



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL TÍTULO
DE INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES**

**TEMA: DETECCIÓN DE CÁNCER DE PIEL ASISTIDO POR
ORDENADOR CON ESPECIALIZACIÓN EN MELANOMAS**

AUTOR: RON DOMÍNGUEZ, DAVID VINICIO

DIRECTOR: DR. CARRERA ERAZO, ENRIQUE VINICIO

SANGOLQUÍ

2018



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA Y TELECOMUNICACIONES

CERTIFICACIÓN

Certifico que el trabajo de titulación, **“DETECCIÓN DE CÁNCER DE PIEL ASISTIDO POR ORDENADOR CON ESPECIALIZACIÓN EN MELANOMAS”** realizado por el señor ***RON DOMÍNGUEZ DAVID VINICIO***, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al Señor ***RON DOMÍNGUEZ DAVID VINICIO*** para que lo sustente públicamente.

Sangolquí, 02 de Febrero de 2018

A handwritten signature in blue ink, which appears to read 'Enrique Vinicio Carrera Erazo', is positioned above the printed name of the director.

PhD Enrique Vinicio Carrera Erazo

DIRECTOR



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA Y TELECOMUNICACIONES

AUTORÍA DE RESPONSABILIDAD

Yo, **RON DOMÍNGUEZ DAVID VINICIO**, con cédula de identidad N° 171800442-5, declaro que este trabajo de titulación **“DETECCIÓN DE CÁNCER DE PIEL ASISTIDO POR ORDENADOR CON ESPECIALIZACIÓN EN MELANOMAS”** ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 02 de Febrero de 2018

David Vinicio Ron Domínguez

CI. 171800442-5



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA EN ELECTRÓNICA Y TELECOMUNICACIONES

AUTORIZACIÓN

Yo, **RON DOMÍNGUEZ DAVID VINICIO**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación **“DETECCIÓN DE CÁNCER DE PIEL ASISTIDO POR ORDENADOR CON ESPECIALIZACIÓN EN MELANOMAS”** cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 02 de Febrero 2018

David Vinicio Ron Domínguez

CI. 171800442-5

DEDICATORIA

El presente proyecto de titulación representa el esfuerzo, la perseverancia y la responsabilidad que se me inculco desde la niñez para cada día ser mejor, sin dejar a un lado los valores éticos y morales aprendidos; también representa la finalización de una meta y el inicio de muchas más.

Dedico con mucho cariño este trabajo a mis padres Marco y Fremiot que me han apoyado incondicionalmente, me han formado en valores, me han enseñado ser una persona con carácter y me han dejado la mejor de las herencias que un ser humano puede desear, la educación.

A mis hermanos Marco e Ignacio quienes hemos estado juntos, acompañándonos en las adversidades que todos los días se han presentado, sin que se llegue a claudicar en el apoyo incondicional.

A mi querida abuelita Matilde y abuelito Nelson quienes estuvieron ayudándome y apoyándome para poder crecer en valores y responsabilidad.

A mis amigos de siempre, por sus ánimos y por compartir esta etapa, en especial a Raquel, Anita, Laura, Diego, René, Danilo, Andrés C., Jesús, Andrés V., Julio, Andrés G., con los que he finalizo la primera meta académica.

David Vinicio Ron Domínguez

AGRADECIMIENTO

Agradezco a Dios, por permitirme seguirlo y por su fidelidad incondicional en todo momento de mi vida, además por permitirme tener una familia reunida.

Agradezco a mis padres, por su ejemplo impartido día a día, su esfuerzo y sus consejos; son los principales actores dentro de mi vida para cada día llegar más lejos.

De igual manera agradezco a mis hermanos, por todo el apoyo, compañía y ayuda que cada día se ha hecho notoria, permitiéndome compartir experiencias, conocimiento y consejos.

Agradezco a mis amigos, que hemos estado en momentos de esfuerzo, en momentos de alegrías, pero también en momentos de angustia, permitiéndonos conocernos y saber quiénes somos amigos.

A los docentes que impartieron enseñanzas, experiencias y consejo, que a más de ser profesores fueron y son mis amigos, en especial agradecerle al Dr. Vinicio Carrera, por guiarme, aconsejarme y motivarme en el desarrollo del mi proyecto de titulación.

David Vinicio Ron Domínguez

ÍNDICE DE CONTENIDOS

CARÁTULA	i
CERTIFICADO	ii
AUTORÍA DE RESPONSABILIDAD.....	iii
AUTORIZACIÓN.....	iv
DEDICATORIA.....	v
AGRADECIMIENTO.....	vi
ÍNDICE DE CONTENIDOS	vii
ÍNDICE DE TABLAS	xi
ÍNDICE DE FIGURAS	xii
RESUMEN.....	xv
ABSTRACT	xvi
CAPÍTULO 1	1
INTRODUCCIÓN	1
1.1 Antecedentes	1
1.2 Justificación e importancia.....	1
1.3 Alcance del proyecto.....	2
1.4 Objetivos	3
1.4.1 General	3
1.4.2 Específicos	3
CAPÍTULO 2	4
MARCO TEÓRICO.....	4
2.1 La piel.....	4
2.2 Lunares normales	4

2.3 Cáncer de piel.....	5
2.3.1 Cáncer de piel no melanoma (CPNM).....	5
2.3.2 Lunares atípicos (nevus displásicos).....	7
2.3.3 Melanoma maligno (MM).....	7
2.4 Identificadores del melanoma.....	9
2.5 Dermatoscopía.....	9
2.6 Técnicas de procesamiento digital de imágenes.....	10
2.6.1 Sistemas de color.....	10
2.6.2 Operaciones con imágenes.....	10
2.6.3 Mejoramiento de imagen en el dominio espacial.....	11
2.6.4 Operaciones morfológicas.....	12
2.7 Detección de bordes.....	13
2.8 Detección de textura.....	13
2.9 Segmentación.....	14
2.10 Aprendizaje de máquina.....	15
2.10.1 Máquina de soporte vectorial (SVM).....	15
2.10.2 Árbol de decisión.....	19
2.10.3 Reducción de características.....	20
2.11 Métodos de validación de resultados.....	20
2.11.1 Validación cruzada k-fold.....	21
2.11.2 Matriz de confusión.....	22
2.11.3 Las curvas ROC.....	23
CAPÍTULO 3.....	25
MATERIALES, MÉTODOS E IMPLEMENTACIÓN.....	25

3.1 Base de datos	25
3.2 Matlab.....	25
3.3 Descripción del sistema.....	26
3.4 Segmentación	27
3.5 Extracción de características	30
3.5.1 Asimetría	31
3.5.2 Diferencia de bordes.....	33
3.5.3 Color medio.....	35
3.5.4 Textura	36
3.5.5 Velo blanco-azul	38
3.6 Normalización y características	40
3.7 Configuración de los clasificadores	44
3.8 Configuración de la técnica PCA	45
CAPÍTULO 4	47
RESULTADOS Y ANÁLISIS	47
4.1 Segmentación de las imágenes	48
4.2 Análisis en términos de sensibilidad y exactitud al utilizar 28 características empleando el clasificador SVM.....	50
4.3 Análisis de reducción de características.....	52
4.4 Análisis en términos de sensibilidad y exactitud al utilizar 28 características empleando el clasificador árbol de decisiones	55
4.5 Análisis de desempeño del sistema	59
CAPÍTULO 5	60
CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS	60
5.1 Conclusiones	60

5.2 Recomendaciones..... 61

5.3 Trabajos Futuros..... 61

REFERENCIA 63

ÍNDICE DE TABLAS

Tabla 1 <i>Descripción de matriz de confusión para dos clases</i>	22
Tabla 2 <i>Valores de textura</i>	38
Tabla 3 <i>Características extraídas para la clasificación</i>	41
Tabla 4 <i>P-VALOR de las 28 características para un $\alpha=6\%$</i>	43
Tabla 5 <i>Nivel de similitud entre las máscaras, empleando el índice de Jaccard</i>	49
Tabla 6 <i>Matriz de confusión en términos de sensibilidad - 28 Características</i>	50
Tabla 7 <i>Matriz de confusión en términos de exactitud - 28 Características</i>	50
Tabla 8 <i>Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 28 características</i>	51
Tabla 9 <i>Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 22 características</i>	52
Tabla 10 <i>Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 13 características</i>	53
Tabla 11 <i>Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para PCA aplicado a 28 características</i>	54
Tabla 12 <i>Matriz de confusión en términos de sensibilidad - 28 Características</i>	55
Tabla 13 <i>Matriz de confusión en términos de exactitud - 28 Características</i>	55
Tabla 14 <i>Resultados del análisis del sistema de detección utilizando los clasificadores SVM y árbol de decisiones con 28 características en término de sensibilidad</i>	57
Tabla 15 <i>Resultados del análisis del sistema de detección utilizando los clasificadores SVM y árbol de decisiones con 28 características en términos de exactitud.</i>	57

ÍNDICE DE FIGURAS

Figura 1 Capas de la piel.....	4
Figura 2 Lunares normales y su evolución: (a) Máculas uniforme marrón entre 1-2 mm de diámetro, (b) Mácula agravado, (c) Pápula pigmentada, puede ser poco elevada, (d) Pápula elevada, (e) Pápula de color rosado o color carne	5
Figura 3 Carcinoma basocelular nodular: lesión tumoral color piel	6
Figura 4 Carcinoma espinocelular: lesión tumoral queratósica, con base infiltrada, Daño actínico del paciente	6
Figura 5 Cáncer de células escamosas: Queratosis actínica. Entre el 40%- 60% de los casos inician con esta forma	6
Figura 6 Lunares atípicos: (a) Nódulo negro de 3mm, (b) Lesión de color marrón oscuro con una superficie elevada, (c) Pápula complejas de superficie irregular y ligeramente rosa, (d) Contextura elevada.....	7
Figura 7 Melanoma lentigo maligno (MLM): mácula hiperpigmentada con diámetro mayor de 3cm, asimétrica, con bordes irregulares, de varios años de evolución	8
Figura 8 Melanoma de extensión superficial: lesión muscular de 7x11mm. Bordes irregulares, ondulados y poco distintivos. Variación en colores (marrón claro y oscuro)	8
Figura 9 Melanoma nodular (MN): lesión nevica existente que, aumentado de tamaño en los últimos años, dos meses de evolución	8
Figura 10 Velo Blanco-azul	9
Figura 11 Ejemplo de histograma: (a) Imagen original en escala de grises, (b) Histograma de la imagen.....	12
Figura 12 Ejemplo de Ecuilización del histograma: (a) Imagen ecualizada en escala de grises, (b) Histograma ecualizado de la imagen	12
Figura 13 Muestras separables linealmente con una maximización del margen.....	16
Figura 14 Muestras no separables linealmente.....	16

Figura 15 Transformación del espacio de entrada.....	17
Figura 16 Sobreajuste y Subajuste	18
Figura 17 Ejemplo de árbol de decisiones.....	19
Figura 18 Ejemplo de validación cruzada K-fold	21
Figura 19 Ejemplo de la curva ROC, elementos rojos enfermos y elementos azules sanos	24
Figura 20 Evaluación del proceso de segmentación del lunar	26
Figura 21 Proceso de detección de cáncer de piel tipo melanoma	26
Figura 22 Evaluación del proceso de segmentación del lunar	27
Figura 23 Componentes: (a) Roja, (b) Verde, (c) Azul.....	29
Figura 24 Lunar binarizado invertido.....	29
Figura 25 Resultado de la segmentación de la imagen: (a) Imagen original, (b) Imagen segmentada	30
Figura 26 Proceso inicial para la extracción de características	30
Figura 27 Proceso general de extracción de asimetría	31
Figura 28 Proceso 1 de extracción de asimetría	31
Figura 29 Proceso 2 de extracción de asimetría	32
Figura 30 Distancia entre el centroide (círculo azul) y el centro de peso (punto rojo): (a) Lunar simétrico, (b) Lunar asimétrico.....	33
Figura 31 Diferencia de bordes: (a) Bordes definidos, (b) Bordes difuminados	33
Figura 32 Proceso de extracción de bordes	34
Figura 33 Diferencia de bordes: (a) Imagen original e índice de Jaccard, (b) Componente R, (c) Componente G, (d) Componente B	35
Figura 34 Proceso de extracción de color medio	35
Figura 35 Proceso de extracción de textura.....	36
Figura 36 Textura: (a) Imagen con textura fuerte, (b) Imagen con textura suave, (c) y (d) Imagen recortada.....	37
Figura 37 Proceso de extracción de estructura velo blanco-azul	38

Figura 38 Operación lógica AND:(a) Componente roja, (b) Componente azul, (c) Intersección entre la componente roja y azul	39
Figura 39 Velo blanco-azul: (a) Imagen original, (b) Representación del resultado del velo blanco-azul.....	40
Figura 40 Influencia de la características en el sistema	42
Figura 41 Segmentación con alto grado de similitud: (a) Imagen original, (b) Imagen segmentada, (c) Máscara de la base de datos, (d) Máscara extraída la imagen	48
Figura 42 Segmentación con bajo índice de similitud: (a) Imagen original, (b) Imagen segmentada, (c) Máscara de la base de dato, (d) Máscara extraída la imagen	49
Figura 43 Curva ROC del clasificador SVM: (a) En términos de sensibilidad, (b) En términos de exactitud.....	51
Figura 44 Gráfica de reducción de características - PCA.....	54
Figura 45 Curva ROC: (a) Árbol de decisiones en términos de sensibilidad, (b) Árbol de decisiones en términos de exactitud	56
Figura 46 Representación gráfica del clasificador árbol de decisiones con 70 jerarquías	58
Figura 47 Representación gráfica del clasificador árbol de decisiones con 30 jerarquías	58

RESUMEN

El cáncer de piel se deriva en múltiples enfermedades, entre ellas la de mayor peligro es el melanoma que es una enfermedad producida principalmente debido a la exposición excesiva al sol. El melanoma es considerado el cáncer de piel más agresivo porque tiene la capacidad de invasión profunda, comprometiendo los órganos vitales y el cerebro. Por esta razón es importante detectar esta enfermedad en etapas tempranas. Existen dos métodos para detectar el melanoma: el método visual que está sujeto a la experiencia de los médicos; y el método automático que emplea un sistema computacional. En la presente investigación se propone el desarrollo de un sistema computacional para asistir al médico al momento de generar un diagnóstico. Con esto se proporciona una segunda opinión que puede reafirmar el análisis obtenido, o puede derivar al paciente a una biopsia de confirmación. El sistema computacional para la detección de cáncer de piel tipo melanoma emplea áreas de la investigación de interés científico y clínico. El sistema vincula el procesamiento digital de imágenes y el aprendizaje de máquina. Se emplea una base de datos confiable que contiene imágenes con sus respectivas máscaras, y el diagnóstico de cada una de ellas. Para analizar cada imagen se realiza el siguiente proceso: Segmentación, donde se divide la imagen en sus partes constituyentes y se extrae el área de interés representada por la máscara de cada imagen; además se obtiene el grado de similitud entre la máscara extraída y su equivalente de la base de datos. La extracción de características, donde se busca las características de asimetría, borde, color, textura y velo blanco-azul. La clasificación, donde se compara el clasificador SVM con el árbol de decisiones, dando como resultado que el SVM tiene mejores prestaciones para el sistema de detección. El resultado final es que el sistema SVM presenta un 98.1% de sensibilidad y 80.1% de exactitud al trabajar con 28 características.

Palabras claves

- **MELANOMA.**
- **PROCESAMIENTO DIGITAL DE IMÁGENES.**
- **APRENDIZAJE DE MÁQUINA.**

ABSTRACT

Skin cancer is caused by multiple diseases. Among these diseases, the one of greater danger is melanoma. This is a disease produced mostly due to excessive sun exposure. Melanoma is considered to be the most aggressive cancer because it has the capacity for deep invasion, compromising the vital organs and the brain. For this reason, it is important to detect this disease in early stages. There are mainly two methods of detecting melanoma: the visual method which is subject to the experience of each doctor and the automatic method which employs a computational system. In this document, the idea is to develop a computational system that can assist the physician for generating a diagnosis. It would provide a second opinion that can reaffirm the analysis obtained, or it can derive in the need of a biopsy confirmation for the patient. The computational system for detection of skin cancer type melanoma uses research areas of scientific and clinical interest. This system links digital image processing and machine learning. The system uses a database of high reliability which contains images with their respective masks and the diagnosis of each one of them. To analyze each image, the following process is performed: Segmentation, this is where the image is divided into its constituent parts, and the area of interest represented by the mask of each image is extracted; In addition, the degree of similarity between the extracted mask and its equivalent in the database is obtained. The extraction of characteristics where the system finds the characteristics of asymmetry, edge, color, texture, and white-blue veil. Classification, here the SVM classifier is compared to the decision tree, resulting in the SVM having better performance for the detection system. Finally the results show that the SVM system presents 98.1% of sensitivity and 80.1% of accuracy when it works with 28 characteristics.

