \left(\frac{\sqrt{2}}{2}\varepsilon\right)^2},$$
(11)

$$p = \sqrt{(I(x+\varepsilon, y) - z)^2 + \left(\frac{\sqrt{2}}{2}\varepsilon\right)^2},$$
(12)

$$q = \sqrt{(I(x+\varepsilon, y+\varepsilon) - z)^2 + \left(\frac{\sqrt{2}}{2}\varepsilon\right)^2},$$
(13)

$$r = \sqrt{(I(x, y + \varepsilon) - z)^2 + \left(\frac{\sqrt{2}}{2}\varepsilon\right)^2}.$$
 (14)

Una vez que se conocen las longitudes de las aristas, se calculan los subperímetros de las cuatro caras triangulares y por medio de la ecuación de Herón (Buchholz y Rathbun, 1998) se obtienen las áreas con las siguientes ecuaciones,

$$s_A = \frac{1}{2} \left(s + p + o \right),$$
 (15)

$$s_B = \frac{1}{2} \left(t + p + q \right),$$
 (16)

$$s_C = \frac{1}{2} \left(u + q + r \right),$$
 (17)

$$s_D = \frac{1}{2} \left(v + o + r \right),$$
 (18)

$$A = \sqrt{s_A (s_A - s) (s_A - p) (s_A - o)},$$
(19)

$$B = \sqrt{s_B (s_B - t) (s_B - p) (s_B - q)},$$
 (20)

$$C = \sqrt{s_C (s_C - u) (s_C - q) (s_C - r)},$$
(21)

$$D = \sqrt{s_D (s_D - v) (s_D - o) (s_D - r)}.$$
 (22)

Una vez que se conocen las áreas de las caras triangulares, se construye el descriptor fractal de la imagen, mediante la siguiente ecuación

$$F(\delta) = \sum_{x=1}^{M-\varepsilon} \sum_{y=1}^{N-\varepsilon} A_{x,y}^{\alpha}(\delta),$$
(23)

donde $\delta = 1, 2, \dots, \lfloor \log_2 (\min(M, N)) \rfloor$, α es un número real, y

$$A_{x,y}^{\alpha}(\delta) = (A + B + C + D)^{\alpha},$$
(24)

es la suma de los valores del área de los cuatro triángulos de la pirámide de base cuadrada en la Figura 1b. ε es un valor par y está dado por

$$\varepsilon = 2^{\delta}.$$
 (25)

En la Figura 2 se muestra un bosquejo del proceso de obtención de las bases de las pirámides. La primera entrada del descriptor fractal se obtiene para $\varepsilon = 2^1$, por lo que el

lado de la base cuadrada de la pirámide es de longitud 3. En la Figura 2(a) se muestra un bosquejo de la construcción de las bases de lado 3 a partir de los pixeles de la imagen. Empezamos a construir las bases de las pirámides por el pixel en la posición (1,1), entonces, para la primer pirámide de base cuadrada de base 3×3 , el vértice a en la Figura 1(b) corresponde al pixel (1,1), el b es el pixel (3,1), el c es el (3,3) y el d es el (1,3), y se barren todos los pixeles de la imagen, por lo que el primer índice del descriptor se calcula como $F(1) = \sum_{x=1}^{M-2} \sum_{y=1}^{N-2} A_{x,y}^{\alpha}(1)$. Análogamente, la segunda entrada del descriptor fractal se obtiene para $\varepsilon = 2^2$, por lo que la base cuadrada de la pirámide tiene longitud 5. En la Figura 2(b) se muestra un bosquejo de la construcción de las bases de lado 5 a partir de los pixeles de la imagen, por lo que $F(2) = \sum_{x=1}^{M-4} \sum_{y=1}^{N-4} A_{x,y}^{\alpha}(2)$. El proceso se repite, hasta llegar al último índice del descriptor, en este ejemplo la octava entrada, esto es, $\varepsilon = 2^8$, por lo que la base cuadrada de la pirámide tiene ahora longitud 257, en la Figura 2(c) se muestra un bosquejo de la construcción de las bases de lado 257 a partir de los pixeles de la imagen, por lo que $F(8) = \sum_{x=1}^{M-256} \sum_{y=1}^{N-256} A_{x,y}^{\alpha}(8)$.

(1, 1)		(1,1+arepsilon)	
	$\left(1+\frac{\varepsilon}{2},1+\frac{\varepsilon}{2}\right)$		
$(1+\varepsilon,1)$		$(1+\varepsilon,1+\varepsilon)$	



(1, 1)		$(1,1+\varepsilon)$		
	$(1+\frac{\varepsilon}{2},1+\frac{\varepsilon}{2})$			
$(1+\varepsilon,1)$		$(1+\varepsilon,1+\varepsilon)$		

(b)



Figura 2: Esquema de la obtención de vecindarios de tamaño impar para la construcción de las bases de las pirámides. (a) $\varepsilon = 2^1$. (b) $\varepsilon = 2^2$. (c) $\varepsilon = 2^8$.



(a)

(b)



(C)

Figura 3: Imágenes dermatoscópicas. (a) Queratosis actínica. (b) Carcinoma de células basales. (c) Melanoma.

En la Figura 3 se muestran ejemplos de imágenes dermatoscópicas de lesiones de la piel. En la Figura 3(a) se tiene un ejemplo de queratosis actínica, en la Figura 3(b) de carcinoma de células basales y en la Figura 3(c) de melanoma. Tanto la imagen de queratosis actínica como la de carcinoma de células basales son de tamaño 600×450 pixeles, por lo que $\delta = 1, 2, \ldots, \lfloor \log_2(\min(600, 450)) \rfloor = \lfloor (8.138) \rfloor = 8$. Entonces, los descriptores de dichas imágenes son de longitud 8. Mientras que la imagen de melanoma es de 722×542 pixeles por lo que $\delta = 1, 2, \ldots, \lfloor \log_2(\min(722, 542)) \rfloor = \lfloor (9.0821) \rfloor = 9$, se tiene un descriptor de longitud 9.
En la Figura 4, se muestran los descriptores fractales para las imágenes de las Figuras 3(a) y 3(b), donde puede apreciarse que para cada descriptor por canal de color RGB de cada imagen, se observa la misma tendencia creciente. Para el caso de la Figura 4(e) el comportamiento de los descriptores por canal RGB es creciente por cada posición del descriptor hasta la posición 8, mientras que el valor del descriptor en la posición 9 decrece. Aunque los descriptores por cada canal de color tienen comportamiento similar, debido a la escala, hay diferencia significativa entre ellos. Para la gráfica de la Figura 4(a), la magnitud de la diferencia de los valores máximos de los descriptores fractales son: máx $\{|F_R - F_G|\} = 1.738 \times 10^8$, máx $\{|F_R - F_B|\} = 2.90 \times 10^8$, $\max\{|F_G - F_B|\} = 1.169 \times 10^8$. Análogamente, en el caso de la gráfica de la Figura 4(c), la magnitud de la diferencia de los valores máximos de los descriptores fractales son: $\max\{|F_R - F_G|\} = 6.81 \times 10^7, \max\{|F_R - F_B|\} = 9.44 \times 10^7, \max\{|F_G - F_B|\} = 2.62 \times 10^7.$ Por último, para la gráfica de la Figura 4(e), la magnitud de la diferencia de los valores máximos de los descriptores fractales son: máx $\{|F_R - F_G|\} = 5 \times 10^8$, máx $\{|F_R - F_B|\} =$ 6.17×10^8 , máx { $|F_G - F_B|$ } = 1.16×10^8 . El uso de estos descriptores por canal RGB es indispensable para el análisis total y exhaustivo de las imágenes.

Dado que la longitud del descriptor fractal depende de las dimensiones espaciales de cada imagen y como no se tienen imágenes del mismo tamaño, es necesario homogeneizar la dimensión de los descriptores. Por esta razón en este trabajo de tesis, se propone el uso de la energía E_k , la varianza σ_k^2 , y la entropía H_k de los descriptores fractales (Deep, 2006), calculadas como

$$E_k = \sum_{j=1}^n F_k^2(j),$$
 (26)

$$\sigma_k^2 = \frac{1}{n-1} \sum_{j=1}^n \left(F_k(j) - \overline{F}_k \right)^2,$$
(27)

$$H_k = -\sum_{j=1}^n \left[F_k(j) \cdot \log_2 \left(F_k(j) \right) \right],$$
(28)

donde \overline{F}_k es la media de F_k , k = R, G, B, y n es la longitud del descriptor F_k , Ecuación (23). La energía mide la fuerza del descriptor. La varianza es la media de la dispersión, mientras que la entropía es el contenido de información esperado de la señal.

Tomando en cuenta que el color es una de las características de la regla ABCDE, es de suma importancia preservar la información del color de las imágenes. Es por ello que, la energía, varianza y entropía de los tres descriptores de la imagen, uno por cada canal RGB, componen el descriptor fractal de la siguiente manera

$$S_F = [E_R, E_G, E_B, \sigma_R^2, \sigma_G^2, \sigma_B^2, H_R, H_G, H_B].$$
 (29)



Figura 4: Ejemplo de descriptores fractales. (a) Descriptores fractales de la imagen de la Figura 3(a). (b) Amplificación del descriptor de la Figura 4(a). (c) Descriptores fractales de la imagen de la Figura 3(b). (d) Amplificación del descriptor de la Figura 4(c). (e) Descriptores fractales de la imagen de la Figura 3(c). (f) Amplificación del descriptor de la Figura 4(e).

Capítulo III. Características CNN

Hoy en día, los avances en la tecnología han permitido la creación de máquinas inteligentes que pueden realizar ciertas tareas como los humanos. Esta es un área llamada Inteligencia Artificial (Di *et al.*, 2018). En el aprendizaje profundo, las máquinas automáticamente descubren las representaciones necesarias para llevar a cabo una detección de características o tareas de clasificación utilizando datos sin procesar, sin embargo, es necesaria una gran cantidad de información para aprender todas las posibles variantes en las que los datos se pueden presentar y con ello obtener resultados satisfactorios. El aprendizaje profundo permite a los modelos computacionales aprender representaciones de datos a través de módulos o capas cada vez más complejas, donde las representaciones de nivel superior permiten extraer características relevantes de los datos de entrada, éstas a su vez pueden ser utilizadas en tareas de discriminación, clasificación o detección (Bermant *et al.*, 2019).

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son variantes biológicamente inspiradas en las redes neuronales artificiales y son utilizadas efectivamente en tareas de reconocimiento de imágenes y clasificación. Las CNN explotan la correlación espacial-local empleando un patrón de conectividad local entre las neuronas de capas adyacentes. Para comprender el funcionamiento de las redes CNN, es necesario definir un conjunto de componentes que conforma su arquitectura: capa de convolución, función de activación, capa de agrupación y capas completamente conectadas (LeCun *et al.*, 1998).

III.1. Arquitectura de redes CNN

III.1.1. Capa de convolución

Para una imagen I(x, y), la capa de convolución emplea la matriz de pesos llamada kernel o filtro, la cual se desplaza por la imagen (Figura 5). Una red neuronal convo-

lucional emplea diferentes filtros dentro de sus capas, lo cual hace que se encuentren diferentes detalles desde líneas y bordes en sus capas inferiores, hasta detalles más elaborados como sombras en sus capas superiores (LeCun *et al.*, 1998).



Figura 5: Ejemplo de convolución con un filtro de tamaño 3×3 con paso 1 sobre una imagen I(x, y).

III.1.2. Capa de agrupamiento

La capa de agrupamiento, reduce progresivamente el tamaño espacial de las capas, esto con la finalidad de reducir el número de parámetros y costo computacional. En imágenes a color se realiza independientemente del canal de color con que se trabaje. Una función empleada comúnmente por su efectividad en la práctica es la función máximo (LeCun *et al.*, 1998). Las capas de agrupamiento con valores máximos, emplean vecindarios de 2×2 con paso de 2 pixeles, donde el valor resultante en la imagen de salida es el máximo en cada vecindario (Figura 6).



Figura 6: Esquema de agrupamiento utilizando la función máx en vecindarios de 2×2 con paso 2.

III.1.3. Función de activación

La función de activación en cada capa decide si las señales de entrada cumplen con un umbral y deben emitir ciertas señales para el siguiente nivel o capa. La correcta elección de una función de activación es crucial para reducir el costo computacional en la búsqueda del mínimo global mediante el método del gradiente. Una de las funciones de activación utilizadas en redes CNNs es la unidad lineal rectificada (Rectified Linear Unit, ReLU, por sus siglas en inglés), la cuál tiene la capacidad de entrenar a la red con mayor eficiencia computacional (Al-masni *et al.*, 2018). En la Figura 7 se muestra una bosquejo de cómo opera la función de activación ReLU en una imagen I(x, y) de 4 × 4 pixeles, donde se observa que los valores negativos en I(x, y), marcados en recuadros de color rojo, después de aplicar la función ReLU son cero en la imagen procesada, indicados en recuadros color verde. La capa de unidad lineal rectificada (ReLu) (Glorot *et al.*, 2011), se define como:

$$\phi(x) = \max(0, x) = \begin{cases} x, & x \ge 0, \\ 0, & x < 0. \end{cases}$$
(30)



Figura 7: Esquema de la función de activación ReLU utilizando una imagen I(x, y) de 4×4 pixeles.

III.1.4. Capa completamente conectada

La capa completamente conectada contiene todas las conexiones de todas las capas precedentes de la red. En un modelo de red CNN, las últimas dos capas completamente conectadas son la vectorización de los mapas de características o funciones de activación de las últimas capas, las cuales son utilizadas para generar probabilidades en una función de suavizado máximo o alimentar a un clasificador probabilístico para la clasificación (LeCun *et al.*, 1998). En la Figura 8 se muestra un bosquejo de la obtención de la capa completamente conectada para una imagen dermatoscópica a color, donde a partir de las capas convolucionales indicadas en recuadros de color gris claro y obscuro, son extraídos los mapas de características, seguidos por capas de agrupamiento que reducen la dimensión espacial de los mapas de características resultantes. Este proceso es realizado sucesivamente hasta la obtención del mapa de características agrupado, donde su vectorización da como resultado la capa completamente conectada, indicada con círculos de color verde.



Figura 8: Esquema de una capa completamente conectada.

Una red neuronal convolucional, típicamente consiste de capas de convolución, seguidas por agrupamiento con valor máximo, mientras que en sus últimas capas son extraídas las capas completamente conectadas. En la Figura 9, se muestra un esquema de una arquitectura de red CNN básica.



III.2. DenseNet-201

Esta propuesta consiste en una metodología híbrida que combina las características de los descriptores fractales con las obtenidas con la red neuronal convolucional DenseNet-201. Al igual que otras redes neuronales, las redes DenseNets, en sus primeras capas, obtienen las líneas y las características de los bordes, mientras que en las capas más profundas, encuentran detalles más específicos en las imágenes (Huang *et al.*, 2017). La red DenseNet-201 fue seleccionada porque demostró manejar mejor tanto los bordes de bajo contraste, como las imágenes de lesiones cutáneas, (Huang *et al.*, 2017). Las redes DenseNets conectan todas las capas en forma secuencial (Figura 10(a)). Por lo tanto, cada capa obtiene información adicional de todos los niveles anteriores y pasa sus mapas de características a las capas posteriores. En lugar de sumar las características, antes de que sean los datos de entrada de la siguiente capa, la DenseNet-201 las concatena. Eso da como resultado un patrón de conectividad más denso en cada una de las capas (Figura 10(b)). El número de canales adicionales en cada capa se denomina tasa de crecimiento, en este caso específico la red DenseNet-201 utiliza 32.

Para cada capa de composición, se utilizan en cada canal la pre-activación de norma por lote, ReLU y convoluciones con vecindarios de 3×3 con paso 1, cuyas operaciones generaron un nuevo mapa de características de salida (Figura 10(c)). En la capa de norma de lote, los valores en la matriz de activación se dividen por la suma de los valores en la matriz del núcleo. Esto se hace para tener todos los valores entre 0 y 1. La función de activación ReLu le da a la CNN la característica de no linealidad.

La DenseNet-201 usa múltiples bloques densos con capas de transición (Figura 10). Primero, utiliza una capa de convolución de tamaño 7×7 con paso de 2 pixeles, seguida de una capa de agrupamiento por valor máximo de 3×3 con paso de 2 pixeles. Entre dos bloques densos contiguos, emplea una capa de convolución de tamaño 1×1 con paso 1, seguido de una capa de agrupación promedio de tamaño 2×2 con paso 2. Esta propuesta de tesis utiliza cuatro bloques densos. La red DenseNet-201 requiere cambiar el tamaño de las imágenes de entrada a 224×224 píxeles por canal. Se usó la función MatLab 2019b augmentedImageDatastore, que utiliza la transformación afín de la escala (Gonzalez *et al.*, 2010). Además, se empleó la operación de interpolación bilineal, porque entre los métodos del vecino más cercano, bilineal y bicúbico, es el que proporciona resultados satisfactorios sin consumir tanto tiempo de cómputo como lo hace el método bicúbico (Gonzalez, 2002). La salida de la red DenseNet-201 es un vector de longitud 1000, llamado vector de características 1D CNN, y denotado por S_{CNN} .



Figura 10: Arquitectura DenseNet-201. (a) Diagrama de bloques de la operación de concatenación, la C dentro del círculo amarillo indica la operación de concatenación. (b) Diagrama de bloques de la propagación hacia adelante. (c) Diagrama de bloques de la composición de la capa.

Capítulo IV. Espacios de clasificación

IV.1. K-vecinos cercanos

El *K* vecinos cercanos, (KNN, por sus siglas en inglés) es un clasificador no probabilístico (Cover y Hart, 1967). Es muy popular debido a su simplicidad y excelente rendimiento. Cuando usamos este tipo de clasificador, tenemos *N* vectores de entrenamiento $\mathbf{x}_j = [S_F, S_{CNN}] \in \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8\}, j = 1, 2, ..., N,$

y { $C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8$ } es el conjunto de clases de lesiones de la piel, x_j es el resultado de la concatenación de las características extraídas del descriptor fractal, denotado por S_F y el vector de características extraídas de la red CNN, nombrado como S_{CNN} . Para determinar la clase a la que pertenece un vector de clase desconocida x, encontramos los K descriptores que son más cercanos a x. La clase mayoritaria en esos vecindarios es la clase asignada a x. En este trabajo de tesis, usamos K = 5 ya que con este valor se obtienen los mejores resultados, y al espacio clasificador lo denotamos por KNN-5. Empleamos la distancia euclidiana para calcular la distancia entre los datos de entrenamiento x_j y el descriptor a clasificar x, dada por

$$d(\mathbf{x}_j, \mathbf{x}) = \sqrt{\sum_{i=1}^{1009} \left[\mathbf{x}_j(i) - \mathbf{x}(i) \right]^2},$$
(31)

donde j = 1, 2, ..., N, Figura 11.



Figura 11: Esquema del espacio clasificador *K*-NN usando 5 vecinos cercanos. La estrella roja representa el punto de prueba. Las estrellas negras, círculos y pentágonos son los puntos de entrenamiento.

IV.2. Máquinas de soporte vectorial

Este es un clasificador no probabilístico. La máquina de soporte vectorial (SVM, por sus siglas en inglés) se basa en las máquinas de entrenamiento lineal con márgenes (Cortes y Vapnik, 1995). La función discriminante lineal tiene la forma general

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0, \tag{32}$$

donde w es un vector de pesos y w_0 es el sesgo o umbral de peso.

En la metodología SVM, cada patrón $\mathbf{x}_j = [S_F, S_{CNN}]$ es transformado por

$$\mathbf{y}_j = \varphi(\mathbf{x}_j),\tag{33}$$

donde j = 1, 2, ..., N. Con un mapeo no lineal apropiado φ , los datos se pueden separar por un hiperespacio, Figura 12. Por cada uno de esos N patrones, tenemos $z_j = \pm 1$. Un discriminante lineal es un espacio aumentado y está dado por

$$g(\mathbf{y}) = \mathbf{a}^T \mathbf{y},\tag{34}$$

donde se aumentan tanto el vector de peso como el vector de patrón transformado (compararlo con la Ecuación (32)). El primer paso para entrenar una SVM es elegir la función no lineal φ . El objetivo es minimizar la magnitud del vector de peso. Para lograrlo, el problema se reescribe como la maximización de la distancia del hiperplano usando la construcción de Kuhn-Tucker (Kuhn y Tucker, 1951), esto es,

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \alpha_i \alpha_j z_i z_j \mathbf{y}_i^T \mathbf{y}_j,$$
(35)

$$\sum_{i=1}^{M} \alpha_i z_i = 0, \quad \alpha_i \ge 0.$$
(36)

Sea $\mathbf{K} = (K_{ij})_{i,j=1,2,...,M}$ una matriz de $M \times M$, y usando la Ecuación (33), obtenemos que

$$K_{ij} = \mathbf{y}_i^T \mathbf{y}_j = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle.$$
(37)

En la Ecuación (35) no se requiere conocer el mapeo $\varphi(\mathbf{x}_i)$, basta con conocer el kernel asociado K_{ij} . Dos de los núcleos más comunes son el núcleo lineal (SVM-L),

$$K_{ij} = \mathbf{y}_i^T \mathbf{y}_j,\tag{38}$$

y el núcleo gaussiano (SVM-G),

$$K_{ij} = exp\left[-\frac{||\mathbf{y}_i - \mathbf{y}_j||^2}{2\sigma^2}\right],\tag{39}$$

donde σ es la desviación estándar.



Figura 12: Esquema de la máquina de soporte vectorial. Las estrellas negras y el círculo negro representan los vectores de soporte.