Keywords:

- **MELANOMA.**
- **DIGITAL IMAGE PROCESSING.**
- **MACHINE LEARNING.**

CAPÍTULO 1

INTRODUCCIÓN

1.1 Antecedentes

El cáncer es una enfermedad genética, sin embargo, existen factores que aumentan su desarrollo tales como: Fumar, mala nutrición, radiación, edad o exposición excesiva a la luz solar. Este último factor es uno de los más influyentes en la aparición de cáncer de piel tipo melanoma, y el riesgo se incrementa en los residentes del Ecuador (Ferlay, 2017). No obstante, el tener uno o varios factores de riesgo no significa que una persona pueda contraer la enfermedad.

El análisis del melanoma se lo realiza habitualmente mediante inspección visual del médico tratante (Hedblad & Lindelöf, 1994) (Mackie & Morton, 1998). Ya que el diagnóstico es subjetivo a la experiencia del médico, es fundamental contar con un sistema médico computacional o CAD (*Computer aided diagnosis*). Este sistema permite asistir al médico al momento de entregar un diagnóstico al paciente (Menzies & Talbot, 2005) (Nachbar, Stolz, & Merkle, 1994) (Johr, 2002) (Zalaudek, Argenziano, & Soyer, 2006) (Menzies, Ingvar, Crotty, & McCarthy, 1996).

1.2 Justificación e importancia

En el Ecuador las estadísticas indican que el cáncer de piel tipo melanoma ha tenido una mayor incidencia entre los años 2001-2010. En este período se ha evidenciado un aumento de 0.9 y 0.1 casos por cada 100.00 habitantes hombres y mujeres, respectivamente (Cueva & Yépez, 2014). Ya que los análisis estadísticos muestran una tendencia ascendente, es necesario que la enfermedad deba ser detectada en etapas tempranas al ser un tipo de cáncer muy agresivo.

Los médicos, al no emplear un instrumento o sistema CAD de diagnóstico, poseen una precisión menor al 31% ya que utilizan un proceso altamente subjetivo y es poco reproducible (Pehamberger, Binder, Steiner, & WolffK, 1993) (Steiner & Wolff, 1987) (Hedblad & Lindelöf, 1994) (Mackie & Morton, 1998).

La detección de melanoma se puede realizar mediante técnicas manuales tales como: Dermatoscopia, microscopia de epiluminiscencia, epiluminiscencia de polarización cruzada y transiluminación lateral. Estas técnicas permiten aumentar los detalles morfológicos que se visualizan, proporcionando una mejora de la sensibilidad entre un 10%-30% (Mayer, 1997)

(Argenziano & Soyer, 2001) (Binder, Schwarz, & Winkler, 1995) (Pehamberger, Binder, Steiner, & WolffK, 1993) (Dhawan, Gordon, & Rangayyan, 1984) (Zouridakis, Doshi, Duvic, & Mullani, 2005).

Los sistemas computacionales existentes para determinar el cáncer de piel tipo melanoma trabajan en función de predicciones y algoritmos tales como: Regla ABCD-E, método Menzies, lista de verificación de siete puntos y lista de verificación de tres puntos. Sin embargo estos sistemas requieren la adquisición de hardware y software propio (Burroni, Corona, & Dell'Eva, 2004) (Abbasi, Shaw, & Rigel, 2004) (Turrisi & Robinson, 2006) (Menzies, Ingvar, Crotty, & McCarthy, 1996) (Menzies & Vestergaard, 2008) (Johr, 2002) (HealSmith, Bourke, Osborne, & Graham-Brown, 1994) (Pozzo, Benelli, & Roscetti, 1999) (Zalaudek, Argenziano, & Soyer, 2006).

La presente investigación propone desarrollar un sistema CAD para la detección de melanomas, empleando características extraídas mediante procesamiento digital de imágenes y técnicas de aprendizaje de máquina. De esta manera, se busca asistir al médico en el diagnóstico de cáncer de piel tipo melanoma en etapa temprana.

1.3 Alcance del proyecto

El presente trabajo de investigación tiene como objetivo desarrollar un sistema que permita detectar el cáncer de piel tipo melanoma. Para esto se emplea una base de datos, algoritmos de procesamiento digital de imágenes, algoritmos de clasificación y métodos para detección de las características de los melanomas.

En primera instancia, mediante el pre-procesamiento de imágenes, el sistema CAD localiza, segmenta y determina la máscara del lunar y las características que destacan a un melanoma (asimetría, bordes, color, textura, y velo blanco-azul).

Posteriormente se entrena al clasificador para detectar el melanoma, empleando técnicas de aprendizaje de máquina de Soporte Vectorial (SVM) y árbol de decisiones. Finalmente se compara estas técnicas empleando el método de validación cruzada para estimar el rendimiento entre ellas. Los resultados se presentan en términos de sensibilidad, especificidad, exactitud y capacidad predictiva.

1.4 Objetivos

1.4.1 General

Diseñar, implementar y evaluar un sistema de detección asistido por ordenador para la detección de cáncer de piel tipo melanoma, empleando procesamiento digital de imágenes y aprendizaje de máquina.

1.4.2 Específicos

- Comprender y analizar el funcionamiento teórico de los algoritmos de clasificación y algoritmos de procesamiento de imágenes.
- Desarrollar un sistema de procesamiento digital de imágenes para localización de lunares, empleando Matlab.
- Implementar en Matlab un sistema para la detección de melanomas en función de las características de asimetría, bordes, color, textura y velo blanco- azul.
- Evaluar y analizar los resultados obtenidos de Matlab, en términos de exactitud, sensibilidad, especificidad y capacidad predictiva.

CAPÍTULO 2

MARCO TEÓRICO

2.1 La piel

La piel o membrana cutánea es el órgano más extenso del ser humano. La dermatoscopía estudia a la piel y la divide en tres capas: la epidermis, la dermis y la hipodermis, como se indica en la Figura 1. La epidermis está conformada por células escamosas, células basales y melanocitos. Los melanocitos producen la pigmentación marrón llamada melanina y son las principales células con probabilidad de convertirse en melanomas (Palacios, 2007) (Pescador, 2012).

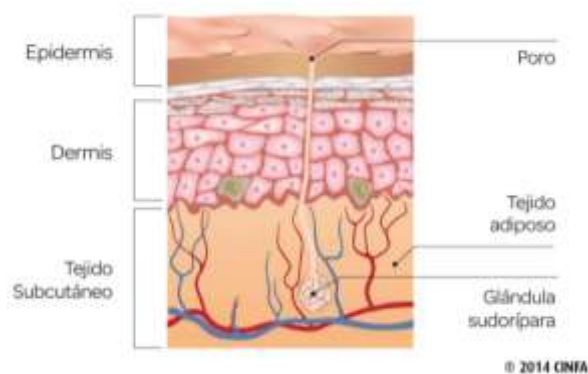


Figura 1 Capas de la piel
Fuente. (Cinfasalud, 2014)

2.2 Lunares normales

Un lunar es un tumor benigno de la piel y se identifica por las siguientes características: son manchas de color café uniforme, canela o negro; tiene forma plana, prominente, redonda u ovalada; sus bordes son lisos, y su diámetro máximo es de 6 milímetros (Vallejos, 2016). En la Figura 2 se exponen 5 tipos de lunares normales de la piel.



Figura 2 Lunares normales y su evolución: (a) Máculas uniforme marrón entre 1-2 mm de diámetro, (b) Mácula agravado, (c) Pápula pigmentada, puede ser poco elevada, (d) Pápula elevada, (e) Pápula de color rosado o color carne

Fuente. (Pérez, 2017)

2.3 Cáncer de piel

El cáncer de piel afecta a las células cutáneas produciendo que las células en el cuerpo aumenten en forma descontrolada. Las partes más habituales para desarrollar melanoma son: pecho, espalda, piernas, orejas, cuello, brazos, cabeza, abdomen, y rostro.

El cáncer de piel se divide en tres principales grupos: cáncer de piel no melanoma, lunares atípicos y melanoma maligno (Magliano, 2015) (Vallejos, 2016).

2.3.1 Cáncer de piel no melanoma (CPNM)

Se tiene tres tipos de cáncer de piel no melanoma: el carcinoma basocelular, el carcinoma espinocelular y el carcinoma celular escamoso.

En el carcinoma basocelular las células muestran una morfología que se asume como el estrato basal de la epidermis, su característica se muestra en la Figura 3 donde se evidencia una invasión local, es de crecimiento lento, tiene mínimo riesgo de presentar metástasis y rara vez es mortal.

El carcinoma espinocelular mostrado en la Figura 4, es considerado como un neoplasma maligno, tiene probabilidad de producirse metástasis, es de crecimiento acelerado y se puede identificar con mucha frecuencia sobre las lesiones precancerosas.

El carcinoma celular escamoso mostrado en la Figura 5, se desarrolla en la piel que ha absorbido los rayos del sol durante varios años, aparece habitualmente en el rostro, orejas, uñas, labios, dorso de las manos, brazos y las piernas (Vidrio, 2003) (Pérez, 2017).



Figura 3 Carcinoma basocelular nodular: lesión tumoral color piel
Fuente. (Magliano, 2015)



Figura 4 Carcinoma espinocelular: lesión tumoral queratósica, con base infiltrada, Daño actínico del paciente
Fuente. (Magliano, 2015)



Figura 5 Cáncer de células escamosas: Queratosis actínica. Entre el 40%- 60% de los casos inician con esta forma
Fuente. (Pérez, 2017)

2.3.2 Lunares atípicos (nevus displásicos)

Los lunares atípicos suelen ser benignos, sin embargo, son propensos a desarrollar melanoma. En la Figura 6 se exponen 4 tipos de lunares atípicos con estructura diferente.



Figura 6 Lunares atípicos: (a) Nódulo negro de 3mm, (b) Lesión de color marrón oscuro con una superficie elevada, (c) Pápula compleja de superficie irregular y ligeramente rosa, (d) Contextura elevada

Fuente. (Pérez, 2017)

2.3.3 Melanoma maligno (MM)

Existen varios tipos de melanoma maligno: melanoma lentigo maligno, melanoma extensivo superficial, melanoma nodular, y melanoma vertiginoso acral (Magliano, 2015) (Vidrio, 2003).

El melanoma lentigo maligno se visualiza como una mácula hiperpigmentada con colores tales como negro, marrón y gris, como se identifica en la Figura 7. Se localizan usualmente en el rostro especialmente en las mejillas y nariz.

El melanoma extensivo superficial mostrado en la Figura 8 es uno de los más comunes, representa a un aproximado del 70% de casos de melanoma maligno, se localiza en el tronco y en las piernas.

El melanoma nodular mostrado en las Figura 9, se desarrolla en forma de lunar; representa entre el 15%-30% de los casos de melanomas malignos. Puede localizarse especialmente en el dorso o en las extremidades.

El melanoma vertiginoso acral se localizan con mayor frecuencia en la palma de las manos, en la planta de los pies y zonas subungueales. Representa entre el 2%-8% de casos de melanoma maligno en personas con tez clara, y el 60%-70% en personas con tez negra.



Figura 7 Melanoma lentigo maligno (MLM): mácula hiperpigmentada con diámetro mayor de 3cm, asimétrica, con bordes irregulares, de varios años de evolución
Fuente. (Magliano, 2015)



Figura 8 Melanoma de extensión superficial: lesión muscular de 7x11mm. Bordes irregulares, ondulados y poco distintivos. Variación en colores (marrón claro y oscuro)
Fuente. (Pérez, 2017)



Figura 9 Melanoma nodular (MN): lesión nevíca existente que, aumentado de tamaño en los últimos años, dos meses de evolución
Fuente. (Magliano, 2015)

2.4 Identificadores del melanoma

Los identificadores empleados en el sistema CAD desarrollado en el presente trabajo de investigación son asimetría, borde, color, textura y velo blanco-azul, descritos a continuación.

- **Asimetría:** Se valora en función de la forma, considerando si la mitad del lunar es diferente de la otra mitad.
- **Borde:** Los bordes son irregulares, desiguales, dentados, ondulados o mal definidos.
- **Color:** El color no es uniforme, diferentes tonalidades en el interior con respecto al exterior, diferentes tonos de marrón y negro, a veces blanco, rojo o azul.
- **Textura:** Se identifica a través la contextura o forma del lunar, permitiendo definir métricas de comparación entre lunares malignos y benignos.
- **Velo blanco-azul:** Las características que presenta el velo blanco-azul están descritas en las Figura 10 y son: pigmentación azul-blanquecina, no tiene estructura en su interior, no ocupa la lesión por completo y otorga una apariencia en vidrio esmerilado (Zaballos, Carrera, Puig, & Malveyh, 2004).



Figura 10 Velo Blanco-azul
Fuente. (Alcántara & Menárguez, 2013)

2.5 Dermatoscopía

La dermatoscopía es una técnica no invasiva, que permite visualizar en profundidad cada lesión de la piel a través de un lente óptico de mano denominando dermatoscopio, cuya lente amplifica 10 o más veces la imagen.

Actualmente se cuenta con la dermatoscopia digital, la cual combina las características ópticas del dermatoscopio con técnicas de captura de imágenes, permitiendo obtener resultados digitales (Alcántara & Menárguez, 2013).

2.6 Técnicas de procesamiento digital de imágenes

2.6.1 Sistemas de color

Las imágenes digitales a color se pueden descomponer en tres modelos: modelo Red, Green, Blue (RGB); modelo Cian, Magenta, Yellow (CMY); y modelo Hue, Saturation, Intensity (HSI).

Una imagen en RGB se modela en un sistema de coordenadas cartesianas de 3 dimensiones. Las dos primeras dimensiones identifican a la imagen en un sistema bidimensional de valores discretos y finitos en función de la intensidad de la luz. La tercera dimensión representa a cada espacio de color.

Los dispositivos modernos permiten capturar imágenes a color, donde cada región de la imagen es representada por la intersección entre una fila y una columna y se denomina pixel. Cada pixel representa un vector común C formado por una combinación lineal de los colores RGB, y la proyección en cada eje cromático (r, g, b) descrito en la ecuación 2.1.

$$C = rR + gG + bB \quad (2.1)$$

En el procesamiento digital de imágenes médicas se emplean imágenes binarizadas para diferenciar múltiples tonalidades. Este procedimiento se lo puede realizar mediante selección de una componente de color, o transformación de la imagen a escala de grises. Esta última es la opción más óptima ya que no pierde información relevante, y es calculada como:

$$I(GS) = 0.299 * R + 0.587 * G + 0.114 * B \quad (2.2)$$

Donde $I(GS)$ representa a la imagen en escala de grises.

2.6.2 Operaciones con imágenes

En una imagen se pueden realizar varias operaciones aritméticas como la suma, resta, multiplicación y la división. Además de operaciones de relación como Min, Max, y operaciones lógicas como AND, NAND, OR, XOR. Estas operaciones se realizan pixel a pixel formando una segunda imagen compuesta por las dos anteriores (Ortiz, 2013).

2.6.3 Mejoramiento de imagen en el dominio espacial

El objetivo del mejoramiento es procesar una imagen para extraer sus características. Este método trabaja sobre el plano de la imagen permitiendo manipular directamente sus píxeles, y es denotado por:

$$g(x,y) = T[f(x,y)] \quad (2.3)$$

Donde $f(x,y)$ es la imagen de entrada del proceso; $g(x,y)$ es la imagen de salida y T es un operador de f .

Las operaciones de mejoramiento espacial empleadas son: binarización, el procesamiento de histograma, la ecualización del histograma y la transformación negativa, descritos a continuación (Vilet, 2005) (Ortiz, 2013).

En la binarización se descompone una imagen digital empleando niveles de umbral, es decir que a partir de un umbral se diferencien dos niveles uno superior y otro inferior.

El procesamiento del histograma representa a una imagen digital como niveles de gris dentro de un rango definido. En la Figura 11 se describe un ejemplo de histograma tomado de la base de datos (*Database* - DTB). En la imagen se identifica un eje horizontal, que representa el valor de los niveles de gris (r_k), y un eje vertical establecido como $h(r_k) = n_k$. Donde $h(r_k)$ corresponde al valor de una función discreta sin normalizar, y n_k representa el número de píxeles en la imagen.