Capítulo V. Diagnóstico asistido por computadora

V.1. Base de datos

Las metodologías de diagnóstico asistido por computadora se tienen que probar con bases de datos avaladas por dermatólogos, sin embargo no existen muchas de ellas. Una de estas bases de datos es la Dermofit Image Library (Ballerini et al., 2013), la cual consta de 1300 imágenes de alta calidad, divididas en diez clases de lesiones. La adquisición de esta base de datos está limitada al uso de universidades con restricciones de investigación solamente. Para poder utilizarla, es necesario que la universidad adquiera una licencia por 75 euros. Además, otra de las limitantes de esta base de datos es que contiene una reducida cantidad de imágenes por lesión: 45 imágenes de gueratosis actínica, 239 de carcinoma de células basales, 331 de nevus, 257 de gueratosis seborreica, 88 de carcinoma de células escamosas, 78 de carcinoma intraepitelial, 24 de granuloma piogénico, 97 de hemangioma, 65 de dermatofibroma y 76 imágenes de melanoma. Como se puede observar al ser la queratosis actínica una enfermedad benigna, solamente se tienen 45 muestras de ella, una cantidad muy baja para poder hacer pruebas apropiadamente. Otra de las bases de datos recientes es la XiangyaDerm (Xie et al., 2019), la cual contiene 107565 imágenes adquiridas con cámaras fotográficas convencionales de alta resolución, divididas en 541 clases de lesiones de la piel. Sin embargo esta base de datos no está disponible para el uso del público en general, sino únicamente para los integrantes del grupo de investigación donde fue desarrollada. Una de las bases de datos ampliamente utilizada es la ISIC (International Skin Imaging Collaboration, por sus siglas en inglés), es gratuita y es resultado de la cooperación entre la academia y la industria para proponer soluciones en la clasificación de enfermedades de la piel. Desde el año 2016, la ISIC organiza diferentes retos, tanto en segmentación como en clasificación de imágenes dermatoscópicas. Esta base de datos crece año con año incorporando más imágenes y clases de las mismas. En 2016, la base de datos contaba solamente con dos clases, melanoma y no melanoma, con un total de 900 imágenes. La versión de ISIC 2017, incrementó a 2000 imágenes, de las cuales 374 son de melanoma, 1372 nevus y 254 de queratosis seborreica. Ya para el 2018 contiene 23665 imágenes de siete clases de lesiones: 334 de queratosis actínica (compuesta de imágenes de queratosis actínica y enfermedad de Bowen), 583 carcinoma de células basales, 1674 queratosis benigna (conformada por léntigo solar, queratosis seborreica y queratosis similar al liquen plano), 122 dermatofibroma, 2177 melanoma, 18618 nevus y 157 de lesión vascular.

La propuesta de tesis de este trabajo fue probada con la base de datos de ISIC 2019 (Tschandl *et al.*, 2018a; Codella *et al.*, 2018a), la cual consta de 25331 imágenes de ocho diferentes clases, además incluye etiquetas de identificación de las lesiones y metadatos, como la edad, el sitio anatómico y el sexo del paciente. Las ocho clases de las lesiones son queratosis actínica (AK, por sus siglas en inglés), carcinoma de células basales (BCC, por sus siglas en inglés), carcinoma de células escamosas (SCC, por sus siglas en inglés), queratosis benigna (BKL, por sus siglas en inglés, conformada por léntigo solar, queratosis seborreica y queratosis similar al liquen plano), melanoma (MEL), nevus (NEV), dermatofibroma (DER), y lesiones vasculares (VASC, por sus siglas en inglés). La Tabla I muestra la distribución de imágenes por clase en la base de datos. Como puede observarse, la cantidad de imágenes para las clases AK, SCC, DER y VASC siguen siendo muy pocas en comparación con las imágenes de lesiones cancerígenas.

Nombre de clase	Abreviación	Imágenes
Queratosis actínica	AK	867
Carcinoma de células basales	BCC	3323
Carcinoma de células escamosas	SCC	628
Queratosis benigna	BKL	2624
Melanoma	MEL	4522
Nevus	NEV	12,875
Dermatofibroma	DER	239
Lesión vascular	VASC	253
Total		25,331

Tabla I: Imágenes por clase en ISIC 2019.

V.2. El ensamblado de clasificadores

Como la base de datos ISIC 2019 tiene un desbalance considerable entre las clases mayoritarias y las minoritarias. En la literatura es posible encontrar diferentes estrategias para el manejo del desbalance de clases. De acuerdo con Gao y colaboradores (Gao et al., 2020) las técnicas de balanceo de datos se pueden catalogar básicamente usando: métodos basados en nivel de datos, métodos a nivel algoritmo y métodos de combinación de datos con mejoras de algoritmo. Dentro de los métodos basados en nivel de datos, el sub-muestreo aleatorio se encarga de balancear la distribución de los datos eliminando instancias de las clases mayoritarias, pero tiene la particularidad de poder perder información valiosa. Otra estrategia a nivel de datos, es el sobre-muestreo aleatorio, el cual consiste en generar aleatoriamente instancias de las clases minoritarias para balancear los datos, sin embargo tiende a sobreajustar los modelos. En cuanto a las técnicas de balanceo de datos a nivel algoritmo, las técnicas de ensamble pueden ayudar a aumentar el nivel de efectividad en la clasificación combinando información de diversos clasificadores, donde entre este tipo de modelos se encuentran el bagging (Breiman, 1996) y boosting (Freund y Shapire, 1996). De acuerdo con Galar y colaboradores (Galar et al., 2012), la idea básica de un método de ensamble es construir diferentes clasificadores y utilizar sus predicciones para construir nuevas instancias. Esto se asemeja al comportamiento humano de observar diferentes opciones para realizar la toma de decisiones. El uso de modelos de ensamble ayuda a las habilidades de generalización de un modelo, lo cual propicia la predicción de nuevas instancias. Según Hosni y colaboradores (Hosni *et al.*, 2019), el uso de modelos de ensamble solventa las debilidades de los modelos individuales de clasificación y consolida sus fortalezas para realizar tareas de predicción. Trabajos recientes como los de Harangi (Harangi, 2018) y Mahbod (Mahbod *et al.*, 2020) han probado la efectividad del uso de modelos basados en ensamble para la clasificación de lesiones de la piel. Considerando el gran desbalance de datos de la base de datos ISIC 2019 y que la queratosis actínica pertenece a las clases minoritarias, no es posible realizar una clasificación binaria, ya que se tendrían 867 imágenes en la clase de la queratosis actínica y 24464 imágenes en la clase imágenes que no son queratosis actínica. Por tal motivo, en esta propuesta de tesis se optó por realizar una clasificación multiclases con el uso de un modelo de ensamble basado en votación por mayoría en la salida de los tres clasificadores KNN-5, SVM-L y SVM-G.

V.2.1. Diagnóstico asistido por computadora para clasificar ocho tipos de lesiones de la piel

El primer paso de la metodología para el diagnóstico asistido por computadora es representar a la imagen en el espacio de color RGB, Figura 13(a). Posteriormente, se extraen tanto las características fractales como las características de la DenseNet-201, Figura 13(b), y se concatenan para formar el descriptor 1D de la imagen, Figura 13(c). El descriptor se clasifica en los espacios KNN-5, SVM-L y SVM-G, Figura 13(d). Los tres resultados se introducen en el clasificador de ensamble, Figura 13(e) y la respuesta será la clasificación de la imagen, Figura 13(f).



Figura 13: Metodología propuesta. (a) Imagen RGB de entrada. (b) Extracción de características fractales y de DenseNet-201. (c) Descriptor 1D. (d) Clasificadores SVM-L, SVM-G y KNN-5. (e) Clasificador de ensamble. (f) Predicción final.

V.3. Métricas de evaluación

Los resultados del diagnóstico asistido por computadora tienen cuatro posibilidades: verdaderos positivos, denotados como TP; falsos positivos, nombrados por FP; falsos negativos, llamados FN; y verdaderos negativos, indicados como TN. Utilizando esos valores, se usaron cuatro métricas para la evaluación del rendimiento de la clasificación, las cuales son (Metz, 1978)

$$exactitud = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%,$$
 (40)

$$precisión = \frac{TP}{TP + FP} \times 100\%, \tag{41}$$

$$sensibilidad = \frac{TP}{TP + FN} \times 100\%, \tag{42}$$

$$especificidad = \frac{TN}{TN + FP} \times 100\%.$$
 (43)

La exactitud es la proporción del total de casos clasificados correctamente como positivos (TP+TN) entre el total de posibilidades que se pueden presentar (TP+TN+FP+FN); es decir, una medida del sesgo en la clasificación. La precisión es la probabilidad condicional que mide la correcta clasificación (TP) entre el número de clasificaciones indicados como positivos (TP + FP), esta es una medida de dispersión. La sensibilidad es la probabilidad condicional que mide la proporción de clasificaciones correctas (TP)entre todos los casos que deberían haber sido clasificados como correctos (TP + FN). La especificidad es la probabilidad condicional que mide la proporción de clasificación de verdaderos negativos (TN) entre el número de casos totales que fueron clasificados como negativos (TN + FP) (Metz, 1978).

Capítulo VI. Resultados

VI.1. Experimentos

Para realizar la validación del sistema de diagnóstico asistido por computadora propuesto, se realizaron cuatro experimentos con la base de datos ISIC 2019. Para la implementación de la metodología se utilizó Matlab 2019b en una laptop HP con procesador Intel i5 con 12GB de memoria RAM. Todos los experimentos fueron realizados variando el parámetro $\alpha = -2$ a 2 de los descriptores fractales, con paso de 0.1. En el primer experimento, Exp-1, las 25331 imágenes se usaron para evaluar y entrenar el modelo. En el segundo experimento, Exp-2, se seleccionaron solamente imágenes de las seis principales categorías más representadas en la base de datos. Para el tercer y cuarto experimento, denotados como Exp-3 y Exp-4, la base de datos se dividió en el 70 % para la parte de entrenamiento y el 30 % restante para la parte de prueba. Como la división es aleatoria, tanto en el Exp-3 y Exp-4 se realizaron cinco repeticiones, empleando particiones aleatorias.

Para el Exp-1, los mejores resultados fueron obtenidos con $\alpha = -2.0$, como exponente en los descriptores fractales. En la Tabla II se muestra la matriz de confusión. La base de datos ISIC 2019 tiene 867 imágenes de AK, 3323 BCC, 628 SCC, 2624 BKL, 4522 MEL, 12875 NEV, 239 DER, 253 VASC. La respuesta del modelo fue de 802 AK, 3956 BCC, 295 SCC, 2285 BKL, 3746 MEL, 14,105 NEV, 48 DER y 94 VASC. Como se esperaba, las clases con menor cantidad de imágenes o con menor representación en la base de datos fueron erróneamente clasificadas, como SCC, DER, y VASC. Siempre debemos tener en cuenta, que como estamos trabajando en un modelo que es supervisado, debemos tener un conjunto de datos representativo D_C para las clases C, que nos permita reproducir o generar el espacio para esa clase. Para D_C , todas las imágenes C deben reescribirse como una combinación lineal de los elementos en D_C , lo cual no aplica para las clases SCC, DER, y VASC. Con base en los valores de la matriz de confusión de la Tabla II, se calcularon los valores de TP, FP, FN y TN, como se muestra en la Tabla III. Estos valores se utilizan en las Ecuaciones (40) a (43) para calcular el rendimiento del sistema del Exp-1. La Tabla IV muestra, por clase, el porcentaje de las métricas: exactitud, precisión, sensibilidad y especificidad. Como se esperaba, la sensibilidad fue muy baja debido a la poca representación en cantidad de imágenes de las clases SCC, DER, VASC. Para el Exp-1, el diagnóstico asistido por computadora arrojó una media \pm desviación estándar de *exactitud* = 97.35 \pm 2.04 %, *precisión* = 91.61 \pm 6.89 %, *sensibilidad* = 66.45 \pm 29.09 %, y *especificidad* = 97.85 \pm 3.98 %.