La ecualización del histograma de una imagen permite mejorar el contraste de la misma. En la Figura 12 (a) se muestra la imagen ecualizada y en las Figuras 12 (b) se observa el histograma ecualizado, el mismo que evidencia una tendencia uniforme.

La transformación negativa es el proceso que transforma la imagen digital original en su equivalente negativo. Este proceso se realiza mediante la inversión de intensidad para mejorar la visualización del lunar, y se describe como:

$$s = L - 1 - r \quad (2.4)$$

Donde s y r son los píxeles que representa un antes y después respectivamente, y $L - 1$ es el nivel de gris máximo que puede tomar la imagen.

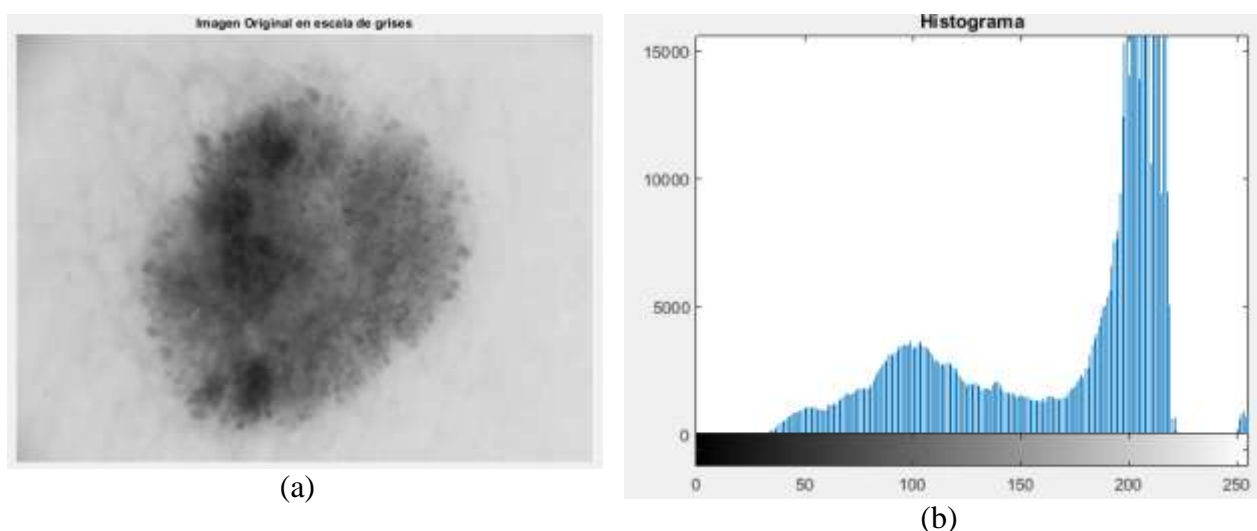


Figura 11 Ejemplo de histograma: (a) Imagen original en escala de grises, (b) Histograma de la imagen

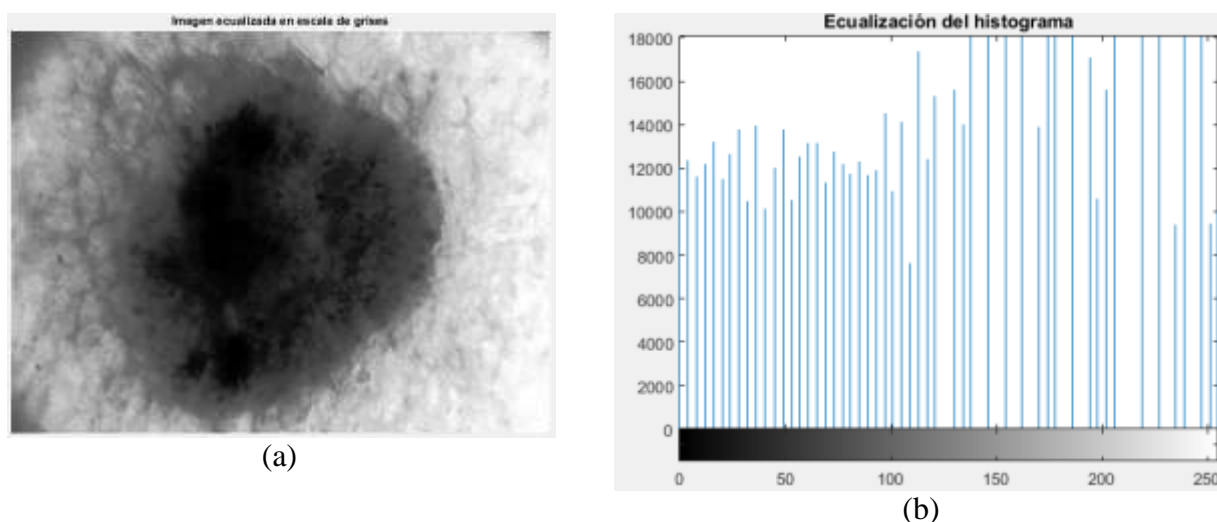


Figura 12 Ejemplo de Ecuación del histograma: (a) Imagen ecualizada en escala de grises, (b) Histograma ecualizado de la imagen

2.6.4 Operaciones morfológicas

Una operación morfológica es una operación no lineal entre la imagen original (binaria) y una matriz denominada elemento estructural. Existen varias operaciones morfológicas como: dilatación, erosión, apertura y cierre. La operación de dilatación permite agregar píxeles, y añadir puntos del fondo para unir líneas discontinuas. La operación de erosión permite filtrar la región periférica de los objetos, y separar un objeto de otro en base a su contorno. La operación apertura aplica una erosión seguida de una dilatación, permite desunir los objetos pequeños de otros

grandes. La operación de cierre permite cerrar los huecos o conectar objetos cercanos, es una operación extensiva e idempotente, y emplea a su vez la operación de dilatación seguida de la erosión.

2.7 Detección de bordes

Los bordes de una imagen son la frontera entre dos regiones de diferentes tonalidades, caracterizadas por presentar fuertes contrastes de intensidad. Para detectar los bordes se debe analizar los cambios bruscos en los niveles de gris de los píxeles vecinos, y suprimir las áreas con valores constantes de gris. Para esto se emplean los operadores de trazo de contorno definidos mediante la función *boundaries* y *signature* de Matlab.

2.8 Detección de textura

Para determinar la textura se emplean los siguientes descriptores basados en el histograma de intensidad:

- **Media:** Medida de intensidad media.

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (2.5)$$

Donde z_i es una variable randomica que indica la intensidad, $p(z)$ es el histograma de los niveles de intensidad en una región, y L es el número de niveles de intensidad posibles.

- **Desviación estándar:** Medida de contraste medio.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - m)^2} \quad (2.6)$$

Donde N es el tamaño de la población y m es la media aritmética de la población.

- **Suavidad:** Mide la suavidad relativa de intensidad en una región. Si la región es igual a 0, la intensidad en la región es constante; si la región es igual a 1 se determina que tiene grandes niveles de intensidad.

$$R = 1 - 1/(1 + \sigma^2) \quad (2.7)$$

- **Tercer momento:** Mide la asimetría de un histograma. Si el histograma es simétrico toma el valor de 0.

$$u_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \quad (2.8)$$

- **Uniformidad:** Esta medida es máxima cuando la imagen tiene los mismos niveles de intensidad.

$$U = \sum_{i=0}^{L-1} p^2(z_i) \quad (2.9)$$

- **Entropía:** Es la cantidad de información que se dispone en la imagen.

$$e = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (2.10)$$

2.9 Segmentación

El objetivo de la segmentación es identificar y separar los objetos de interés de su entorno. La segmentación basada en regiones divide la imagen en particiones con características similares, permitiendo detectar un objeto.

La similitud entre las máscaras extraídas en la investigación y las máscaras de la base de datos es determinada mediante el coeficiente de Jaccard. El coeficiente de Jaccard emplea la razón del conjunto intervalo al conjunto de unión para obtener una métrica de similitud, si no

existen elementos que se intercepten la métrica será de cero, caso contrario si todos los elementos se interceptan la métrica será igual a uno (Reyes & Torres, 2009).

El coeficiente de Jaccard se identifica como:

$$I_j = \frac{c}{a + b - c} \quad (2.11)$$

Donde a, b y c, son los números de elementos presentes en cada región.

2.10 Aprendizaje de máquina

El aprendizaje de máquina (*Machine Learning*) es una rama de la inteligencia artificial que busca realizar procesos que permitan al ordenador aprender o generalizar comportamientos, para poder desarrollar tareas en función de parámetros matemáticos de forma autónoma, basándose en la experiencia que ha obtenido con información no estructurada suministrada en forma de ejemplos (Padilla, 2010) (Mitchell, 1997).

Existen múltiples tipos de modelos de aprendizaje autónomo. Uno de estos modelos es el aprendizaje supervisado binario, que emplea el conjunto de patrones de las imágenes de la base de datos y de la clase o diagnóstico de las imágenes.

Dentro del aprendizaje autónomo se cuenta con una amplia lista de algoritmos, sin embargo, en el presente trabajo se emplean SVM y árbol de decisión. SVM es un algoritmo que brinda una mejor exactitud y generalización pese a que cuenta con una base de datos limitada y proporciona un alto espacio de dimensión. El algoritmo de árbol de decisión se emplea ya que trabaja en el área de la biología y la medicina (Padilla, 2010) (Moscovitz & Rengifo, 2010) (Martínez, 2006) (Betancourt, 2005) (Saegerman, 2004) (Crichton, 1997) (Álvarez, 2012) (Perrier, 2003).

2.10.1 Máquina de soporte vectorial (SVM)

Las SVM (*Support Vector Machine*) resuelven problemas de clasificación binaria permitiendo separar las clases, minimizar el error y maximizar el margen de separación, por tal razón se denominó hiperplano óptimo de separación (OSH- *Optimal Separating Hyperplane*), el cual se basa en realizar una transformación de los puntos de entrada a un espacio con características de dimensión superior.

Las SVM se dividen en: separable linealmente, no separable linealmente y no lineal.

La SVM separable linealmente se puede dividir a través de una frontera lineal delimitando a los datos, para este proceso cada punto utilizado en el entrenamiento es etiquetado y pertenece a una de dos clases. Se requiere determinar un hiperplano que cometa un menor número de errores y maximice el margen de la clase, a través del mapeo del espacio de entrada al espacio de características, de tal forma que permita separar cada punto dependiendo de su etiqueta, esto se indica en la Figura 13. La solución del hiperplano se realiza combinando los puntos de entrada, y para maximizar el margen de la clase se introduce el multiplicador de LaGrange.

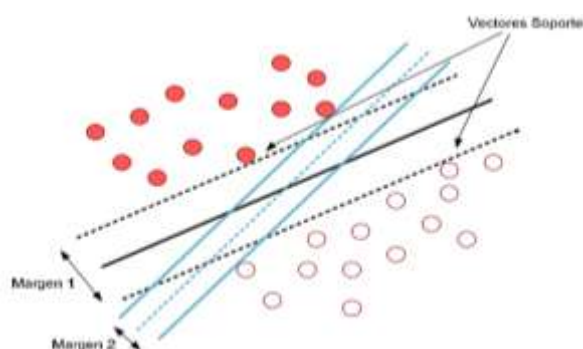


Figura 13 Muestras separables linealmente con una maximización del margen
Fuente. (Padilla, 2010)

La SVM no separable linealmente tienen muestras que no son separables, para este caso se puede buscar una frontera no lineal, o tratar de buscar un hiperplano que permita tener menos restricciones en el espacio de características, teniendo como resultado una frontera de clasificación no lineal que tenga dimensiones superiores, como se indica en la Figura 14.

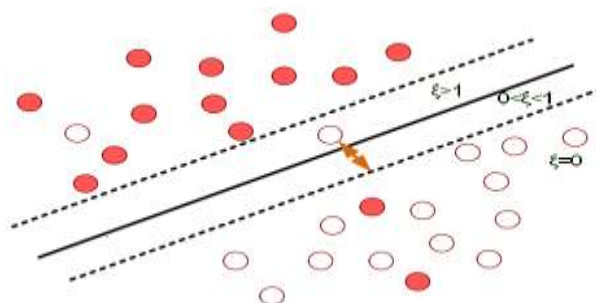


Figura 14 Muestras no separables linealmente
Fuente. (Padilla, 2010)

La SVM no lineal se interpreta como un caso de generalización del hiperplano óptimo de decisión, permitiendo así resolver el problema de no ser separables y se puede trazar fronteras de clasificación no lineal, esto se muestra en la Figura 15 (Padilla, 2010). En el caso de que los datos no sean linealmente separables, se realiza la transformación no lineal a un espacio de dimensión mayor con el fin de aumentar las posibilidades de obtener un conjunto de datos separables, esto se justifica a través del Teorema de Cover (Cover, 1965). Para construir un SVM debe ser un Espacio Hilbert denotado por:

$$K(x, x_i) = \langle \phi(x), \phi(x_i) \rangle \quad (2.12)$$

Donde K que se denomina función núcleo o función Kernel, a través de la cual es posible aplicar el algoritmo de entrenamiento de las SVM sin la necesidad de conocer ϕ .

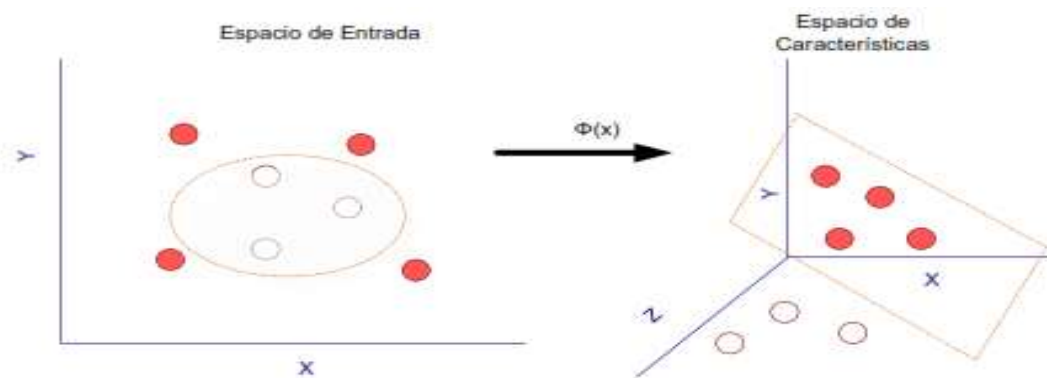


Figura 15 Transformación del espacio de entrada
Fuente. (Padilla, 2010)

Las SVMs se pueden adaptar a diferentes funciones Kernel, las más empleadas son:

- Kernel lineal:

$$K(x, x_i) = x \cdot x_i \quad (2.13)$$

- Kernel polinómico:

$$K(\bar{x}, \bar{y}) = (\bar{x} \cdot \bar{y} + 1)^p \quad (2.14)$$

Donde p es el grado del polinomio.

- Kernel Gaussiano:

$$K(\bar{x}, \bar{y}) = \exp\left(-\|\bar{x} - \bar{y}\|^2 / 2\sigma^2\right) \quad (2.15)$$

Donde $\gamma(\sigma)$ permite definir la escala del parámetro del Kernel Gaussiano. γ se relaciona con la extensión de los puntos de datos. Si estos puntos están alejados el uno del otro (si el punto y se aleja del punto x) el kernel tiende a decrecer rápidamente.

Para obtener una clasificación con un alto grado de generalización, se debe evitar dos efectos típicos del entrenamiento que son: el sobreajuste y el subajuste. El sobreajuste se describe cuando los parámetros obtenidos en el entrenamiento, se ajustan con mucha exactitud a las muestras. Por otro lado, el subajuste se describe cuando el hiperplano trazado es muy simple y no permite tener una buena generalización, esto sucede cuando se han empleado muy pocas muestras de entrenamiento, por lo tanto, la máquina no genera buenos resultados al ingresar muestras nuevas. Esto se muestra en la Figura 16.

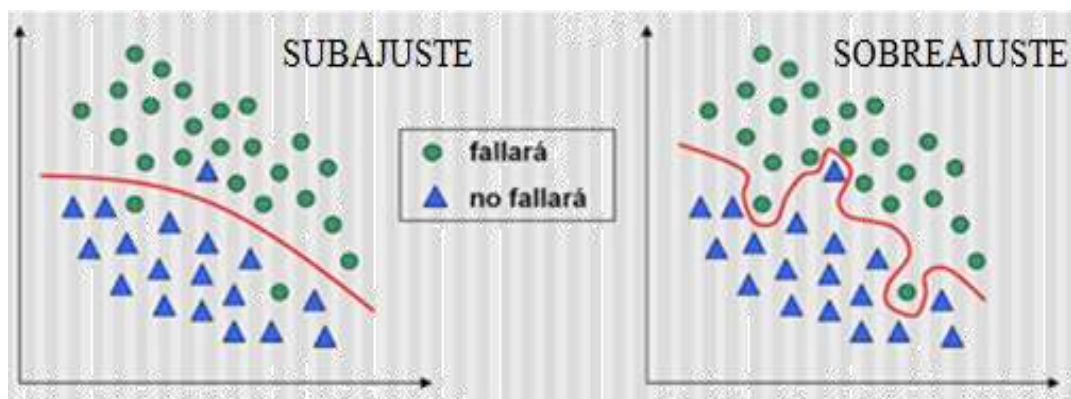


Figura 16 Sobreajuste y Subajuste

Fuente. (Carbon, 2017)

2.10.2 Árbol de decisión

Un árbol de decisión es un algoritmo recursivo de clasificación jerárquico cuya función es descomponer una decisión compleja en múltiples decisiones simples. Este modelo se descompone en una serie de registros o nodos interiores, cada uno de ellos representa una decisión tomada en función del valor de la variable constituida por valores predictivos y de resultado.

El funcionamiento de este aprendizaje de máquina emplea criterios de optimización y dos etapas continuas: la etapa de selección de subconjuntos y la etapa de evaluación del funcionamiento. Para crear el árbol de decisiones se realiza una sucesiva iteración de preguntas, por ejemplo (a_1, a_2, b_1, b_2) , permitiendo buscar las variables, para este ejemplo $(X_1$ y $X_2)$, cuyo valor divide al conjunto de muestras en subconjuntos homogéneos o $(W_1, W_2$ o $W_3)$, tal como se observa en la Figura 17. Para obtener el valor máximo de las muestras se emplea el índice de Gini o la entropía. El proceso se realiza sucesivamente en todos los subconjuntos hasta lograr la clasificación de todas las muestras, o hasta alcanzar un umbral de separación que se determine como aceptable (Álvarez, 2012).

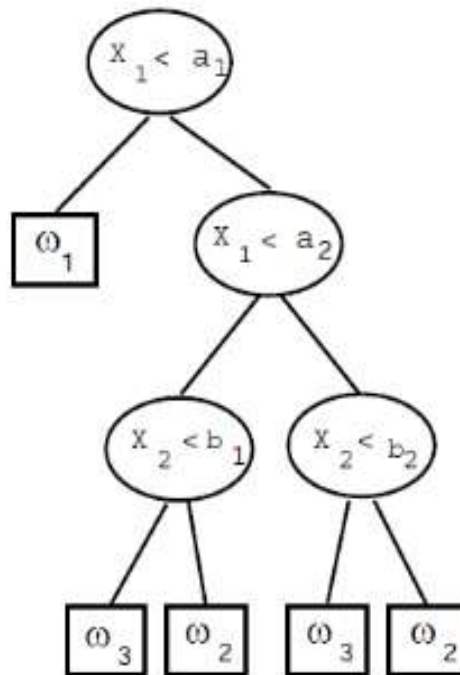


Figura 17 Ejemplo de árbol de decisiones
Fuente. (Álvarez, 2012)

Se realiza la selección de características a través de árbol de decisión tipo CART (*Classification and Regression Trees*) que forma parte de la distribución de Matlab (función *classregtree*). El proceso de desarrollo del clasificador se basa en la construcción, parada, podado y selección de clasificador óptimo. El proceso inicia en el nodo raíz donde se encuentra la variable que representa la clasificación; para determinar esta variable se mide su pureza y se asegura la pureza de los nodos hijos (Trujillano & Sarria, 2008) (Álvarez, 2012).

2.10.3 Reducción de características

Se tiene varias técnicas de reducción de características o dimensionalidad. Estas son utilizadas para realizar la transformación lineal del espacio original a un nuevo espacio.

PCA es una herramienta matemática de análisis de datos, que utiliza una transformación ortogonal para convertir de un conjunto de variables correlacionadas a un conjunto de variables no correlacionadas, maximizando la varianza de los datos. Su objetivo es eliminar el ruido y los datos redundantes, obteniéndose así los componentes principales. Se utiliza en la neurociencia, reconocimiento de patrones e imágenes, entre otras. Proporciona un método que parte de un conjunto de datos complejos y genera un nuevo espacio de datos con dimensión menor.

PCA realiza una transformación lineal del espacio de medidas definidas con dimensión n , a un nuevo espacio con dimensión m , donde $m \leq n$, a través de la matriz de transformación $U = [u_1 \ u_2 \ u_3 \ \dots \ u_m]_{n \times m}$, donde las columnas son linealmente independientes y forman un espacio ortogonal. Esta transformación lineal determina estructuras simplificadas que contengan información relevante al buscar proyecciones de datos representados en términos de mínimo cuadrado. Se elige un nuevo sistema de coordenadas para el conjunto de datos originales, donde se tiene que la mayor varianza se encuentra en el primer eje, la segunda mayor varianza se encuentra en el segundo eje, y así sucesivamente. (Romón & Enriquez, 2011), (Abril de Mur, 2017), (Zaccaro, Cárdenas, & Salazar, 2011), (Barra, 2008).

2.11 Métodos de validación de resultados

Existen dos métodos para evaluar los resultados del clasificador. El primero consiste en evaluar teóricamente el error en base de las características del algoritmo de clasificación. El

segundo método consiste en analizar un conjunto de áreas test obtenidas de la misma manera que las áreas de entrenamiento. En este método se obtiene una estimación más realista de los errores siempre y cuando la muestra de pixels para la estimación del error sea suficientemente grande (Kohavi, 1995) (Benitez, 2008) (Trujillano, March, & Sorribas, 2004).

2.11.1 Validación cruzada k-fold

La validación cruzada *k-fold* o de *k* iteraciones se utiliza para evaluar, validar y estimar la exactitud de los métodos de clasificación o predicción. Consta de un modelo generado a través de la media aritmética de diferentes particiones.

La validación divide las muestras en dos conjuntos: uno de entrenamiento (*training set*) y otro de verificación (*test set*), para posteriormente dividirlos en *k* subconjuntos equitativos. El conjunto de entrenamiento realiza la función de ajuste de pesos. El conjunto de verificación decide cuándo detener el proceso de entrenamiento, es decir cuando el error es mínimo, permitiendo determinar si el sistema es capaz de predecir correctamente los resultados.

Este modelo de validación realiza una combinación de los datos de entrenamiento y verificación, donde *k-1* subconjuntos son de entrenamiento y uno es de validación, esta metodología es empleada de manera que todos los subconjuntos hayan sido de entrenamiento y de verificación. En la Figura 18 se muestra un ejemplo de validación cruzada con *k=3* iteraciones, por lo tanto se divide los datos en 3 subconjuntos equitativos y se realiza la combinación (Devijver, 1982) (Refaeilzadeh, Tang, & Liu, 2009) (Trujillano, March, & Sorribas, 2004).

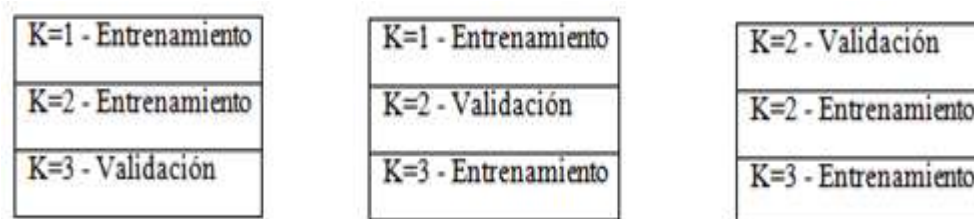


Figura 18 Ejemplo de validación cruzada K-fold

2.11.2 Matriz de confusión

La matriz de confusión permite visualizar la información, obtenida de los resultados de la clasificación, mediante una matriz conocida como tabla de contingencia de la distribución, tal como se observa en la Tabla 1 (Corso, 2009).

Tabla 1

Descripción de matriz de confusión para dos clases

Matriz de Confusión		Clase Predicha	
		Negativo	Positivo
Clase Actual	Negativo	VN	FP
	Positivo	FN	VP

Donde:

- **VN:** Verdaderos negativos, son el número de predicciones correctas al ser la respuesta negativa.
- **VP:** Verdaderos positivos, son el número de predicciones correctas al ser la respuesta positiva.
- **FN:** Falsos negativos, son el número de predicciones incorrectas al ser la respuesta negativa.
- **FP:** Falsos Positivos, son el número de predicciones incorrectas al ser la respuesta positiva.

A partir de la matriz de confusión se pueden obtener detalles que determinan qué tan eficiente es el sistema de clasificación (Hamilton, Findlater, & Olive, 2013) (Corso, 2009). Los atributos empleados para este propósito son:

- **Sensibilidad:** Permite determinar la capacidad del modelo del clasificador para detectar los verdaderos positivos.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \times 100 \quad (2.16)$$

- **Especificidad:** Se considera como el porcentaje de muestras sanas que se determinó correctamente.

$$\text{Especificidad} = \frac{VN}{FP + VN} \times 100 \quad (2.17)$$

- **Exactitud:** Es el resultado de la proporción entre los valores de los verdaderos positivos y los verdaderos negativos para el total de casos examinados.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \times 100 \quad (2.18)$$

2.11.3 Las curvas ROC

Las curvas ROC (*Receiver Operating Characteristic*) son una forma de visualización gráfica del análisis de todos los posibles puntos de corte que provienen de la matriz de confusión. Representan los pares de especificidad y sensibilidad, donde el eje X indica la tasa de falsos positivos y el eje Y indica la tasa de los verdaderos positivos, tal como se muestra en la Figura 19. Estas curvas permiten evaluar el desempeño del clasificador a través del área bajo la curva (AUC), expresada en la ecuación 2.19, convirtiéndose en un indicador de capacidad predictiva del modelo. Si el área está más cerca de la diagonal (área cercana a 0.5) el modelo tendrá una menor precisión, pero si al área se aproxima a 1 el modelo se considera como "óptimo" (Amorim, Corominas, & Zarzosa, 2009).

Las curvas ROC se localizan en función de la importancia de las medidas, lo que hace que cada una tenga un grado de solapamiento mayor en términos de sensibilidad, exactitud, y especificidad respectivamente.

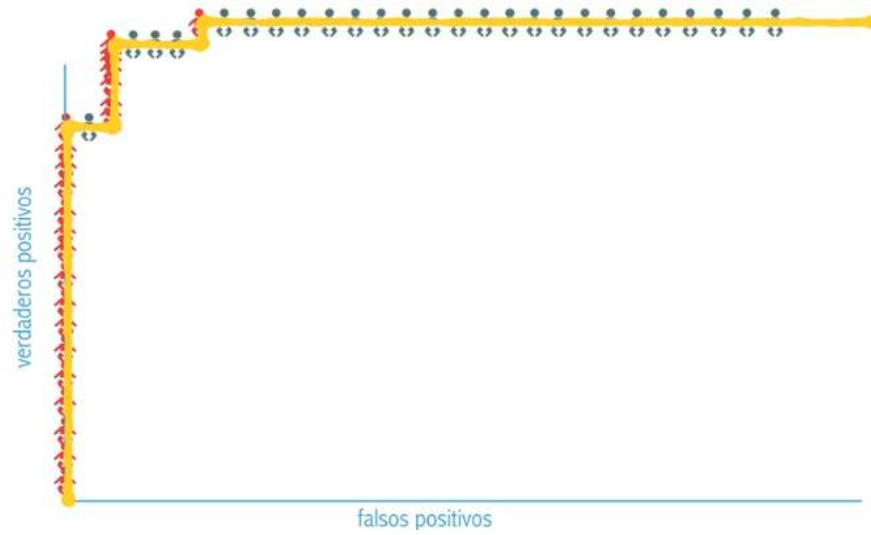


Figura 19 Ejemplo de la curva ROC, elementos rojos enfermos y elementos azules sanos
Fuente. (Cuello, 2016)

$$AUC = \int_0^1 ROC(t) dt \quad (2.19)$$

CAPÍTULO 3

MATERIALES, MÉTODOS E IMPLEMENTACIÓN

3.1 Base de datos

La base de datos empleada es obtenida de la “*International Skin Imaging Collaboration: Melanoma Project 2017*” (ISIC). El ISIC es un proyecto de la “*International Society for Digital Imaging of the Skin*”, cuyo fin es publicar un archivo de código abierto (base de datos) de imágenes de lunares, y de esta manera capacitar y ayudar a los médicos a mejorar las habilidades de diagnóstico sobre el cáncer de piel tipo melanoma. Las imágenes de la base de datos son una colección de diferentes estudios realizados en varias clínicas del mundo, y basan su diagnóstico mediante confirmación por biopsia, historial clínico y la determinación de expertos.

La base de datos *Skin Lesion Analysis Towards Melanoma Detection* de ISIC, fue seleccionada por su resolución, sus características, la existencia de la máscara de los lunares, y la cantidad de imágenes de melanomas malignos o benignos que se acoplan al sistema de detección desarrollado. La base de datos consta de 2000 imágenes a color de entrenamiento y prueba, de las cuales 374 son melanomas, 254 queratosis seborreica (tumor benigno de la piel, derivada de queratinocitos), y 1374 nevus benignos (tumor benigno de la piel, derivado de los melanocitos). Con el fin de obtener datos reales, en la presente investigación se emplea una base de datos equilibrada conformada por las 374 imágenes de melanoma y 374 imágenes seleccionadas al azar del grupo de nevus benignos, dando un total de 748 imágenes.

3.2 Matlab

Matlab es un software aplicativo matemático de cálculo numérico, tiene un lenguaje propio, y está orientado a trabajar con matrices. Matlab cuenta con bibliotecas, herramientas y aplicaciones internas que permiten realizar el entrenamiento, la verificación y la validación de datos o muestras. Las herramientas disponibles y el detalle de cada función se describen en el libro de (Gonzalez & Woods, 2002).

Para desarrollar el sistema de detección de cáncer de piel tipo melanoma, se emplea el *software* aplicativo Matlab en la versión 2015^a, ejecutado en computador portátil con las siguientes características: Procesador Intel(R) Core(TM) i5 de primera generación, velocidad del procesador de 2.53Ghz, memoria RAM de 8.00GB, sistema operativo Windows 7 de 64 bits.

3.3 Descripción del sistema

El sistema desarrollado cumple con los procesos descritos en las Figuras 20 y 21. Este inicia con la obtención de la imagen digital del lunar tomada desde la base de datos ISIC-Archive. Posteriormente la imagen digital ingresa a la etapa de pre-procesamiento, permitiendo acoplarla para poder extraer las características de interés. A continuación, se procede a realizar la localización, segmentación y cuantificación de los lunares en función de sus características. Las características elegidas ingresan a la etapa de clasificación, donde se utilizan los algoritmos de SVM y árbol de decisiones. Finalmente se realiza la evaluación de los resultados obtenidos mediante la matriz de confusión. Adicionalmente se determina la capacidad predictiva a través de la curva ROC.

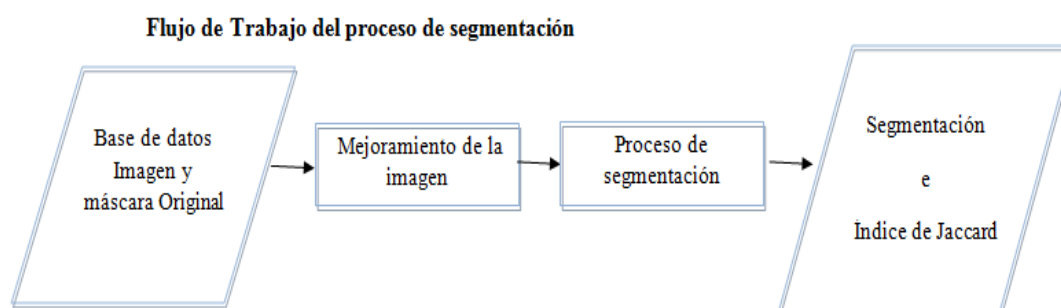


Figura 20 Evaluación del proceso de segmentación del lunar

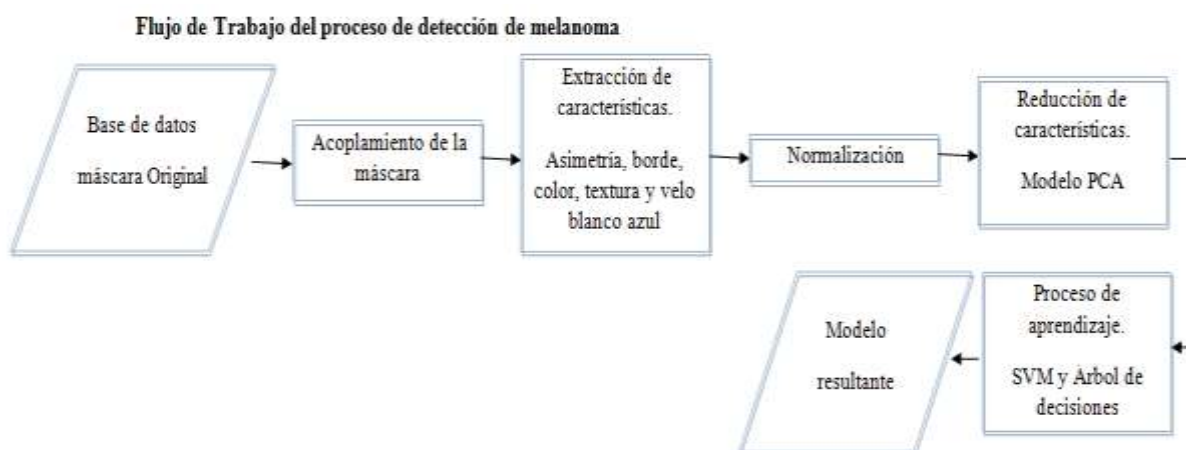


Figura 21 Proceso de detección de cáncer de piel tipo melanoma

3.4 Segmentación

El proceso de segmentación permite localizar el lunar dentro de una imagen. Además, ejecuta la tarea de separación y extracción de información relevante dentro de la imagen.