		Datos de entrenamiento						
	AK	BCC	SCC	BKL	MEL	NEV	DER	VASC
	867	3323	628	2624	4522	12,875	239	253
AK	662	91	3	16	5	90	0	0
BCC	22	3203	0	13	14	70	1	0
SCC	34	224	276	11	2	81	0	0
BKL	27	144	4	2105	39	305	0	0
MEL	37	111	5	85	3597	687	0	0
NEV	14	96	1	40	65	12,659	0	0
DER	3	54	3	10	14	108	47	0
VASC	3	33	3	5	10	105	0	94
Respuesta	802	3956	295	2285	3746	14,105	48	94
Exp-1								

Tabla II: Matriz de confusión del Exp-1 con $\alpha = -2.0$. Utilizando las 25331 imágenes.

Clase	Datos entrenamiento	Respuesta Exp-1	TP	FP	FN	TN
AK	867	802	662	140	205	24,324
BCC	3323	3956	3203	753	120	21,255
SCC	628	295	276	19	352	24,684
BKL	2624	2285	2105	180	519	22,527
MEL	4522	3746	3597	149	925	20,660
NEV	12,875	14,105	12,659	1446	216	11,010
DER	239	48	47	1	192	25,091
VASC	253	94	94	0	159	25,078

Tabla III: Rendimiento del Exp-1 obtenido de la matriz de la Tabla II.

Tabla IV: Métricas del rendimiento de la Tabla III para el Exp-1 con $\alpha = -2.0$.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.64	82.54	76.36	99.43
BCC	96.55	80.97	96.39	96.58
SCC	98.54	93.56	43.95	99.92
BKL	97.24	92.12	80.22	99.21
MEL	95.76	96.02	79.54	99.28
NEV	93.44	89.75	98.32	88.39
DER	99.24	97.92	19.67	100.00
VASC	99.37	100.00	37.15	100.0
$\textbf{Media} \pm \textbf{DE}$	$\textbf{97.35} \pm \textbf{2.04}$	$\textbf{91.61} \pm \textbf{6.89}$	$\textbf{66.45} \pm \textbf{29.09}$	$\textbf{97.85} \pm \textbf{3.98}$

Para el Exp-2, la base de datos es de 24839 imágenes: 867 AK, 3323 BCC, 628 SCC, 2624 BKL, 4522 MEL, y 12875 imágenes de NEV. Las 239 imágenes de DER y las 253 imágenes de VASC no son consideradas para este experimento. Los mejores resultados para el modelo se obtuvieron con $\alpha = -1.9$. Para este experimento, se tiene *exactitud* = 96.85 ± 1.71%, *precisión* = 90.12 ± 6.07%, *sensibilidad* = 79.25 ± 19.06%, y *especificidad* = 97.42 ± 3.93%, tal como se muestra en la Tabla V. Como se pue-de observar, al no considerar las imágenes menos representadas en la base de datos, la sensibilidad del modelo aumenta, mientras que la desviación estándar disminuye. Lo cual, hace visible la importancia de tener clases de imágenes más representadas en la base de datos.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.62	82.68	76.47	99.42
BCC	96.78	82.63	96.21	96.88
SCC	98.55	94.98	45.22	99.94
BKL	97.22	92.78	79.88	99.27
MEL	95.73	96.53	79.39	99.27
NEV	94.16	91.12	98.31	89.69
$\textbf{Media} \pm \textbf{DE}$	$\textbf{96.85} \pm \textbf{1.71}$	$\textbf{90.12} \pm \textbf{6.07}$	$\textbf{79.25} \pm \textbf{19.06}$	$\textbf{97.42} \pm \textbf{3.93}$

Tabla V: Métricas de rendimiento obtenidas para el Exp-2 con $\alpha=-1.9.$ La base de datos es de 24839 imágenes.

Para el Exp-3, la base de datos se dividió aleatoriamente en 70 % para entrenamiento (17732 imágenes) y 30 % para prueba (7599 imágenes). El mejor rendimiento se obtuvo con $\alpha = 0.4$. La media \pm una desviación estándar (DE) de cinco repeticiones del experimento se muestra en la Tabla VI. Para el Exp-3, en la evaluación de la partición de entrenamiento se obtiene una *exactitud* = 97.31 \pm 2.14 %, *precisión* = 91.90 \pm 6.28 %, *sensibilidad* = 64.39 \pm 32.93 %, *especificidad* = 97.80 \pm 4.15 %. Para la partición de prueba, el modelo logró una *exactitud* = 92.20 \pm 6.74 %, *precisión* = 63.79 \pm 21.56 %, *sensibilidad* = 38.36 \pm 31.20 %, y una *especificidad* = 93.60 \pm 12.22 %, Tabla VII.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.87 ± 0.07	85.78 ± 0.81	$\textbf{80.21} \pm \textbf{1.80}$	99.54 ± 0.02
BCC	96.56 ± 0.13	81.00 ± 0.44	96.62 ± 0.29	96.55 ± 0.12
SCC	98.45 ± 0.05	$\textbf{92.94} \pm \textbf{1.35}$	$\textbf{40.43} \pm \textbf{2.05}$	99.92 ± 0.02
BKL	97.36 ± 0.07	$\textbf{92.92}\pm\textbf{0.49}$	80.72 ± 0.65	99.29 ± 0.05
MEL	95.58 ± 0.10	95.54 ± 0.13	78.92 ± 0.60	99.20 ± 0.02
NEV	$\textbf{93.14} \pm \textbf{0.08}$	89.34 ± 0.09	98.21 ± 0.12	87.91 ± 0.07
DER	99.16 ± 0.03	98.00 ± 2.75	10.87 ± 1.67	100.00 ± 0.00
VASC	99.33 ± 0.01	99.69 ± 0.70	$\textbf{29.16} \pm \textbf{385}$	100.00 ± 0.00
$\textbf{Mean} \pm \textbf{DE}$	$\textbf{97.31} \pm \textbf{2.14}$	$\textbf{91.90} \pm \textbf{6.28}$	$\textbf{64.39} \pm \textbf{32.93}$	$\textbf{97.80} \pm \textbf{4.15}$

Tabla VI: Media \pm DE de cinco repeticiones del Exp-3 con $\alpha=0.4.$ Utilizando datos de entrenamiento de 17732 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.61 ± 0.26	$\textbf{38.81} \pm \textbf{2.15}$	$\textbf{47.56} \pm \textbf{2.79}$	$\textbf{97.33} \pm \textbf{0.13}$
BCC	89.48 ± 0.28	$\textbf{57.72} \pm \textbf{1.29}$	69.07 ± 0.63	92.51 ± 0.28
SCC	$\textbf{97.19} \pm \textbf{0.13}$	$\textbf{34.48} \pm \textbf{4.10}$	$\textbf{13.84} \pm \textbf{1.19}$	99.32 ± 0.11
BKL	90.30 ± 0.21	55.55 ± 3.59	$\textbf{28.97} \pm \textbf{1.90}$	97.34 ± 0.29
MEL	$\textbf{86.85} \pm \textbf{0.18}$	$\textbf{81.83} \pm \textbf{2.11}$	$\textbf{33.84} \pm \textbf{0.59}$	98.37 ± 0.19
NEV	80.06 ± 0.55	$\textbf{73.36} \pm \textbf{0.65}$	95.58 ± 0.39	63.95 ± 0.86
DER	99.07 ± 0.07	$\textbf{70.00} \pm \textbf{44.72}$	$\textbf{2.55} \pm \textbf{2.40}$	99.99 ± 0.006
VASC	99.06 ± 0.07	$\textbf{98.57} \pm \textbf{3.14}$	15.49 ± 1.70	99.99 ± 0.01
$\textbf{Media} \pm \textbf{DE}$	$\textbf{92.20} \pm \textbf{6.74}$	$\textbf{63.79} \pm \textbf{21.56}$	$\textbf{38.36} \pm \textbf{31.21}$	$\textbf{93.60} \pm \textbf{12.22}$

Tabla VII: Media \pm DE de cinco repeticiones del Exp-3 para los datos de prueba con $\alpha = 0.4$. Los datos de prueba contienen 7599 imágenes.

Por último, en el Exp-4 no se consideran las clases con menor cantidad de imágenes DER y VASC, por lo que la base de datos consta de 24839 imágenes, la cual fue aleatoriamente dividida en 70 % para entrenamiento (17387 imágenes) y el 30 % restante para la prueba (7452 imágenes). Los mejores resultados en este experimento se obtuvieron con $\alpha = 1.6$. Se realizaron cinco repeticiones del experimento y al evaluar el modelo desarrollado en la partición de entrenamiento se obtuvo una $exactitud = 96.82 \pm 1.85\%$, $precisión = 90.57 \pm 5.19\%$, $sensibilidad = 79.35 \pm 20.24\%$, y una $especificidad = 97.39 \pm 10\%$ 4.06%, como se muestra en la Tabla VIII. Para la partición de prueba se obtuvo una $exactitud = 89.95 \pm 6.05\%$, precisión = $58.62 \pm 17.54\%$, sensibilidad = $47.89 \pm 29.30\%$, y una $especificidad = 91.60 \pm 13.62\%$, como se muestra en la Tabla IX. En este grupo de experimentos (Exp-4), podemos observar que las métricas en la partición de entrenamiento, al no considerar las clases con pocas imágenes, los valores de la exactitud, precisión, y especificidad permanecen con valores muy similares a las del Exp-3, pero incrementa la *sensibilidad*. En el Apéndice A y Apéndice B, se muestran los valores que cada una de las métricas obtuvieron para las cinco repeticiones realizadas para el Exp-3 y el Exp-4.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	$\textbf{98.87} \pm \textbf{0.08}$	$\textbf{85.88} \pm \textbf{1.56}$	80.52 ± 1.21	99.52 ± 0.07
BCC	$\textbf{96.83} \pm \textbf{0.04}$	$\textbf{82.93}\pm\textbf{0.14}$	$\textbf{96.29} \pm \textbf{0.23}$	$\textbf{96.91} \pm \textbf{0.05}$
SCC	98.48 ± 0.05	94.68 ± 1.86	$\textbf{41.87} \pm \textbf{1.05}$	99.93 ± 0.02
BKL	$\textbf{97.28} \pm \textbf{0.04}$	$\textbf{92.88} \pm \textbf{0.28}$	80.55 ± 0.53	99.27 ± 0.03
MEL	95.55 ± 0.07	96.21 ± 0.30	$\textbf{78.69} \pm \textbf{0.67}$	99.31 ± 0.06
NEV	$\textbf{93.92}\pm\textbf{0.19}$	$\textbf{90.83} \pm \textbf{0.29}$	98.14 ± 0.09	89.39 ± 0.37
$\textbf{Media} \pm \textbf{DE}$	$\textbf{96.82} \pm \textbf{1.85}$	$\textbf{90.57} \pm \textbf{5.19}$	$\textbf{79.35} \pm \textbf{20.24}$	$\textbf{97.39} \pm \textbf{4.06}$

Tabla VIII: Media \pm DE de cinco repeticiones del Exp-4 para los datos de entrenamiento con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.

Tabla IX: Media \pm DE de cinco repeticiones del Exp-4 para los datos de prueba con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.38 ± 0.30	$\textbf{38.25} \pm \textbf{3.57}$	$\textbf{45.15} \pm \textbf{1.91}$	97.26 ± 0.40
BCC	89.66 ± 0.21	59.15 ± 0.30	68.38 ± 0.83	$\textbf{92.87} \pm \textbf{0.14}$
SCC	$\textbf{97.23} \pm \textbf{0.23}$	$\textbf{40.12} \pm \textbf{6.49}$	$\textbf{16.26} \pm \textbf{3.14}$	99.36 ± 0.11
BKL	$\textbf{90.39} \pm \textbf{0.16}$	58.13 ± 1.94	$\textbf{29.27} \pm \textbf{0.80}$	97.53 ± 0.20
MEL	$\textbf{86.43} \pm \textbf{0.28}$	$\textbf{81.66} \pm \textbf{1.43}$	$\textbf{32.62} \pm \textbf{1.15}$	98.37 ± 0.14
NEV	80.59 ± 0.45	74.40 ± 0.40	95.68 ± 0.28	64.17 ± 0.83
$\textbf{Media} \pm \textbf{SD}$	$\textbf{89.95} \pm \textbf{6.05}$	$\textbf{58.62} \pm \textbf{17.54}$	$\textbf{47.89} \pm \textbf{29.30}$	$\textbf{91.60} \pm \textbf{13.62}$

El objetivo planteado para el desarrollo de esta tesis fue diseñar una metodología de reconocimiento de patrones en imágenes digitales a color para la clasificación de manchas de la piel de queratosis actínica. El objetivo se logró, ya que la metodología de diagnóstico asistido por computadora desarrollada clasifica simultáneamente ocho clases de lesiones de la piel. Uno de los grandes retos al trabajar con imágenes de lesiones de la piel, es el desbalance que presentan las clases en las bases de datos. Existen dos maneras de trabajar con esta problemática, la primera es balancear las clases, es decir, agregar imágenes generadas por computadora y eliminar imágenes que pudieran ser relevantes en la extracción de características, esto no es viable para la queratosis actínica, ya que es una clase minoritaria. Por lo que, en este trabajo se optó por el segundo caso, que es el ensamblado de clasificadores, esta técnica es más complicada que la primera, pero es más aceptada en el área de la dermatología porque en las bases de datos no se eliminan imágenes, ni se agregan imágenes generadas por computadora.

Capítulo VII. Comparativa con otras metodologías

En este capítulo se revisarán algunos de los trabajos relevantes en el estado del arte en la clasificación de imágenes dermatoscópicas de lesiones de la piel. El primero de ellos es el propuesto por Garnavi y colaboradores (Garnavi *et al.*, 2012) quienes desarrollaron un sistema de diagnóstico asistido por computadora para la clasificación binaria de melanoma. Trabajaron con descomposición de ondículas y características geométricas de la lesión. Para la clasificación emplearon los clasificadores SVM, Bosque Aleatorio, Árbol de Modelo Logístico y Naive Bayes. Para el entrenamiento del modelo usaron 40 imágenes dermatoscópicas de melanoma y 59 de lesiones catalogadas como benignas. Para la parte de validación seleccionaron subconjuntos de características y diferentes configuraciones de umbrales y nodos en los clasificadores de 30 imágenes de melanoma y 57 imágenes de lesiones benignas, este paso lo repitieron hasta encontrar la configuración óptima. Para esta etapa, el modelo obtuvo una *exactitud* = 87.36 % y área bajo la curva (AUC, por sus siglas en inglés) AUC de 0.913. Finalmente, la etapa de la prueba constó de 44 imágenes de melanoma y 59 de lesiones benignas, para ella, reportando *exactitud* = 91.26 %, y *AUC* = 0.937.

Barata y colaboradores (Barata *et al.*, 2014) construyeron dos sistemas de diagnóstico asistido por computadora de imágenes dermatoscópicas para la clasificación binaria de melanoma: uno usa características globales y el otro características locales. Para la clasificación aprovecharon las ventajas de los clasificadores KNN, AdaBoost, SVM-RBF, y las características extraídas de bolsa de palabras (BoW, por sus siglas en inglés). La base de datos consistió de 176 imágenes dermatoscópicas: 25 MEL y 151 NEV. Para manejar el desbalance de las clases, realizaron copias de los descriptores de melanoma agregando ruido gaussiano con $\sigma = 0.0001$ a cada copia del descriptor. El método global mostró una *sensibilidad* = 96 % y *especificidad* = 80 %. El método local obtuvo una *sensibilidad* = 100 % y una *especificidad* = 75 %. Shimizu y colaboradores (Shimizu *et al.*, 2015) usaron clasificadores lineales y una estrategia binaria, la cual consistió en la clasificación de cada clase contra todas las demás, extrayendo una a una las clases dentro del grupo de las demás clases combinadas. La base de datos empleada contenía 968 imágenes dermatoscópicas, empleando la validación cruzada reportaron una tasa de detección de 90.48 % para melanoma, 82.51 % para NEV, 82.61 % para BCC, y un 80.61 % para queratosis seborreica (SK).

Wahba y colaboradores (Wahba *et al.*, 2018) propusieron un método utilizando diferencias de niveles de gris en conjunto con características ABCDE y el espacio de clasificación SVM. Emplearon la base de datos ISIC 2018 de la cual seleccionaron solamente 300 imágenes por clase de MEL, NEV, BCC y BKL. Al emplear la validación cruzada los resultados reportados fueron resultados de *exactitud* = 100 %, *sensibilidad* = 100 %, *especificidad* = 100 % y una *precisión* = 100 %. Sin embargo, esta estrategia solamente clasifica cuatro clases de lesiones, las clases están balanceadas y no dicen cómo seleccionaron las 300 imágenes de cada una de las clases, además de que la metodología depende de la correcta segmentación de las imágenes para lograr resultados satisfactorios.

Khan y colaboradores (Khan *et al.*, 2019) aplicaron la matriz de co-ocurrencia de niveles de gris (GLCM, por sus siglas en inglés), los patrones binarios locales (LBP, por sus siglas en inglés) y características extraídas del color. La clasificación binaria de 146 imágenes de melanoma y 251 de nevus fue realizada con los clasificadores SVM, KNN, Naive Bayes, y árboles de decisión, consiguiendo una *exactitud* = 96 %, *sensibilidad* = 97 %, *especificidad* = 96 %, y una *precisión* = 97 %. Albahar y colaboradores (Albahar, 2019) emplearon una red neuronal convolucional (CNN) para clasificar de forma binaria imágenes dermatoscópicas de MEL, NEV, SK, SCC, BCC, y léntigo. La base de datos fue dividida en tres partes iguales de 8000 imágenes en categoría maligna y benigna. En la parte de entrenamiento usaron 5600 imágenes, reportaron una *exactitud* = 97.49 %. Para la parte de la prueba lograron un AUC = 0.98, *exactitud* = 97.49 %, *especificidad* = 93.6 %, y una *sensibilidad* = 94.3 %.

Marka y colaboradores (Marka *et al.*, 2019) presentaron una revisión exhaustiva del estado del arte en la detección automatizada de cáncer de piel del tipo no melanoma, analizaron las técnicas de diagnósticos asistidos por computadora para imágenes dermatoscópicas y no dermatoscópicas.

Wu y colaboradores (Wu *et al.*, 2019) entrenaron cinco redes CNNs con imágenes de queratosis seborreica (SK, por sus siglas en inglés), AK, rosácea (ROS), lupus eritematoso, (LE, por sus siglas en inglés), BCC, y SCC. En su mejor configuración, declaran en la prueba valores de sensibilidad de 92.9 %, 89.2 %, y 84.3 % para las clases de LE, BCC, y SK, respectivamente.

Gessert y colaboradores (Gessert *et al.*, 2020) propusieron un modelo de ensamble de diferentes redes CNNs: SENet154 (Hu *et al.*, 2018), ResNext (Mahajan, 2018) y EfficientNet-B0 a B7 (Tan y Le, 2019), para clasificar imágenes dermatoscópicas en la base de datos ISIC 2019. La metodología requiere procesar las imágenes en forma binaria y escalar el tamaño de las imágenes de mayor dimensión a un máximo de 600 pixeles por lado, conservando la proporción de aspecto de las imágenes. Adicionalmente, para solventar el desbalance de clases, se valieron del aumento artificial de los datos realizando transformaciones geométricas aleatoriamente. Informan una *sensibilidad* = 72.5%, mientras que al utilizar los metadatos, la sensibilidad incrementó a 74.2 %, esto haciendo uso de la totalidad de imágenes de la base de datos ISIC 2019. Mientras que al usar la prueba oficial de ISIC 2019 obtienen *sensibilidad* = 63.6 % y una *sensibilidad* = 63.4 % al utilizar metadatos.

Recientemente, Bajwa y colaboradores (Bajwa *et al.*, 2020) desarrollaron un sistema de diagnóstico asisitido por computadora para el ensamblado de cuatro redes CNNs: ResNet-152, DenseNet-161, SE-ResNeXt-101 y NASNet. Los experimentos fueron realizados en la base de datos ISIC 2018. Para la etapa de entrenamiento, las imágenes fueron sometidas a diferentes transformaciones aleatorias, como recorte y reflexión horizontal. También utilizaron una validación cruzada estratificada de *k* pliegues con k = 5. Sólo informan resultados para la prueba, los cuales son *precisión* = 85.02 ± 09.10 %, *sensibilidad* = 80.46 ± 09.38 %, *especificidad* = 96.57 ± 07.15 % y una métrica F (media harmónica de precisión y sensibilidad (Chinchor, 1992)), $F = 82.45 \pm 08.38$.

Capítulo VIII. Discusiones

El diagnóstico de una lesión de la piel por un dermatólogo sigue siendo subjetivo, el rango de exactitud medido en centros dermatológicos especializados es de 64 % a 80 % (MacKenzie-Wood *et al.*, 1998; Miller, 1992; Lindelöf y Hedblad, 1994; Grin *et al.*, 1990). Hoy en día, muchos de los diagnósticos son realizados por asistentes médicos, los cuales obtienen una menor exactitud que especialistas dermatólogos (Anderson *et al.*, 2018).

En este trabajo de tesis se desarrolló una metodología de diagnóstico asistido por computadora para clasificar simultáneamente ocho clases de imágenes dermatoscópicas a color de la base de datos ISIC 2019. Las clases son la queratosis actínica, carcinoma de células basales, carcinoma de células escamosas, queratosis benigna, melanoma, nevus, dermatofibroma y lesión vascular. Dado que las imágenes de las diferentes clases con las que cuenta la base de datos no son todas del mismo tamaño (por lo que los descriptores fractales son de diferente tamaño), se utilizó la energía, la varianza y la entropía de los descriptores fractales por cada canal de color RGB.

En total se emplearon nueve características de los descriptores fractales, las cuales son concatenadas con las mil características extraídas de la última capa completamente conectada de la red DenseNet-201. Para realizar la clasificación se usaron tres espacios para construir un ensamble de clasificadores basado en votación por mayoría: el K-NN, SVM-L y SVM-G. El trabajar con una metodología híbrida combinando los descriptores fractales por cada canal de color con las características obtenidas por la red DenseNet-201, permite aumentar la efectividad en la clasificación. Dado que no necesitamos realizar una segmentación, contamos con mayor información al trabajar con los descriptores fractales y además, no tenemos que reducir las imágenes, a diferencia de lo que sucede en las redes CNNs, por ejemplo la red DenseNet-201 requiere imágenes de tamaño 224×224

pixeles.