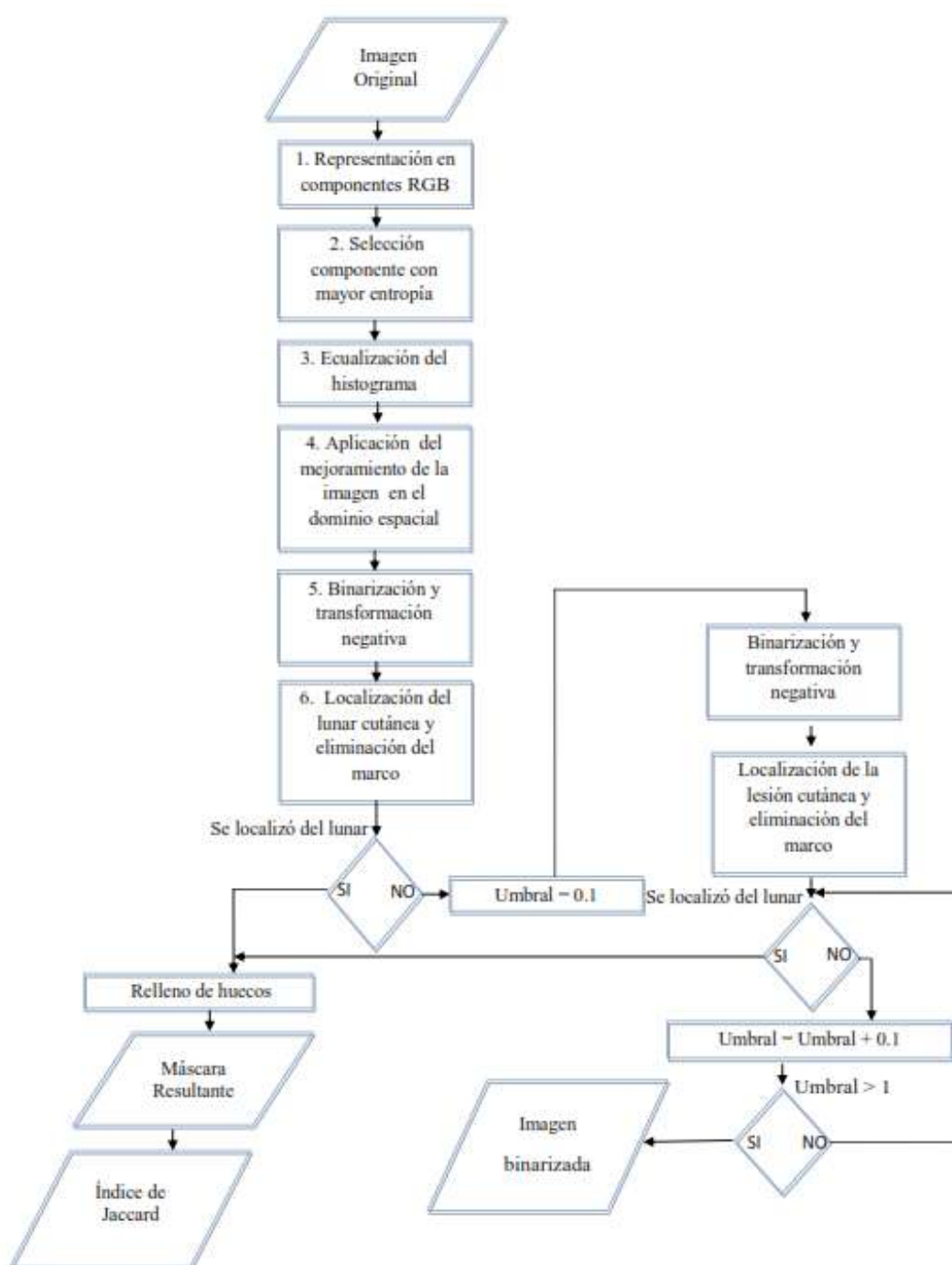


Figura 22 Evaluación del proceso de segmentación del lunar

El proceso de segmentación de la imagen de un lunar consta de seis pasos consecutivos. Adicionalmente se cuenta con tres bucles desarrollados para localizar el lunar en las imágenes que no son totalmente compatibles con el sistema. Se parte de la imagen original, hasta llegar a entregar la máscara resultante del lunar, obtener el índice de Jaccard o entregar la imagen binarizada. En la Figura 22 se observa el proceso de segmentación de un lunar detallado a continuación:

El proceso inicia con la representación de la imagen original en sus componentes básicos del modelo RGB. A continuación, se comparan las tres componentes y se devuelve la que posee mayor entropía. En la Figura 23 se muestra que el lunar que posee mayor información es la componente azul; confirmando que cada componente de la lesión cutánea es independiente.

Posteriormente se mejora el contraste de la imagen a través de la ecualización del histograma, permitiendo que todas las imágenes de la base de datos tengan la misma intensidad. Una vez ecualizado el histograma, se aplica la operación morfológica de dilatación, permitiendo que se elimine la información basura (vello o imperfecciones) al expandir el lunar.

Posteriormente se procede a la binarización empleando un umbral óptimo. Si no se logra localizar el lunar, la imagen entra a un sub-proceso donde el umbral va aumentando, desde un valor definido hasta detectar el lunar usando el proceso de localización. Si el umbral alcanza un valor de 1 sin poder localizar el lunar, el sub-proceso termina y se asigna como resultado la imagen binarizada con umbral automático. Una vez terminado el sub-proceso se obtiene una matriz binaria del mismo tamaño que la imagen, cuyos pixeles toman el valor de 1 en la región de la piel, y el valor de 0 en la región donde posiblemente se ubica un lunar.

Además, se realiza la transformación negativa de la imagen del lunar, para lo cual se invierte los valores, es decir, que la región de la piel tome el valor 0 y la del lunar tome el valor de 1. El objetivo este procedimiento es multiplicar la imagen resultante por la imagen original, para obtener solo el lunar y eliminar la mayor parte de la piel. En la Figura 24 se muestra la imagen binarizada del lunar aplicando transformación negativa, en la cual se observan elementos que no representan el lunar y deben ser eliminados. El primer paso del proceso es eliminar los marcos de la imagen; posteriormente se realiza una búsqueda en función de los objetos dispuestos en la imagen, para esto, se la divide en los posibles objetos encontrados y se identifica su centroide (centro de simetría del lunar) junto con su área.

Se concluye que el objeto es un lunar si se cumplen las dos validaciones: la primera validación define que su tamaño sea mayor a un valor específico, y la segunda es que el centroide de los lunares se debe localizar siempre en el centro de la imagen.

Al tener un resultado positivo del sub-proceso, se procede a llenar los huecos del lunar, eliminando así las discontinuidades dentro de la imagen.

De todo el procedimiento descrito se retorna la máscara del lunar, la cual se multiplica por la imagen original permitiendo obtener la segmentación de la imagen resultante mostrada en la Figura 25. Caso contrario se retorna la imagen binarizada.

Finalmente, se compara la máscara resultante con la proporcionada en la base de datos empleando el índice de Jaccard. Esto permite obtener el porcentaje de similitud para validar la segmentación.

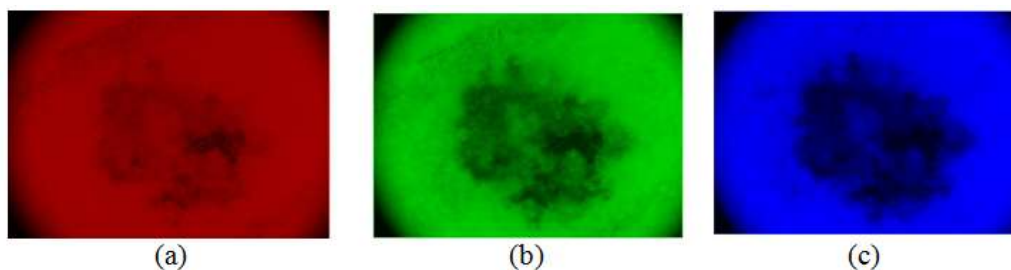


Figura 23 Componentes: (a) Roja, (b) Verde, (c) Azul



Figura 24 Lunar binarizado invertido

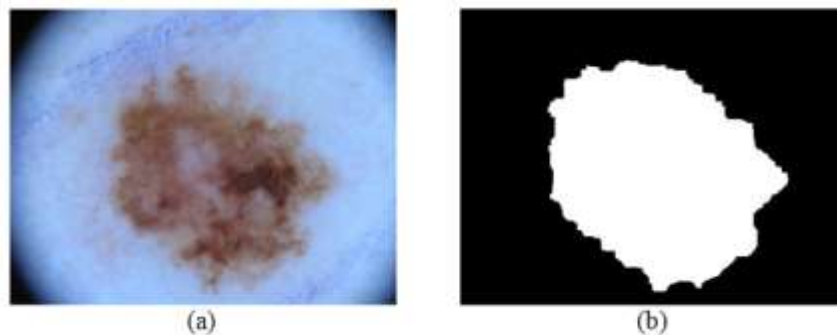


Figura 25 Resultado de la segmentación de la imagen: (a) Imagen original, (b) Imagen segmentada

3.5 Extracción de características

En esta sección se describen los procesos implementados para la extracción de las características de asimetría, bordes, color, textura, y velo blanco-azul, para su posterior uso en el entrenamiento y validación de los clasificadores.

El proceso de mejoramiento de la imagen, descrito en la Figura 26, parte de la lectura de las imágenes y las máscaras originales de la base de datos. Posteriormente se binariza la máscara y se realiza el cierre morfológico. A continuación, se realiza la selección del lunar, a través de la búsqueda de la mayor área. Finalmente se concatena cada componente RGB con un mismo tipo de datos, para seleccionar la zona de interés.

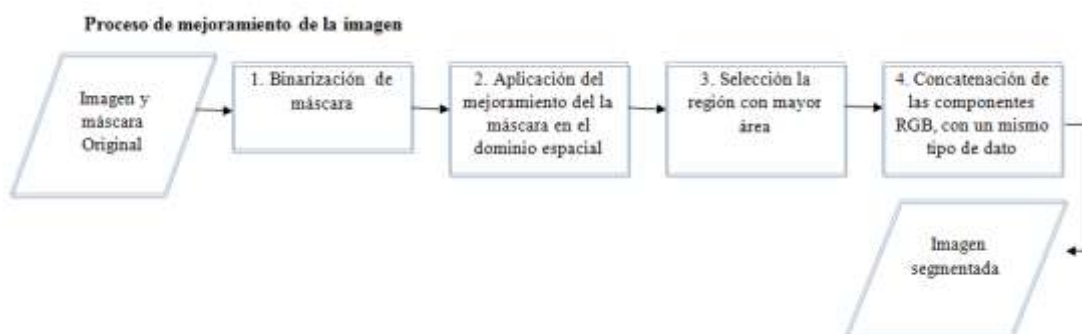


Figura 26 Proceso inicial para la extracción de características

A continuación, se detalla el proceso de extracción de las características que serán ingresadas en el clasificador.

3.5.1 Asimetría

Para extraer las características que representan la asimetría de un lunar se realizan dos procesos detallados en la Figura 27.

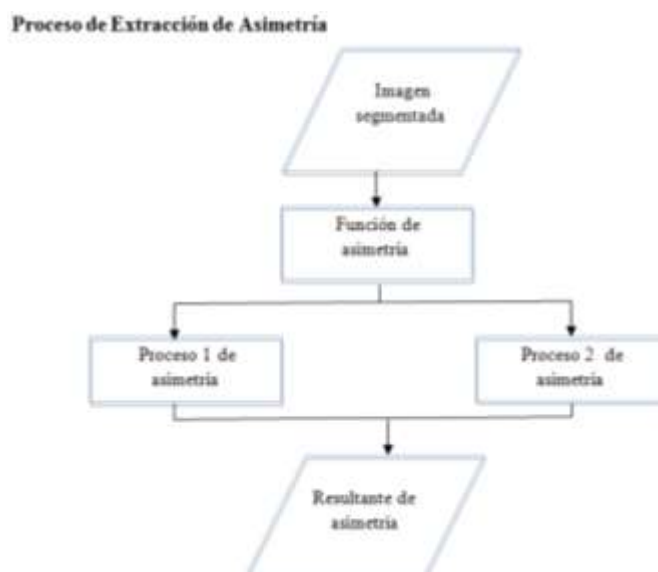


Figura 27 Proceso general de extracción de asimetría

Del proceso 1 se obtiene el número de veces que varía la distancia entre el centroide y los bordes de la imagen. Del proceso 2 se obtiene el valor de la distancia entre el centroide y el centro de peso. En las Figuras 28 y 29 se muestran los procesos empleados para la extracción de las características de asimetría.

El proceso 1 consta de cuatro pasos descritos a continuación:

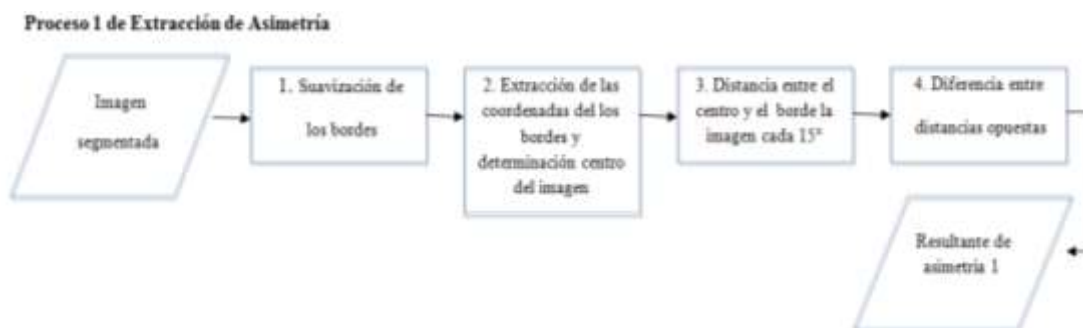


Figura 28 Proceso 1 de extracción de asimetría

El proceso inicia con la suavización de la imagen segmentada, evitando que los datos de respuesta tengan cambios bruscos. Se traza los bordes utilizando la media de un vector dado y se eliminan los posibles puntos innecesarios.

Luego de la suavización, se extraen las coordenadas de los bordes para conocer la distribución del lunar y determinar su centroide. A continuación, se procede a estimar la distancia entre el centro encontrado y el borde de la imagen. Para esto se divide el borde del lunar cada 15° en el rango de 0° a 360° .

Finalmente, se procede a determinar la relación entre la distancia existente cada 15° y su punto diametralmente opuesto. En este proceso se buscan variaciones bruscas que son cuantificadas dando una representación de la simetría del lunar.

El proceso 2 consta de tres pasos descritos a continuación:



Figura 29 Proceso 2 de extracción de asimetría

Se encuentra el centroide de la imagen segmentada. A continuación, se ubica el centro de peso del lunar, considerando los valores de intensidad de los píxeles, a partir de la máscara binarizada y la unión de la imagen en escala de grises.

Finalmente, se realiza el cálculo de la distancia entre el centroide y el centro de peso, tal como se muestra en la Figura 30. Si la distancia tiende a cero se considera que el lunar es simétrico, caso contrario el lunar es asimétrico. El valor resultante de la distancia es devuelto al proceso general de asimetría.