La metodología propuesta, presenta excelentes resultados, tal como los obtenidos en trabajos como Gessert y colaboradores (Gessert *et al.*, 2020), y Bajwa y colaboradores (Bajwa *et al.*, 2020). A diferencia de las metodologías de Gessert y colaboradores, y Bajwa y colaboradores, en este trabajo de tesis, no se requiere la generación y manipulación de imágenes mediante la computadora. Dado que existen varias alternativas de código abierto para la implementación de las CNNs, la metodología propuesta no necesita de equipo de cómputo de alto rendimiento, es por esto que la hace viable para reproducir los mismos resultados con diferentes lenguajes de programación en varias plataformas.

Es bien conocido que una red neuronal se entrena con una amplia base de datos para obtener un excelente desempeño, ya que se basa en la convolución de imágenes con varios filtros para extraer todas las formas posibles en las que se puede presentar la imagen. Esto implica que el costo computacional de extraer todas esas características sea alto, ya que la operación de convolución de imágenes no es invariante a transformaciones geométricas, tales como escala, rotación y traslación. Además, para segmentar adecuadamente imágenes con bordes de bajo contraste y vello capilar, como los de lesiones cutáneas, se requieren redes neuronales especializadas en el tema (Al-masni *et al.*, 2018). Debido a estas razones, una opción viable para la clasificación de imágenes de lesiones de la piel y a su vez reducir el costo computacional del ensamblado de varias redes neuronales convolucionales, es trabajar con una metodología híbrida que combina características extraídas con descriptores fractales y CNNs, como la presentada en esta propuesta de tesis.

58

Capítulo IX. Conclusiones

La metodología propuesta en este trabajo de tesis para la clasificación de imágenes dermatoscópicas de lesiones de la piel basada en características fractales y de la red CNN cumple el objetivo planteado, los objetivos específicos y responde a cada una de las preguntas de investigación.

La primera parte de este trabajo consiste en la extracción de características fractales por medio de pirámides de base cuadrada. La metodología propuesta emplea tres descriptores fractales, un descriptor por cada canal de color en el espacio RGB. La energía, varianza y entropía de cada descriptor fractal por canal RGB son concatenadas con las características obtenidas de la red CNN DenseNet-201. El uso de la textura de las imágenes vía análisis fractal y redes CNNs permitió la obtención de características confiables para lograr un método con un nivel de exactitud elevado en la clasificación. El uso de descriptores fractales basados en pirámides de base cuadrada proporcionó un análisis completo de las imágenes, tanto global como local al analizar pixel a pixel la imagen, usando diferentes tamaños de pirámides. Esto facilitó que la totalidad de la información contenida en las imágenes fuera tomada en cuenta y no hubiera pérdida de información. Dado que no siempre es posible obtener una correcta segmentación de lesiones de la piel, debido a su bajo contraste y bordes difusos, el uso de descriptores fractales permitió extraer características representativas de las imágenes para su clasificación, sin necesidad de pre-procesamiento o segmentación.

Es común que en las bases de datos de lesiones de la piel se tengan clases desbalanceadas, lo cual implica un reto para la clasificación. En la actualidad existen alternativas para solucionar el desbalance de clases, una de ellas es la generación de imágenes por computadora, sin embargo ésta no es una práctica válida para el área de la dermatología. Una solución más compleja, pero significativamente más eficiente y válida es el uso de ensamblado de clasificadores, la cual no necesita la generación de imágenes adicionales. Por esta razón, en este trabajo se optó por clasificar a las imágenes usando el ensamblado de los clasificadores KNN-5, SVM-L y SVM-G. Este ensamblado de clasificadores permitió combinar las predicciones de tres de los espacios clasificadores más efectivos para formar una clasificación más robusta. Esto ayudó a generar un modelo para la clasificación sin agregar tiempo adicional en segmentar, preprocesar o utilizar múltiples redes neuronales convolucionales para aumentar la efectividad.

La metodología de diagnóstico asisitido por computadora propuesta en este trabajo de tesis permitió una clasificación efectiva para imágenes dermatoscópicas de lesiones de la piel en la base de datos ISIC 2019. La metodología propuesta, clasifica simultáneamente las ocho clases de lesiones de la base de datos ISIC 2019: queratosis actínica (AK), carcinoma de células basales (BCC), carcinoma de células escamosas (SCC), queratosis benigna (BKL), melanoma (MEL), nevus (NEV), dermatofibroma (DER) y lesiones vasculares (VASC), con *exactitud* = 97.35%, *precisión* = 91.61%, *sensibilidad* = 66.45% y *especificidad* = 97.85%. Una metodología eficiente de diagnóstico asistido por computadora para la clasificación de imágenes dermatoscópicas de lesiones de la piel, como la propuesta en este trabajo de tesis, puede ser una herramienta valiosa que ayude a los dermatólogos en la toma de decisiones acerca de las acciones necesarias para agilizar los diagnósticos clínicos y con ellos tomar las medidas necesarias para mejorar la salud de sus pacientes.

Capítulo X. Trabajo a futuro

En este trabajo de tesis se propuso una metodología híbrida que combina información de descriptores fractales de pirámides de base cuadrada con las características obtenidas de la red CNN DenseNet-201. Esta metodología permitió lograr un excelente rendimiento en la base de datos ISIC 2019 sin necesidad de realizar preprocesamiento, segmentación y aumento artificial de los datos. Aún con los resultados obtenidos existen posibles variantes que pueden realizarse para mejorar el desempeño de la metodología propuesta. El primero de ellos es considerar nuevos espacios de color. Otra de las posibles variantes es considerar cada descriptor fractal por banda como un histograma y extraer nuevos atributos. En lo referente a las redes CNN, como trabajo a futuro se estudiarán las características extraídas con el uso de redes más recientes como las EfficientNet-B0 a EfficientNet-B7. Por otra parte, para el paso del ensamblado se puede analizar el uso de Bagging y Boosting.
Bibliografía

- Al-antari, M., Al-Masni, M., Park, S., Park, J., Metwally, M., Kadah, Y., Han, S., y Kim, T. (2018). An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. *J. Med. Biol.*, **38**: 443–456.
- Al-antari, M., Han, S., y Kim, T. (2020). Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital x-ray mammograms. *Comput. Meth. Prog. Biomed.*, **196**.
- Al-masni, M., Al-antari, M., Choi, M., Han, S., y Kim, T. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Meth. Prog. Biomed.*, 162: 221–231.
- Albahar, M. A. (2019). Skin lesion classification using convolutional neural network with novel regularizer. *IEEE Access*, **7**: 38306–38313.
- Amadasun, M. y King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**(5): 1264–1274.
- Amelard, R., Glaister, J., Wong, A., y Clausi, D. A. (2015). High-level intuitive features (hlifs) for intuitive skin lesion description. *IEEE Transactions on Biomedical Engineering*, 62(3): 820–831.
- Anderson, A., Matsumoto, M., Saul, M., Secrest, A., y Ferris, L. (2018). Accuracy of skin cancer diagnosis by physician assistants compared with dermatologists in a large health care system. *JAMA Dermatol.*, **154**: 569–573.
- Arivazhagan, S., Ganesan, L., y Priyal, S. P. (2006). Texture classification using gabor wavelets based rotation invariant features. *Pattern Recognition Letters*, **27**(16): 1976– 1982.
- Ashour, M. W., Khalid, F., Halin, A. B., y Darwish, S. H. (2016). Multi-class support vector machines for texture classification using gray-level histogram and edge detection features. En: *Proceedings of 57th the IIER International Conference*. pp. 1–5.
- Backes, A. R., Casanova, D., y Bruno, M. O. (2009). Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, **23**: 1145–1160.
- Bajwa, M., Muta, K., Malik, M., Siddiqui, S., Braun, S., Homey, B., Dengel, A., y Ahmed, S. (2020). Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences*, **10**(7): 2488.
- Ballerini, L., Fisher, R., Aldridge, R., y Rees, J. (2013). A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. En: M. E. Celebi y G. Schaefer (eds.), *Color Medical Image Analysis*, Vol. 6. Lecture Notes in Computational Vision and Biomechanics.
- Barajas-Garcia, C., Solorza-Calderón, S., y Gutiérrez-López, E. (2019). Scale, translation and rotation invariant wavelet local feature descriptor. *Applied Mathematics and Computation*, **363**.

- Barata, C., Ruela, M., Francisco, M., Mendoza, T., y Marques, J. S. (2014). Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*, **8**(3): 965–979.
- Bay, H., Ess, A., Tuytelaars, T., y LucVan Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, **110**(3): 346–359.
- Benesty, J., Chen, J., Huang, Y., y Cohen, I. (2009). *Noise Reduction in Speech Processing*, capítulo Pearson Correlation Coefficient. Springer Topics in Signal Processing.
- Bermant, P., Bronstein, M. M., Wood, R. J., Gero, S., y Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific Reports*, **9**(1): 1–10.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24: 123–140.
- Buchholz, R. H. y Rathbun, R. L. (1998). Heron triangles and elliptic curves. *Bulletin of the Australian Mathematical Society*, **58**(3): 411–421.
- Celebi, M., Kingravi, H., Uddin, B., Iyatomi, H., Aslandogan, Y., Stoecker, W., y Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Comput Med Imaging Graph*, **6**(31): 362–373.
- Celebi, M., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., y Schaefer, G. (2015). *Dermos-copy Image Analysis*, capítulo A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images, pp. 97–129. CRC Press, Boca Raton, FL, USA.
- Chen, X. W., Zeng, X., y van Alphen, D. (2006). Multi-class feature selection for texture classification. *Pattern Recognition Letters*, **27**(14): 1685–1691.
- Chinchor, N. (1992). Muc-4 evaluation metrics. En: *Proceedings of the 4th Conference on Message Understanding*, USA. Association for Computational Linguistics, MUC4 '92, pp. 22–29.
- Chu, A., Sehgal, C., y Greenleaf, J. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, **11**(6): 415–419.
- Chu, K. (1999). An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emergency Medicine*, **11**: 175–181.
- Codella, N., Gutman, D., Celebi, M.E., H. B., Marchetti, M., Dusza, S., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., y Halpern, A. A. (2017). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi). Hosted by the International Skin Imaging Collaboration (ISIC).
- Codella, N., Gutman, D., Celebi, M., Helba, B., Marchetti, M., Dusza, S., Kalloo, A., Liopyris, K., Mishra, N., y Kittler, H. (2018a). Skin lesion analysis toward melanoma detection:
 A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). 4-7 de abril, Washington, DC, USA. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172.

- Codella, N., Rotemberg, V., Tschandl, P., Celebi, M., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., y Halpern, A. (2018b). Skin lesion analysis toward melanoma detection 2018. A Challenge Hosted by the International Skin Imaging Collaboration (ISIC).
- Cortes, C. y Vapnik, V. N. (1995). Support-vector networks. *Machine Learning*, **20**(3): 273–297.
- Cover, T. M. y Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, **13**(1): 21–27.
- Csurka, G., Dance, C., Fan, L., Willamowski, J., y Bray, C. (2004). Visual categorization with bags of keypoints. Proceedings of the InternationalWorkshop on Statistical Learning in Computer Vision, pp. 1–16.
- Dasarathy, B. V. y Holder, E. B. (1991). Image characterizations based on joint gray levelrun length distributions. *Pattern Recognition Letters*, **12**(8): 497–502.
- Dash, S. y Senapati, M. R. (2018). Gray level run length matrix based on various illumination normalization techniques for texture classification. *Evolutionary Intelligence*, pp. 1–10.
- Deep, R. (2006). Probability and Statistics. Academic Press.
- Depeursinge, A., Püspöki, Z., Ward, J. P., y Unser, M. (2017). Steerable wavelet machines (swm): Learning moving frames for texture classification. *IEEE Transactions on Image Processing*, **26**(4): 1626–1636.
- Di, W., Bhardwaj, A., y Wei, J. (2018). *Deep Learning Essentials*, capítulo 1, pp. 91–101. Packt Publishing.
- Dodson, J., DeSpain, J., Hewett, J., y Clark, D. (1991). Malignant potential of actinic keratoses and the controversy over treatment. *Arch Dermatol*, **127**.
- Emerson, C., Lam, N., y Quattrochi, D. (1999). Multi-scale fractal analysis of image texture and pattern. *Photogrammetric Engineering and Remote Sensing*, **65**: 51–61.
- Fitzpatrick, T. B. (1975). Soleil et peau. J Med Esthet, 2: 33-34.
- Florindo, J. B. y Bruno, M. O. (2019). Fractal descriptors of texture images based on the triangular prism dimension. *Journal of Mathematical Imaging and Vision*, **61**: 140–159.
- Florindo, J. B., Bruno, O. M., y Landini, G. (2017). Morphological classification of odontogenic keratocyst using bouligand-minkowski fractal descriptors. *Computer in Biology* and Medicine, 81: 1–10.
- Foody, G. M. (2009). Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment*, **113**: 1658–1663.
- Freund, Y. y Shapire, R. E. (1996). Experiments with a new boosting algorithm. En: M. Kaufmann (ed.), *Machine Learning: Proceedings of the Thirteenth International Conference*. pp. 148–156.

- Galar, M., Fernández, A., Barrenechea, E., Bustince, H., y Herrera, F. (2012). A review on ensembles for the class imbalance problem: Bagging, boosting, and hybrid-based approaches. *EEE T. Syst. Man Cybern.*, **42**: 463–484.
- Galloway, M. (1975). Texture analysis using gray level run lengths. *Computer Graphics* and *Image Processing*, **4**(2): 172–179.
- Gao, X., Ren, B., Zhang, H., Sun, B., Li, J., Xu, J., He, Y., y Li, K. (2020). An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling. *Expert Systems with Applications*, **160**: 113660.
- Garnavi, R., Aldeen, M., y Bailey, J. (2012). Computer-aided diganosis of melanoma using border- and wavelet-based texture analysis. *IEEE Transactions on Information Technology in Biomedicine*, **16**(6): 1239–1252.
- Garza-Flores, E., Guerra-Rosas, E., y Álvarez-Borrego, J. (2019). Spectral indexes obtained by implementation of the fractional fourier and hermite transform for the diagnosis of malignant melanoma. *Biomed. Opt. Express*, **10**: 6043–6056.
- Gessert, N., Nielsen, M., Shaikh, M., Werner, R., y Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. *Methods X*.
- Glorot, X., Bordes, A., y Bengio, Y. (2011). Deep sparse rectifier neural networks. *AIS*-*TATS*, **3**.
- Gonzalez, R. (2002). Digital image processing. Pearson Education, segunda edición.
- Gonzalez, R., Wood, R., y Eddins, S. (2010). *Digital Image Processing with MATLAB*, pp. 237–240, 253–259. Tata McGraw-Hill, New Delhi, India, segunda edición.
- Grin, C., Kopf, A., Welkovich, B., Bart, R., y Levenstein, M. (1990). Accuracy in the clinical diagnosis of malignant melanoma. *Arch. Dermatol.*, **126**(6): 763–766.
- Guerra-Rosas, E. y Álvarez-Borrego, J. (2015). Methodology for diagnosis of skin cancer on images of dermatologic spots by spectral analysis. *Biomed. Opt. Express*, **6**(10): 3876–3891.
- Guerra-Rosas, E., Álvarez-Borrego, J., y Angulo-Molina, A. (2017). Identification of melanoma cells: A method based in mean variance of signatures via spectral densities. *Biomed. Opt. Express*, **8**(4): 2185–2194.
- Guo, Z., Wang, X., Zhou, J., y You, J. (2016). Robust texture image representation by scale selective local binary patterns. *IEEE Transactions on Image Processing*, **25**(2): 687–699.
- Gutman, D., Codella, N. C. F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., y Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016. hosted by the International Skin Imaging Collaboration (ISIC).
- Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, **67**(5): 786–804.

- Haralick, R. M., Shanmugam, K., y Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-3**(6): 610–621.
- Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics*, **86**: 25–32.
- Hay, R., Johns, N., Williams, H., Bolliger, I., Dellavalle, R., Margolis, D., Marks, R., Naldi, L., Weinstock, M. A., Wulf, S. K., Michaud, C., Murray, C., y Naghavi, M. (2014). The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *Journal of Investigative Dermatology*, **134**(6): 1527–1534.
- Hosni, M., Abnane, I., Idri, A., Carrillo de Gea, J. M., y Fernández Alemán, J. L. (2019). Reviewing ensemble classification methods in breast cancer. *Computer Methods and Programs in Biomedicine*, **177**: 89–112.
- Hu, J., Shen, L., y Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Huang, G., Liu, Z., Van Der Maaten, L., y Weinberger, K. (2017). Densely connected convolutional networks. Honolulu, HI, USA. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269.
- Kabbai, L., Abdellaoui, M., y Douik, A. (2019). Image classification by combining local and global features. *Vis Comput*, **35**: 679–693.
- Khan, M., Hussain, A., Rehman, S. U., Khan, U., Maqsood, M., Mehmood, K., y Khan, M. A. (2019). Classification of melanoma and nevus in digital images for diagnosis of skin cancer. *IEEE Access*, 7: 90132–90144.
- Kuhn, H. W. y Tucker, A. W. (1951). Nonlinear programming. En: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, Calif. University of California Press, pp. 481–492.
- Kumar, B. y Hassebrook, L. (1990). Performance measures for correlation filters. *Applied optics*, **29**(20): 2997–3006.
- Lam, N., Qiu, H., Quattrochi, D., y Emerson, C. (2002). An evaluation of fractal methods for characterizing image complexity. *Cartography and Geographic Information Science*, 29(1): 25–35.
- LeCun, Y., Bottou, L., Bengio, Y., y Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. of the IEEE*.
- Lindelöf, B. y Hedblad, M. (1994). Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic. *J. Dermatol.*, **21**: 461–464.
- Liu, L., Lao, S., Fieguth, P. W., Guo, Y., Wang, X., y Pietikäinen, M. (2016). Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing*, **25**(3): 1368–1381.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, **60**(2): 91–110.

- MacKenzie-Wood, A., Milton, G., y Launey, J. (1998). Melanoma: Accuracy of clinical diagnosis. *Australas. J. Dermat*, **39**(1): 31–33.
- Mahajan, D. (2018). Exploring the limits of weakly supervised pretraining. Proceedings of the European Conference on Computer Vision (ECCV).
- Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., y Ellinger, I. (2020). Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Comput. Meth. Prog. Biomed*, **193**: 105475.
- Mandelbrot, B. B. (1977). *Fractals: form, chance, and dimension*. San Francisco: W.H. Freeman.
- Marka, A., Carter, J. B., Toto, E., y Hassanpour, S. (2019). Automated detection of nonmelanoma skin cancer using digital images: a systematic review. *BMC Medical Imaging*, **19**(21): 1–12.
- Martínez, J. y Otley, C. (2001). The management of melanoma and nonmelanoma skin cancer: A review for the primary care physician. *Mayo Clin Proc.*, **76**(12): 1253–1265.
- Metz, C. E. (1978). Basic principles of roc analysis. *Seminars in Nuclear Medicine*, **8**(4): 283–298.
- Miller, M., A. A. (1992). How accurate are dermatologists in the diagnosis of melanoma? degree of accuracy and implications. *Arch. Dermatol.*, **128**(4): 559–560.
- Nachbar, F., Stolz, M., T., M., B., C. A., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., y Plewig, G. (1994). The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, **30**(4): 551–559.
- of Health, D. (2020). Department of health. national skin cancer prevention plan 2019-2022. https://www.gov.ie/en/publication/ 4655d6-national-skin-cancer-prevention-plan-2019-2022/.
- Ojala, T., Pietikäinen, y Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, **29**(1): 51–59.
- Otsu, N. (1979). A tlreshold selection method a tlreshold selection method from graylevel histograms. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, **SMC-9**(1): 62–66.
- Pardo, A., Gutiérrez, J., Lihacova, I., López-Higuera, J., y Conde, O. (2018). On the spectral signature of melanoma: A non-parametric classification framework for cancer detection in hyperspectral imaging of melanocytic lesions. *Biomed. Opt. Express*, 9(12): 6283–6301.
- Regniers, O., Bombrun, L., Lafon, V., y Germain, C. (2016). Supervised classification of very high resolution optical images using wavelet-based textural features. *IEEE Transactions on Geoscience and Remote Sensing*, **54**(6): 3722–3735.
- Schroeder, M. (1991). *Fractals, Chaos, Power Laws: Minutes From an Infinite Paradise*. Freeman, New York.

- Shimizu, K., Iyatomi, H., Celebi, M. E., Norton, K. A., y Tanaka, M. (2015). Four-class classification of skin lesions with task decomposition strategy. *IEEE Transactions on Biomedical Engineering*, **62**(1): 274–283.
- Silveira, M., Nascimento, J., Marques, J., Marçca, T., Yamauchi, S., Maeda, J., y Rozeira, J. (2009). Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J. Sel. Top. Signal Process.*, **3**(1): 35–45.
- Society, A. C. (2019). American cancer society. skin cancer. https://www.cancer.org/ cancer/skin-cancer.html.
- Soekhoe, D., van der Putten, P., y Plaat, A. (2016). On the impact of data set size in transfer learning using deep neural networks. En: H. Boström, A. Knobbe, C. Soares, y P. Papapetrou (eds.), Advances in Intelligent Data Analysis XV. IDA 2016, Vol. 9897 de Lecture Notes in Computer Science. Springer, Cham.
- Spyridonos, P., Gaitanis, G., Likas, A., y Bassukas, I. D. (2017). Automatic discrimination of actinic keratoses from clinical photographs. *Computers in Biology and Medicine*, **88**: 50–59.
- Sun, C. y Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, **23**(3): 341–352.
- Tan, M. y Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. 09-15 Jun, Long Beach, California, USA. PMLR, Vol. 97 de *Proceedings of Machine Learning Research*, pp. 6105–6114.
- Tang, X. (1998). Texture information in run-length matrices. *IEEE Transactions on Image Processing*, **7**(11): 1602–1609.
- Theodoris, S. (2008). Pattern Recognition. Academic Press.
- Tou, J. T. y Gonzalez, R. C. (1977). Pattern Recognition Principles. Addison-Wesley.
- Tschandl, P., Rosendahl, C., y Kittler, H. (2018a). Data descriptor: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, **5**: 180161(1)–180161(9).
- Tschandl, P., Rosendahl, C., y Kittler, H. (2018b). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, **5**(180161).
- Wahba, M. A., Ashour, A. S., Guo, Y., Napoleon, S. A., y Abd-Elnaby, M. M. (2018). A novel cumulative level difference mean based gldm and modified abcd features ranked using eigenvector centrality approach for four skin lesion types classification. *Computer Methods and Programs in Biomedicine*, **165**: 163–174.
- Wu, Z., Zhao, S., Peng, Y., He, X., Zhao, X., Huang, K., Wu, X., Fan, W., Li, F., Chen, M., Li, J., Huang, W., Chen, X., y Li, Y. (2019). Studies on different cnn algorithms for face skin disease classification based on clinical images. *IEEE Access*, **7**: 66505–66511.