Figura 30 Distancia entre el centroide (círculo azul) y el centro de peso (punto rojo): (a) Lunar simétrico, (b) Lunar asimétrico

3.5.2 Diferencia de bordes

La diferencia de bordes está definida por la uniformidad del contorno del lunar, que a su vez depende de su color. Si el borde es marcado se mantendrá el color del lunar y se podrá diferenciar de la piel, como se puede ver en la Figura 31 (a). Caso contrario si el borde es difuso este irá perdiendo su color y se mezclará con el color de la piel, como se indica en la figura 31 (b).

Si un lunar tiene los bordes marcados o bien definidos se visualizarán tres imágenes similares de las componentes (RGB). Caso contrario, si el lunar presenta bordes difuminados existirá una diferencia entre dichas componentes. Ambos casos son representados a través de los valores del índice de Jaccard como se indica en la Figura 31 (a) y (b), donde se observa que al tener los bordes marcados el valor tiende a ser menor que cuando los bordes son difuminados.

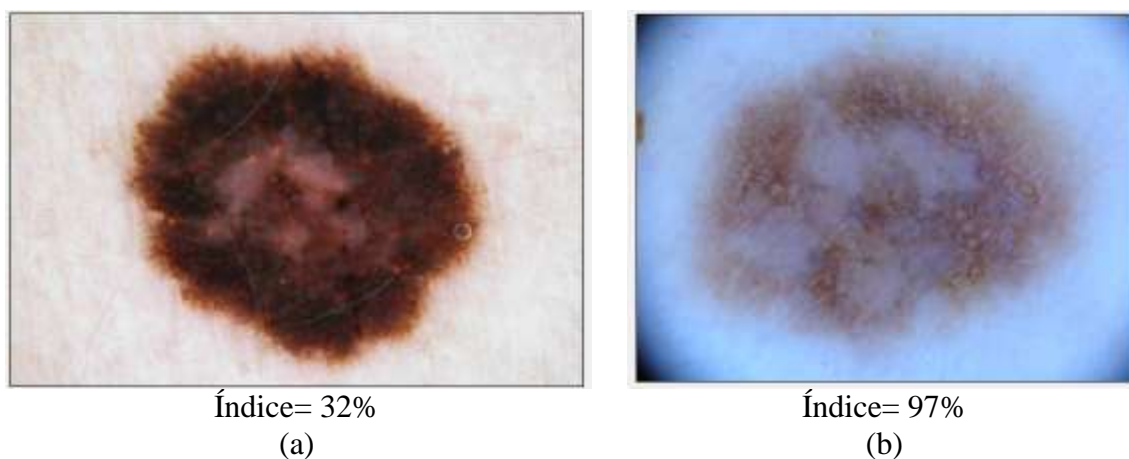


Figura 31 Diferencia de bordes: (a) Bordes definidos, (b) Bordes difuminados

El proceso para determinar la diferencia entre los bordes del lunar consta de cinco pasos detallados en Figura 32.

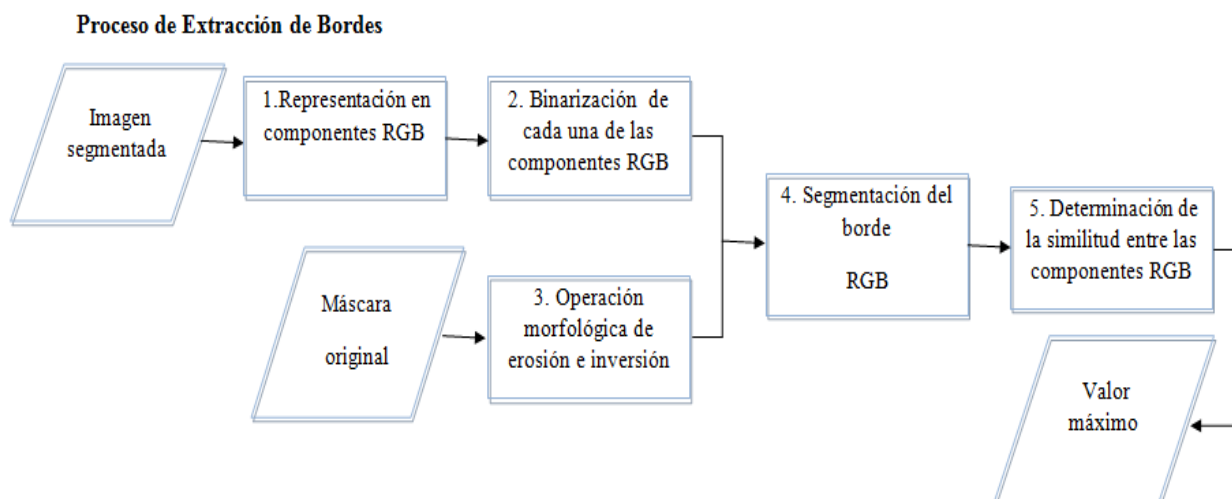


Figura 32 Proceso de extracción de bordes

El proceso inicia con la representación de la imagen segmentada del lunar en sus componentes básicos RGB. A continuación, se binariza cada una de estas componentes empleando un umbral óptimo.

Posteriormente se requiere extraer los bordes de importancia. Para esto se parte de la máscara del lunar, se aplica la operación morfológica de erosión y se invierten los valores obtenidos.

A continuación, se combinan los dos resultados anteriores para obtener tres nuevas componentes que representan los bordes del lunar, como se indica en la Figura 33 (b), (c) y (d). Además, se realiza el enmarcado del lunar para que el índice de Jaccard no contemple los píxeles que corresponden a la piel (píxeles negros).

Finalmente, se compara la similitud entre los compontes RGB empleando el índice de Jaccard, como se observa en la Figura 33 (a). Se obtiene el porcentaje máximo de similitud entre dichos compontes cuando el lunar presenta bordes desiguales, irregulares o mal definidos.

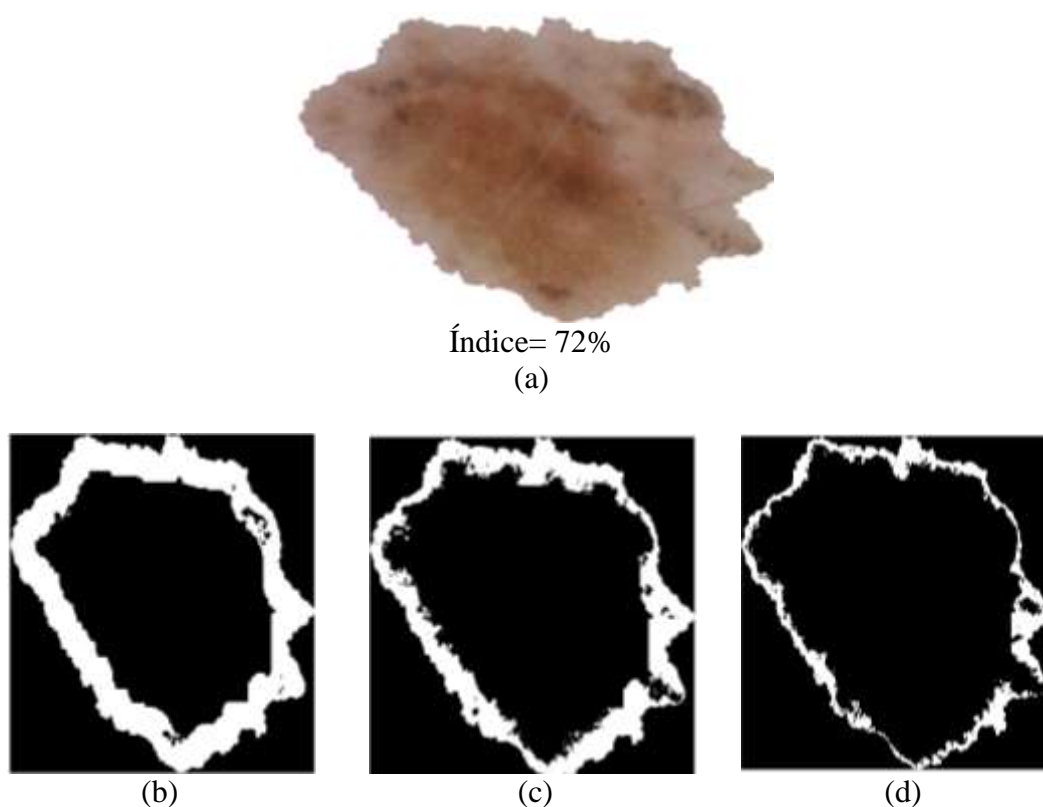


Figura 33 Diferencia de bordes: (a) Imagen original e índice de Jaccard, (b) Componente R, (c) Componente G, (d) Componente B

3.5.3 Color medio

La característica de color provee tres valores al clasificador. El proceso de extracción de color, mostrado en la Figura 34, toma la imagen segmentada y la representa en sus componentes RGB. Posteriormente, de cada componente se recorta el contorno del lunar y se selecciona los pixeles diferentes de cero. Finalmente se calcula el valor medio para obtener la intensidad de los tres componentes.

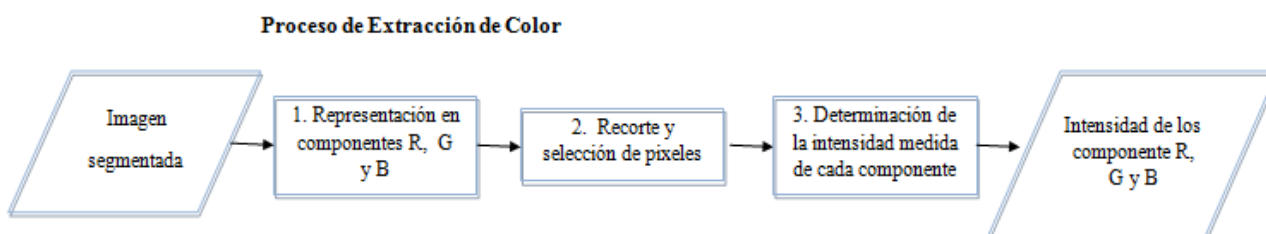


Figura 34 Proceso de extracción de color medio

El proceso de extracción del color del lunar tiene una gran importancia en la detección del melanoma, ya que un lunar normal debe ser una mancha de color uniforme, sin presencia de sombras de color marrón, rosada, roja, azul o blanca.

3.5.4 Textura

El proceso de extracción de textura del lunar consta de cuatro pasos, descritos en la Figura 35.

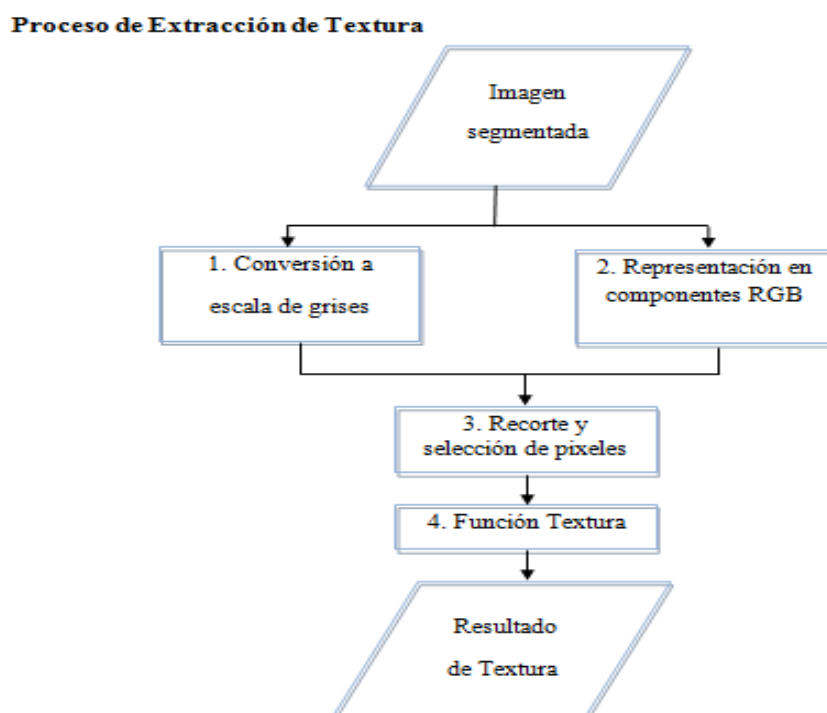


Figura 35 Proceso de extracción de textura

Debido que la función de textura trabaja únicamente en 2 dimensiones, se realiza la conversión de la imagen segmentada a escala de grises o se trabaja sobre cada componente R, G, o B, en función del número de características que se desean extraer. En la Figura 36 (a) y (b) se muestran dos imágenes de lunares con texturas diferentes, y en la Figura 36 (c) y (d) su representación respectiva.

Posteriormente, tanto para escala de grises o componentes RGB, se recorta el contorno del lunar y se busca los píxeles diferentes de cero. El resultado de la función textura es descrita en

seis valores, tomados de los cálculos detallados en (Gonzalez & Woods, 2002), que son: La media, que representa la intensidad promedio; la desviación estándar, que describe el contraste promedio; la suavidad o la variación del lunar a través del histograma; el tercer momento, que cuantifica la simetría o no del histograma en función de las repeticiones; la uniformidad, que describe la intensidad del lunar; y la entropía, que representa la cantidad de información presente en el lunar. Estos parámetros han sido descritos en las ecuaciones 2.5-2.10 del Capítulo 2 de la presente investigación.

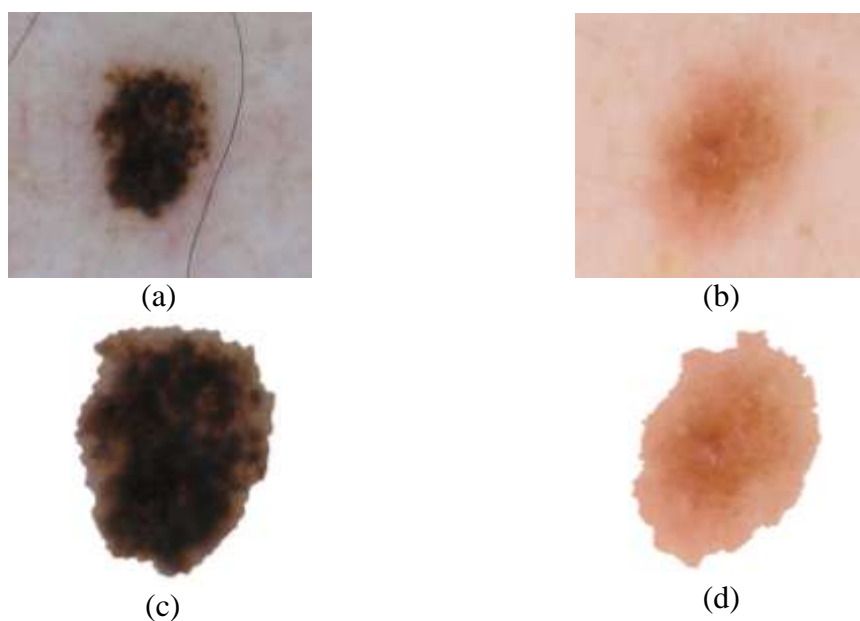


Figura 36 Textura: (a) Imagen con textura fuerte, (b) Imagen con textura suave, (c) y (d) Imagen recortada

Los valores tabulados que representan la textura del lunar de la Figura 36 (c) y (d) se muestran en la Tabla 2. Se observa que los valores de la media, desviación estándar, suavidad, y uniformidad tienden a ser más altos cuando se trata de una imagen uniforme. Por otro lado, los valores de la entropía y tercer momento tienden a ser menores en el mismo caso. Finalmente, se devuelven los valores de los seis parámetros de la textura al clasificador.

Tabla 2
Valores de textura

Figura	Media	Desviación estándar	Suavidad	Tercer momento	Uniformidad	Entropía
44 (a)	32.7195	26.8951	0.0110	0.3586	0.0515	5.4634
44 (b)	76.6639	53.0961	0.0416	-2.8722	0.0805	5.6839

3.5.5 Velo blanco-azul

La característica de velo blanco-azul se define mediante la relación entre el volumen de velo blanco-azul comparado con el volumen completo del lunar. El proceso consta de tres pasos generales descritos en la Figura 37.