- Xie, B., Xiaoyu, H., Zhao, S., Li, Y., Su, J., Zhao, S., Li, Y., Su, J., Zhao, X., Kuang, Y., Wang, Y., y Chen, X. (2019). Xiangyaderm: A clinical image dataset of asian race for skin disease aided diagnosis. En: L. Zhou (ed.), *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*, Vol. 11851 de *LABELS 2019, HAL-MICCAI 2019, CuRIOUS 2019. Lecture Notes in Computer Science*. Springer, Cham.
- Yan, L., Zhu, R., Mo, N., y Liu, Y. (2017). Improved class-specific codebook with two-step classification for scene-level classification of high resolution remote sensing images. *Remote Sens*, 9(223).
- Zhang, L., Zhang, L., Zhang, D., y Zhu, H. (2011). Ensemble of local and global information for finger–knuckle-print recognition. *Pattern Recognition*, **44**(9): 1990–1998.
- Zhu, W. y Zeng, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. http://www.nesug.org/ Proceedings/nesug10/h1/h107.pdf.
- Zortea, M., Flores, E., y Scharcanski, J. (2017). A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images. *Pattern Recognition*, **64**: 92–104.

Apéndice A. Valores de las métricas para el Experimento 3 (Exp-3)

En este capítulo se muestran los valores obtenidos en cada una de las cinco repeticiones del Exp-3, Tabla X a XIX, utilizando una proporción de 70% de los elementos de la base de datos para el entrenamiento (17732 imágenes) y el 30% para la prueba (7599 imágenes).

Tabla X: Primer muestra del Exp-3 con $\alpha=0.4.$ La base de datos de entrenamiento contiene 17732 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.82	85.15	79.75	99.53
BCC	96.72	81.63	96.77	96.72
SCC	98.47	91.13	42.24	99.90
BKL	97.24	92.37	79.87	99.24
MEL	95.68	95.57	79.36	99.20
NEV	93.27	89.49	98.32	88.03
DER	99.21	100	12.52	100
VASC	99.34	100	30.77	100

Tabla XI: Primer muestra del Exp-3 con $\alpha=0.4.$ La base de datos de prueba consta de 7599 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.99	38.78	51.49	97.41
BCC	89.67	59.13	69.37	92.74
SCC	97.07	30.49	13.16	99.23
BKL	90.12	55.15	27.05	97.44
MEL	86.59	81.46	33.31	98.33
NEV	79.30	72.45	95.43	62.75
DER	98.96	50	1.27	99.99
VASC	99.05	100	14.29	100

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.86	86.48	79.28	99.56
BCC	96.48	80.64	96.73	96.44
SCC	98.44	92.35	40.86	99.91
BKL	97.38	92.43	81.65	99.22
MEL	95.45	95.54	78.08	99.21
NEV	93.13	89.30	98.18	87.94
DER	99.16	100.00	10.78	100.00
VASC	99.31	100.00	25.90	100.00

Tabla XII: Segunda muestra del Exp-3 con $\alpha = 0.4$. La base de datos de entrenamiento contiene 17732 imágenes.

Tabla XIII: Segunda muestra del Exp-3 con $\alpha = 0.4$. con $\alpha = 0.4$. La base de datos de prueba consta de 7599 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.72	38.94	49.21	97.33
BCC	89.88	58.63	69.83	92.81
SCC	97.14	31.82	15.14	99.19
BKL	90.60	56.90	30.48	97.39
MEL	86.97	83.42	34.29	98.51
NEV	80.25	73.72	95.54	64.15
DER	99.07	100.00	1.39	100.00
VASC	99.01	100.00	13.79	100.00

Tabla XIV: Tercer muestra del Exp-3 con $\alpha=0.4.$ La base de datos de entrenamiento contiene 17732 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.90	85.50	80.07	99.54
BCC	96.39	80.52	96.25	96.41
SCC	98.38	94.80	37.02	99.95
BKL	97.38	93.37	80.47	99.34
MEL	95.62	95.48	79.26	99.18
NEV	93.17	89.29	98.32	87.87
DER	99.14	94.74	10.59	99.99
VASC	99.32	100.00	26.67	100.00

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.60	42.27	47.02	97.50
BCC	89.27	55.79	69.33	92.13
SCC	97.39	40.58	15.14	99.45
BKL	90.30	56.43	27.35	97.56
MEL	86.89	82.07	33.56	98.42
NEV	80.44	73.61	96.24	63.93
DER	99.14	100.00	5.80	100.00
VASC	99.04	100.00	17.05	100.00

Tabla XV: Tercer muestra del Exp-3 con $\alpha=0.4.$ La base de datos de prueba consta de 7599 imágenes.

Tabla XVI: Cuarta muestra del Exp-3 con $\alpha=0.4.$ La base de datos de entrenamiento contiene 17732 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.98	86.28	83.28	99.53
BCC	96.55	81.02	96.41	96.57
SCC	98.50	93.44	40.24	99.93
BKL	97.42	93.07	80.76	99.32
MEL	95.65	95.73	79.39	99.22
NEV	93.05	89.34	98.04	87.87
DER	99.13	100.00	8.28	100.00
VASC	99.32	100.00	27.27	100.00

Tabla XVII: Cuarta muestra del Exp-3 con $\alpha=0.4.$ La base de datos de prueba consta de 7599 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.42	36.84	45.25	97.22
BCC	89.35	57.51	68.43	92.47
SCC	97.08	36.62	12.81	99.39
BKL	90.41	59.48	31.36	97.45
MEL	86.75	78.45	33.41	98.05
NEV	79.69	72.97	95.20	63.77
DER	99.12	100.00	4.29	100.00
VASC	99.01	100.00	14.77	100.00

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.80	84.48	78.66	99.50
BCC	96.64	81.14	96.95	96.60
SCC	98.47	92.96	41.76	99.92
BKL	97.37	93.34	80.83	99.32
MEL	95.50	95.38	78.50	99.18
NEV	93.10	89.28	98.20	87.85
DER	99.18	95.24	12.20	99.99
VASC	99.34	98.44	35.20	99.99

Tabla XVIII: Quinta muestra del Exp-3 con $\alpha=0.4.$ La base de datos de entrenamiento contiene 17732 imágenes.

Tabla XIX: Quinta muestra del Exp-3 con $\alpha=0.4.$ La base de datos de prueba consta de 7599 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.32	37.20	44.85	97.19
BCC	89.24	57.55	68.41	92.38
SCC	97.24	32.88	12.97	99.34
BKL	90.09	49.77	28.63	96.83
MEL	87.02	83.75	34.65	98.52
NEV	80.60	74.03	95.49	65.10
DER	99.01	0.00	0.00	100.00
VASC	99.18	92.86	17.57	99.99

Apéndice B. Valores de las métricas para el Experimen-

to 4 (Exp-4)

En este capítulo se muestran los valores obtenidos en cada una de las cinco repeticiones del Exp-4, Tabla XX a XXIX, utilizando una proporción de 70 % de los elementos de la base de datos para el entrenamiento (17387 imágenes) y el 30 % para la prueba (7452 imágenes).

Tabla XX: Primer muestra del Exp-4 con $\alpha=1.6.$ Los datos de entrenamiento contienen 17,387 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.87	85.93	81.95	99.50
BCC	96.86	83.14	96.05	96.98
SCC	98.55	96.37	43.16	99.96
BKL	97.22	92.78	79.88	99.27
MEL	95.47	95.96	78.31	99.27
NEV	93.75	90.64	98.08	89.10

Tabla XXI: Primer muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.18	32.95	47.30	96.78
BCC	89.35	58.77	67.17	92.75
SCC	97.05	34.67	13.20	99.32
BKL	90.41	60.11	28.23	97.78
MEL	86.59	83.64	33.60	98.52
NEV	81.12	74.67	96.19	64.90

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.82	84.17	81.23	99.45
BCC	96.77	82.80	96.04	96.88
SCC	98.45	92.72	42.83	99.91
BKL	97.32	92.48	81.34	99.22
MEL	95.57	96.51	78.56	99.37
NEV	94.13	91.12	98.19	89.79

Tabla XXII: Segunda muestra del Exp-4 con $\alpha = 1.6$. Los datos de entrenamiento contienen 17387 imágenes.

Tabla XXIII: Segunda muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.38	37.54	44.91	97.25
BCC	89.72	59.25	67.90	92.99
SCC	97.53	48.65	19.78	99.48
BKL	90.30	57.00	29.74	97.38
MEL	86.19	81.17	30.96	98.41
NEV	80.25	74.19	95.52	63.45

Tabla XXIV: Tercer muestra del Exp-4 con $\alpha=1.6.$ Los datos de entrenamiento contienen 17387 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.94	87.38	79.86	99.60
BCC	96.82	82.96	96.50	96.87
SCC	98.43	92.93	41.53	99.92
BKL	97.25	92.89	80.36	99.27
MEL	95.50	96.51	78.02	99.37
NEV	93.69	90.45	98.14	88.94

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.34	40.26	44.06	97.39
BCC	89.90	58.94	68.66	93.00
SCC	97.26	35.82	12.97	99.41
BKL	90.66	59.64	30.09	97.65
MEL	86.12	79.78	32.40	98.16
NEV	80.25	74.13	95.56	63.48

Tabla XXV: Tercer muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Tabla XXVI: Cuarta muestra del Exp-4 con $\alpha=1.6.$ Los datos de entrenamiento contienen 17387 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.77	84.46	78.83	99.48
BCC	96.83	82.79	96.35	96.90
SCC	98.51	96.69	40.89	99.96
BKL	97.31	93.00	80.56	99.29
MEL	95.66	95.89	79.76	99.23
NEV	94.02	91.10	98.05	89.70

Tabla XXVII: Cuarta muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.88	42.54	42.70	97.86
BCC	89.56	59.26	69.05	92.71
SCC	96.97	35.87	16.50	99.19
BKL	90.24	58.51	28.63	97.58
MEL	86.81	81.37	33.83	98.32
NEV	80.29	74.03	95.56	63.79

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	98.96	87.45	80.75	99.60
BCC	96.85	82.96	96.50	96.91
SCC	98.46	94.71	40.96	99.94
BKL	97.30	93.23	80.61	99.30
MEL	95.57	96.15	78.81	99.30
NEV	94.00	90.87	98.28	89.43

Tabla XXVIII: Quinta muestra del Exp-4 con $\alpha=1.6.$ Los datos de entrenamiento contienen 17387 imágenes.

Tabla XXIX: Quinta muestra del Exp-4 con $\alpha = 1.6$. Los datos de prueba contienen 7452 imágenes.

Clase	Exactitud (%)	Precisión (%)	Sensibilidad (%)	Especificidad (%)
AK	95.13	37.97	46.79	97.02
BCC	89.79	59.52	69.09	92.91
SCC	97.34	45.57	18.85	99.41
BKL	90.37	55.39	29.66	97.28
MEL	86.42	82.33	32.30	98.46
NEV	81.05	74.97	95.57	65.23