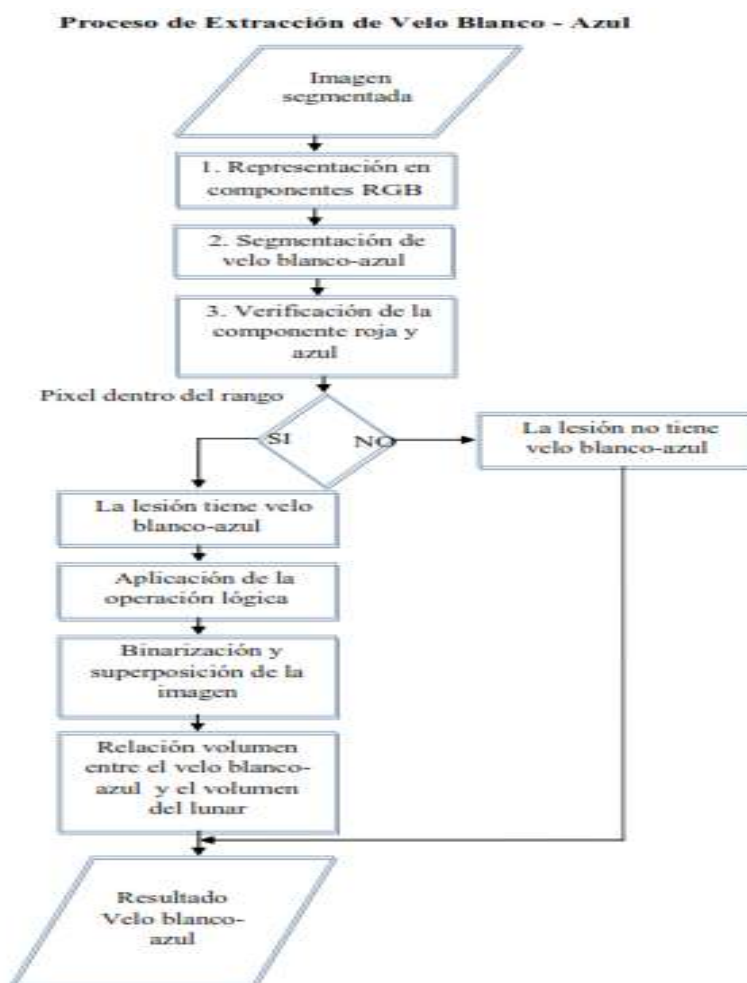


Figura 37 Proceso de extracción de estructura velo blanco-azul

A partir de la imagen segmentada, se extraen las componentes RGB. Si un pixel está dentro de un rango predefinido de intensidad se considera parte de velo blanco-azul. Se emplea únicamente la componente roja y azul, ya que después de varias pruebas se encontró que su diferencia estructural alcanza valores de intensidad entre 80 a 170 para las dos componentes. Para detectar el velo blanco-azul se procede a realizar la operación lógica *AND* entre estos dos resultados, como se muestra en la Figura 38.

Posterior a la detección y segmentación del velo blanco-azul se entrega una imagen binarizada con su respectiva máscara, la cual para visualización, demostración y comprobación se superpone en la imagen original con un color rojizo. En la Figura 39 (a) se observa la imagen original que presenta velo blanco-azul en su estructura, y en la Figura 39 (b) se observa la misma imagen luego de haber pasado por el proceso de detección.

Finalmente se determina la relación entre el volumen de velo blanco-azul y el volumen total del lunar, de la siguiente manera:

$$Rv = \frac{\text{Volumen del velo blanco_azul}}{\text{Volumen total del lunar}} \quad (3.1)$$

Si el valor de la relación es menor a uno, el lunar posee velo blanco-azul, característica que es ingresada al clasificador.

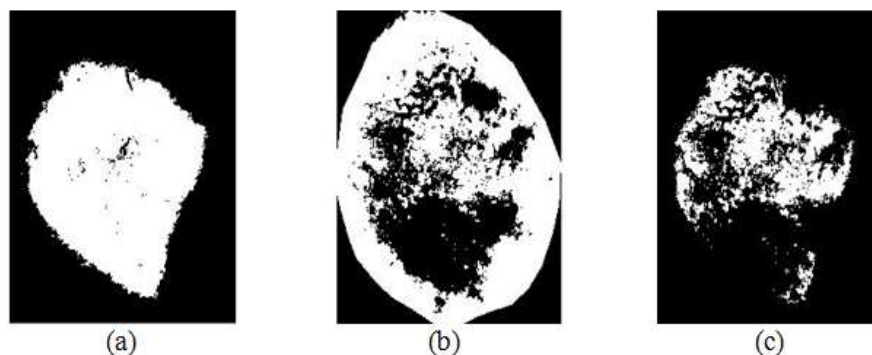


Figura 38 Operación lógica *AND*: (a) Componente roja, (b) Componente azul, (c) Intersección entre la componente roja y azul



Figura 39 Velo blanco-azul: (a) Imagen original, (b) Representación del resultado del velo blanco-azul

3.6 Normalización y características

El sistema de detección de cáncer de piel tipo melanoma requiere la extracción de ciertas características, que permiten determinar si un lunar es benigno o maligno. En la presente investigación se extrajeron un total de 29 características, 28 de las cuales son predictores y fueron tomadas a partir de los procesos aplicados en las imágenes. La última característica corresponde al diagnóstico clínico de cada lunar, fue proporcionada por la base de datos y se utiliza como respuesta del clasificador.

Las 28 características, detalladas en la Tabla 3, poseen valores en distintas escalas numéricas y siguen diferentes distribuciones estadísticas. Por esta razón deben ser normalizadas con una distribución de media cero y variancia unitaria, para permitir entrenar y evaluar el sistema. Un procedimiento muy sensible a valores pequeños es la normalización *Z-Core* mostrada en la ecuación 3.2. Esta normalización utiliza la media (μ) y la desviación estándar (σ) para calcular el nuevo valor (v) (Berrocal, Figuerola, & Ángel, 2006) (Farrús, Anguita, & Hernando, 2007) .

$$v = \frac{v - \mu_x}{\sigma_x} \quad (3.2)$$

Se definen las siguientes características:

- 28 características analizadas al sumar: 2 características de asimetría, 1 de borde, 24 de textura en escala de grises y componentes RGB, y 1 de velo blanco-azul.
- 22 características analizadas al sumar: 2 características de asimetría, 1 de borde, 18 de textura en componentes RGB y 1 de velo blanco-azul.

- 13 características analizadas al sumar: 2 características de asimetría, 1 de borde, 3 de color, 6 de textura en escala de grises y 1 de velo blanco-azul.

Tabla 3

Características extraídas para la clasificación

CARACTERÍSTICAS		
Nombre	Definidas	Descripción
Asimetría	2	Se valora la característica considerando si existe simetría en el lunar, mediante el siguiente proceso: a) Se divide la lesión cada 15° en el rango 0 °- 180°, analizando los bordes de cada mitad del lunar, b) Se mide la distancia entre el centroide y el centro de peso.
Borde	1	Se valora la característica considerando si los bordes son irregulares, dentados o mal definidos. Para esto se emplea el índice de Jaccard en las componentes RGB de la imagen.
Color	3	Se analiza los compontes RGB de la imagen y se valora sus características considerando si los colores son uniformes o diferentes.
Textura	6- En escala de grises 18- En componentes RGB	Se valora la característica considerando si los parámetros se encuentran dentro del rango de lesiones benignas o malignas. Para esto se determina la media, desviación estándar, suavidad, tercer momento, uniformidad y la entropía, en escala de grises y en cada componente RGB.
Velo blanco - azul	1	Se valora la característica considerando si existe velo blanco-azul. Para esto se realza el color azul y se obtiene la relación de volumen.

Para verificar el comportamiento del sistema se identificaron un total de 28 características, las cuales se fueron reduciendo en función de los parámetros que representa la textura que son: la media, desviación estándar, suavidad, tercer momento, uniformidad y entropía. Ya que estas características de textura son obtenidas en escala de grises y en cada componente R,G y B.

Se determinó que al emplear un mayor número características se obtiene un mayor valor en la sensibilidad, sin embargo, como se observa en la Figura 40, se produce un efecto inversamente proporcional para los valores de la exactitud.

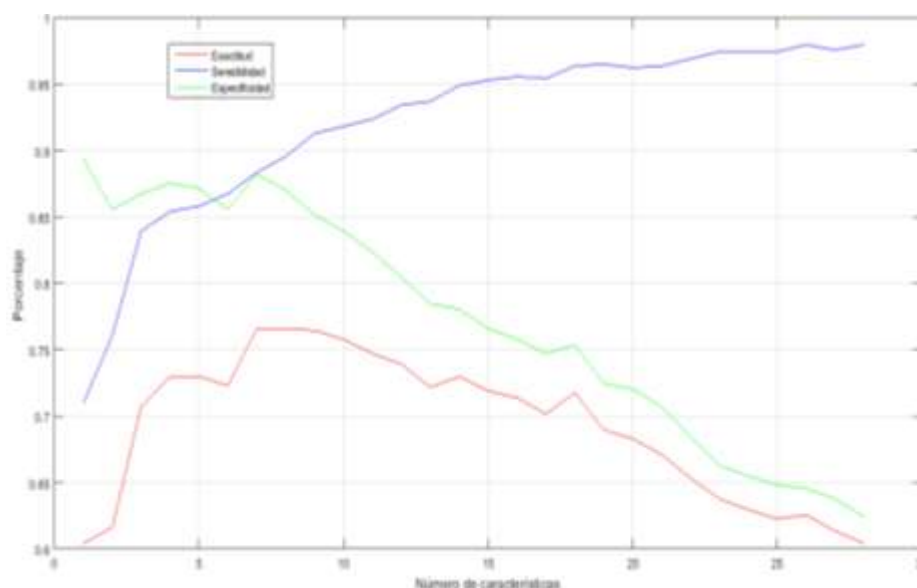


Figura 40 Influencia de la características en el sistema

Para definir las características relevantes se utiliza el p-valor (pv). El p-valor es una medida probabilística directa de lo verosímil que resulta una muestra. Este valor corresponde al nivel mínimo de aceptación o de riesgo (α) de una hipótesis. La decisión se realiza en función de la siguiente regla:

$$\text{Si } pv \leq \alpha, H = 1 \quad (3.3)$$

$$\text{Si } pv > \alpha, H = 0 \quad (3.4)$$

El p-valor está definido por la estadística de contraste indicada en la ecuación 3.5.

$$Z^* = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (3.5)$$

Donde:

\bar{x} es la población que sigue una distribución cualquiera, μ es la media, σ es la desviación estándar y n es el número de muestras (Vila, Sedano, & Anna, 2003).

Tabla 4

P-VALOR de las 28 características para un $\alpha=6\%$

Características	P-VALOR	Hipótesis H=1/H=0
1	0,01459647	1
2	0,00164986	1
3	2,89E-20	1
4	2,52E-31	1
5	0,96228494	0
6	0,01095878	1
7	0,01095878	1
8	0,01492317	1
9	0,28934757	0
10	0,05634657	1
11	0,18583306	0
12	0,53806813	0
13	0,53806813	0
14	0,00038101	1
15	0,88825842	0
16	0,35577974	0
17	0,00020482	1
18	0,30547425	0
19	0,30547425	0
20	0,00015733	1
21	0,02211855	1
22	0,02575453	1
23	0,03577128	1
24	1,35E-07	1
25	1,35E-07	1
26	0,37093307	0
27	4,14E-06	1
28	0,04617404	1

En la tabla 4, se indica el p-valor de las 28 características, donde se visualiza que 18 características son influyentes para el sistema de detección, ya que cumplen con la hipótesis planteada en la ecuación 3.3, y 10 aportan un mínimo de información.

3.7 Configuración de los clasificadores

Para determinar la correcta discriminación de características y los umbrales de decisión, se emplean las técnicas de aprendizaje de máquina SVM, y árbol de decisiones bajo el modo supervisado.

A continuación se detalla el proceso de configuración en el clasificador SVM:

- Se ejecuta la función *classificationLearner* en Matlab
- Se importa el conjunto de entrenamiento conformado por los predictores y la respuesta.
- Se determina el modelo de validación, para este caso la validación cruzada *k-fold* que divide los datos en el conjunto de entrenamiento y el conjunto de validación.
- Se determina el tipo de clasificador que se va a emplear conjuntamente con el tipo de kernel que se va a utilizar.
- Se ajusta los parámetros disponibles en clasificador y del kernel en función de los resultados que se desean.
- Se procede a entrenar obteniendo como respuesta la matriz de confusión, la curva ROC y el modelo del clasificador.

Los parámetros configurados en el clasificador SVM son: el Kernel, la escala, los predictores, el número de iteraciones que definen al método *k-fold* de la validación cruzada y el tipo de respuesta.

En el clasificador SVM se evalúa el rendimiento de los Kernels de tipo lineal, polinómico y Gaussiano, teniendo como resultado que el Kernel Gaussiano fino presta y define las mejores características.

Se configura la escala del clasificador SVM en el rango de 0.1 a 20 en saltos de 0.1. Se obtiene que, para lograr mayor sensibilidad la escala es de 0.6, para lograr mayor exactitud es de 4.5 al emplear 28, 22 y 13 características. Por otra lado, para reducción de características con

PCA se emplea la escala es de 0.6 para mayor sensibilidad y 2.5 para mayor exactitud con 8 características.

A continuación se describe el proceso de configuración del clarificador árbol de decisiones es:

- Se recibe como entrada un conjunto de aprendizaje (S), que está conformado por los vectores de características y etiquetas de clase.
- Se divide el conjunto S en subconjuntos (L, R), a través de un criterio de partición (C), con lo cual se realiza una disminución de impurezas (ΔI).
- Se analiza un condición: Si ΔI es un valor mínimo, su valor no supera un ϵ , por lo tanto es un nodo terminal que es etiquetado con la clase más frecuente; pero si ΔI es significativo, su valor supera un ϵ , se crea un nodo interno del árbol sin etiqueta de clase y es dividido en función del criterio de partición C . Este proceso se repite hasta obtener todo los nodos terminales.

Los clasificadores de árbol de decisión se diferencian por el número máximo de divisiones posibles para jerarquizar la ramificación del árbol, teniendo como mínimo 4 en el *Simple Tree*, hasta 20 en el *Medium Tree* y máximo 100 en el *Complex Tree*. Tras realizar varias pruebas se determina que el clasificador más eficiente para la investigación es el árbol de decisiones *Complex Tree*, donde se realiza el proceso de podado y se obtienen mejores resultados en términos de sensibilidad al emplear 70 jerarquías, y en términos de exactitud al emplear 30 jerarquías.

Se evalúa el rendimiento de los clasificadores para valores de 5, 10 y 20 iteraciones. Definiendo que el valor de la validación cruzada que presenta mejores prestaciones es con 10 iteraciones, ya que con este valor se evita el subajuste y se controla el sobreajuste.

Como predictores se utilizaron las 28, 22 y 13 características detalladas anteriormente, y como respuesta se empleó el diagnóstico médico tabulado en la base de datos.

3.8 Configuración de la técnica PCA

Para implementar la técnica de reducción del espacio de características PCA, se emplea la función 'PCA' de MathWoks. Esta función recibe la matriz de características de entrenamiento X en filas y retorna una matriz con los componentes principales *score*. Luego se agrega a la matriz

de componentes principales, la columna de cada fila del diagnóstico clínico. Posteriormente se determina qué tipo de algoritmo se empleará de entre 'svd', 'eig' y 'als' en función de los requerimientos del sistema. Por defecto está determinado el algoritmo 'svd'. Finalmente se especifica el número de características que se tendrá de la reducción n , y se ingresa la matriz agrupada a la técnica de aprendizaje de máquina SVM (Matlab, 2017).

CAPÍTULO 4

RESULTADOS Y ANÁLISIS

El sistema desarrollado de detección de cáncer de piel tipo melanoma debe contar con un alto porcentaje de sensibilidad, sin perjudicar los valores de especificidad y exactitud. Es decir que el sistema debe tener la capacidad de determinar si un lunar es maligno o benigno, sin subestimar el diagnóstico médico.

Para desarrollar el sistema se emplean las características descritas en el Capítulo 3 de la presente investigación. Para validar y evaluar el sistema la investigación conlleva a realizar cuatro tipos de análisis, detallados a continuación:

El primer análisis consiste en evaluar el proceso de segmentación detallado en la Sección 3.4. Para esto se emplea el método de comparación de similitud entre las máscaras existentes en la base de datos y las máscaras extraídas al procesar las imágenes mediante algoritmos. El análisis es representado porcentualmente empleando el índice de Jaccard, permitiendo establecer los resultados de la similitud en términos de exactitud promedio.

El segundo análisis evalúa el rendimiento del sistema de detección utilizando el mecanismo de clasificación SVM. Los parámetros configurados en el clasificador están descritos en la sección 3.7 y presentan las mejores prestaciones para el sistema, influyendo en la sensibilidad y exactitud del mismo. Se utilizaron las imágenes de la base de datos y las máscaras de la misma. Adicionalmente se utiliza las 28 características descritas en la Tabla 3. Los resultados del análisis se representan en términos de sensibilidad, especificidad, exactitud y capacidad predictiva.

El tercer análisis se realiza un estudio de reducción de características. Se emplean las 22 y 13 características, detalladas en la sección 3.6; y la técnica PCA aplicada a 28 características. En esta sección se utiliza el mismo clasificador y configuración que en el segundo análisis, además se influye sobre los mismos parámetros. Se emplean las imágenes y máscaras de la base de datos. Los resultados se representan en términos de sensibilidad, especificidad, exactitud y capacidad predictiva.

El cuarto análisis evalúa el rendimiento del sistema de detección utilizando los clasificadores SVM y Árbol de decisión con 28 características. Los parámetros configurados en los clasificadores presentan las mejores prestaciones para el sistema. Se emplearon las imágenes y las máscaras de la base de datos. Los resultados para ambos clasificadores se representan en

términos de sensibilidad, especificidad, exactitud y capacidad predictiva. Este análisis permite establecer que clasificador tiene mejores prestaciones y rendimiento.

Finalmente, se realizó un análisis que evalúa el desempeño del sistema. En este se determina el tiempo de duración de los procesos principales y del sistema global, permitiendo estimar el tiempo requerido del proceso dependiendo de los recursos computacionales.

4.1 Segmentación de las imágenes

Una vez empleado el proceso de segmentación en la imagen original se obtienen la región de interés definida por la máscara, tal como se observa en las Figuras 41 y 42. A su vez, en la Figura 41 se observa un alto grado de concordancia entre las máscaras, ya que la segmentación realizada por los especialistas sigue su contorno real, al igual que la segmentación realizada en la presente investigación. En la Figura 42 se aprecia poca concordancia entre las máscaras, ya que en este caso los especialistas no siguen el contorno real del lunar, sino que toman parte de la piel, o el proceso implementado para la segmentación no reconoce toda la forma del lunar.

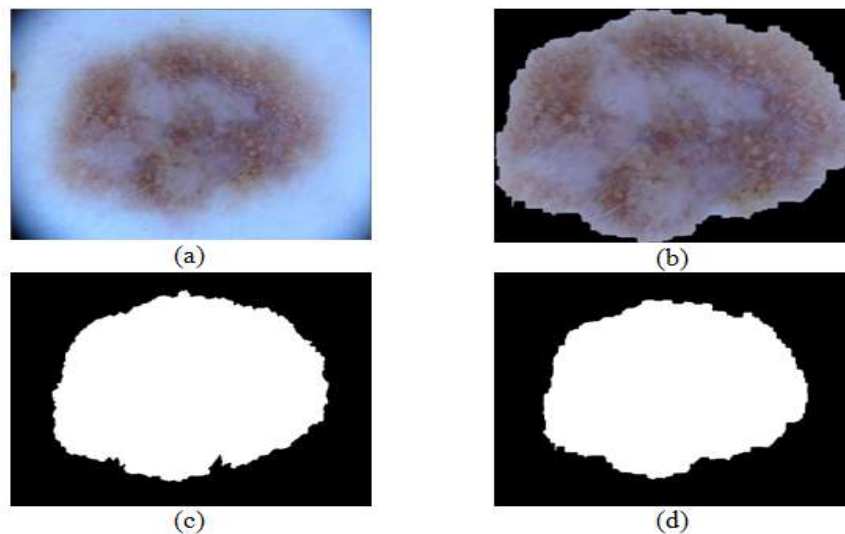


Figura 41 Segmentación con alto grado de similitud: (a) Imagen original, (b) Imagen segmentada, (c) Máscara de la base de datos, (d) Máscara extraída la imagen

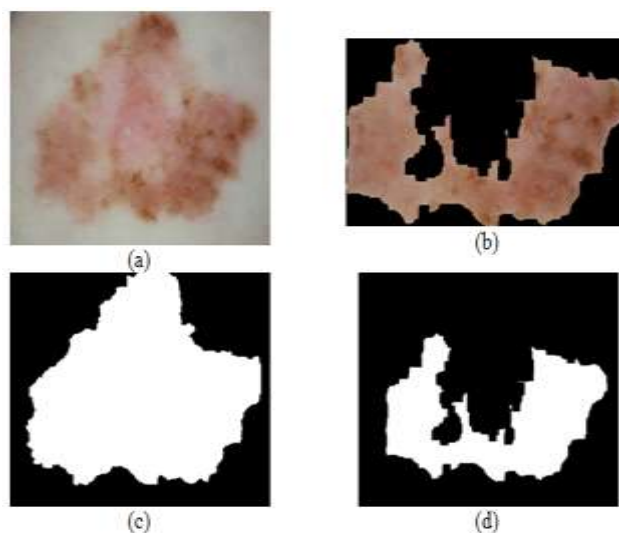


Figura 42 Segmentación con bajo índice de similitud: (a) Imagen original, (b) Imagen segmentada, (c) Máscara de la base de dato, (d) Máscara extraída la imagen

A continuación, se tabula el grado de similitud de las máscaras analizadas empleando el índice de Jaccard, y los resultados se muestran en la siguiente Tabla 5.

Tabla 5

Nivel de similitud entre las máscaras, empleando el índice de Jaccard

Máscaras base de datos VS Máscara extraída	Índice de Jaccard
Figura 38	90.30%
Figura 39	49.38%
Promedio total	59.75 %

Los resultados globales del proceso de segmentación indican que existe una similitud del 59.75%. Este valor es relativo ya que la segmentación realizada por los especialistas, en la base de datos, no es de forma precisa debido a que pueden confundir la piel como una sección del lunar. Además, pueden existir imágenes que no son compatibles con el proceso.

4.2 Análisis en términos de sensibilidad y exactitud al utilizar 28 características empleando el clasificador SVM

En esta sección se comparan dos resultados: en el primero, donde se tiene un mayor valor de sensibilidad, y el segundo donde se tiene un mayor valor de exactitud. Para este análisis se utilizaron las imágenes y máscaras de la base de datos.

Los resultados de este análisis se muestran en las matrices de confusión de las Tablas 6 y 7. Se obtiene que, al ser más influyente el valor de la sensibilidad se tienen menos casos de falsos negativos y mejores resultados de verdaderos positivos. Por otro lado, al ser más influyentes los valores de exactitud se tienen menos casos de falsos positivos. Con lo que se define que es directamente proporcional la influencia del marcador de detección para el diagnósticos, es decir que cuando una medida aumenta las otras disminuyen.

Tabla 6

Matriz de confusión en términos de sensibilidad - 28 Características

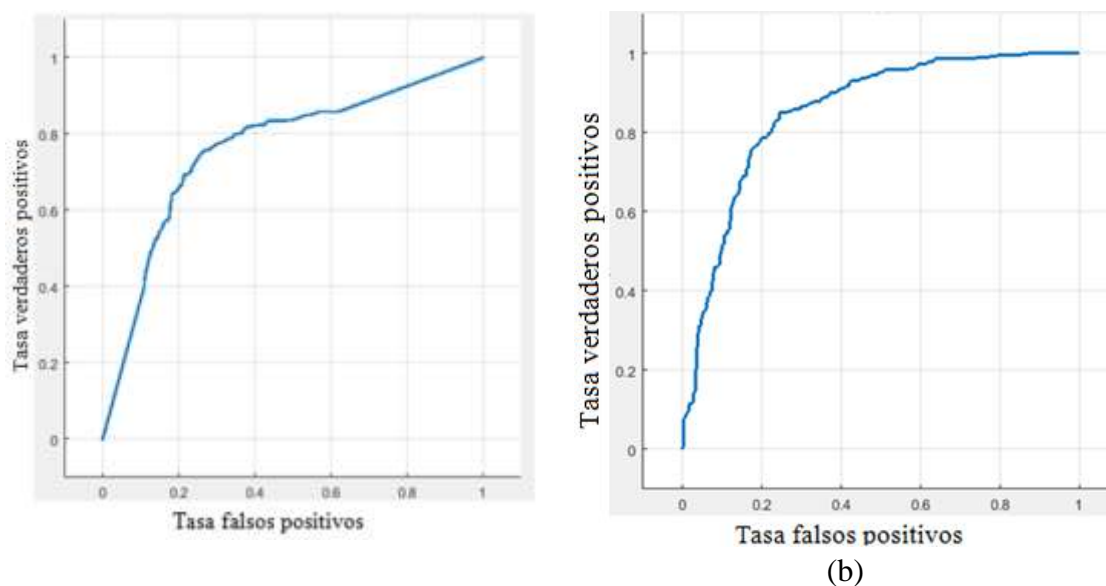
		Predicción	
		NO	SI
Especialista	NO	244	130
	SI	7	367

Tabla 7

Matriz de confusión en términos de exactitud - 28 Características

		Predicción	
		NO	SI
Especialista	NO	346	28
	SI	47	327

La capacidad predictiva, representada por la curva ROC, se interpreta como la probabilidad de detección correcta que tiene el sistema. Las curvas ROC para cada análisis se muestran en la Figura 43 (a) y (b) en términos de sensibilidad y exactitud respectivamente.



(a)
Figura 43 Curva ROC del clasificador SVM: (a) En términos de sensibilidad, (b) En términos de exactitud

Con los resultados de los parámetros de sensibilidad, especificidad, exactitud y capacidad predictiva se procede a realizar la Tabla 8.

Tabla 8

Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 28 características

Parámetro	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
En términos de sensibilidad	98.1	65.2	63.4	74.5
En términos de exactitud	87.4	92.6	80.1	80.7

Al utilizar 28 características se evidencia que la mayor capacidad predictiva se produce cuando se influye en términos de exactitud. El valor de la exactitud es del 80.1%, es decir, tiene un error del 19,9%, equivalente a fallar en 149 de las 748 imágenes.

Los resultados muestran que el sistema tiene un 98.1% en términos de sensibilidad, lo que indica que se cuenta con una tasa de falsos negativos del 1.9%, equivalente a fallar en 14 de las 748 imágenes, consiguiendo un alto grado de detección.

4.3 Análisis de reducción de características

En este estudio, al igual que en la sección 4.2, se obtiene la matriz de confusión y la curva ROC para 22 y 13 características, y además para PCA aplicado a 28 características. Sin embargo, al ser un proceso complementario donde no se obtuvieron datos relevantes en comparación con la sección anterior, se detallará únicamente la tabla de resumen de resultados para cada caso.

- **Análisis con 22 características**

Los resultados de los parámetros de sensibilidad, exactitud y capacidad predictiva se describen en la Tabla 9.

Tabla 9

Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 22 características

Parámetro	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
En términos de sensibilidad	96.8	68.7	65.8	73.3
En términos de exactitud	88.4	90.6	79.1	79.2

Al emplear 22 características se evidencia que la mayor capacidad predictiva se da al influir en términos de exactitud. Además, el valor de la exactitud es del 79.1%, es decir que tiene un error del 20.9%, equivalente a fallar en 157 de las 748 imágenes.

Los resultados muestran que el sistema tiene un 96,8% en términos de sensibilidad, lo que indica que se cuenta con una tasa de falsos negativos del 3.2%, equivalente a errar en 24 de las 748 imágenes.

- **Análisis con 13 características**

Los resultados de los parámetros de sensibilidad, exactitud y capacidad predictiva se describen en la Tabla 10.

Tabla 10

Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para el clasificador SVM con 13 características

Parámetro	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
En términos de sensibilidad	94.9	77.4	72.3	75.5
En términos de exactitud	85.1	92.3	77.3	77.8

El valor de sensibilidad es del 94.9%, es decir, se tiene una tasa de falsos negativos del 5.1% equivalente a tener 39 errores de un total de 748 imágenes.

Los resultados muestran que el sistema tiene un 77.3% en términos de exactitud, lo que indica que se cuenta con un error del 22.7%, equivalente a errar en 170 de las 748 imágenes.

- **Análisis empleando PCA aplicado a 28 características**

Al utilizar la herramienta PCA en el rango de 1 a 28 características, se obtienen la gráfica de la Figura 44. Esta indica el valor de reducción con mejores resultados para el sistema, siendo 8 características.

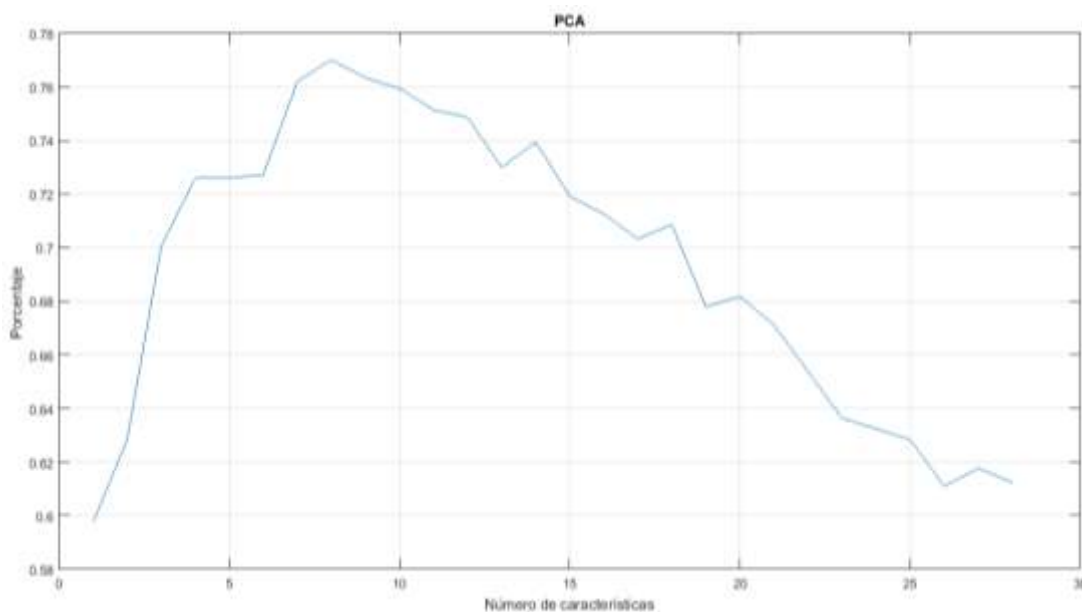


Figura 44 Gráfica de reducción de características - PCA

Los resultados de los parámetros de sensibilidad, exactitud y capacidad predictiva se describen en la Tabla 11.

Tabla 11

Resultados del análisis del sistema de detección en términos de sensibilidad y exactitud para PCA aplicado a 28 características

Parámetro	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
En términos de sensibilidad con 8 características	94.3	80.8	75.1	77.1
En términos de exactitud con 8 características	86.5	92.5	79.4	79.9

Los resultados muestran que el sistema tiene un 94.3% en términos de sensibilidad, lo que indica que se cuenta con una tasa de falsos negativos del 5.7%, equivalente a errar en 43 de las 748 imágenes.

El valor de la exactitud es del 79.4%, es decir se tiene un error del 20.6%, equivalente a fallar en 155 de las 748 imágenes.

Al aplicar la herramienta PCA a 28 características, se observa que los valores de exactitud y sensibilidad decrecieron al compáralos con el caso estudiado en la sección 4.2 donde no se aplica dicha herramienta.

4.4 Análisis en términos de sensibilidad y exactitud al utilizar 28 características empleando el clasificador árbol de decisiones

Los resultados obtenidos en términos de sensibilidad y exactitud para el clasificador árbol de decisiones se muestran en las matrices de confusión de las Tablas 12 y 13.

Tabla 12

Matriz de confusión en términos de sensibilidad - 28 Características

		Predicción	
		NO	SI
Especialista	NO	320	54
	SI	52	322

Tabla 13

Matriz de confusión en términos de exactitud - 28 Características

		Predicción	
		NO	SI
Especialista	NO	337	37
	SI	63	311

En las Tablas 12 y 13 se denota un aumento en los valores de falsos negativos y una disminución en los falsos positivos, provocado un aumento en la exactitud para el primer caso y una disminución de la exactitud en el segundo caso, en comparación con el clasificador SVM.

En las curvas ROC para el clasificador árbol de decisiones se evidencia un AUC en términos de sensibilidad del 72.7%; y en términos de exactitud de 74.5%. Se denota un decrecimiento de las posibles parejas formadas por un paciente con diagnóstico positivo y uno con diagnóstico negativo. A continuación, en las Figuras 45 (a) y (b), se muestran las curvas ROC al utilizar el clasificador árbol de decisiones.

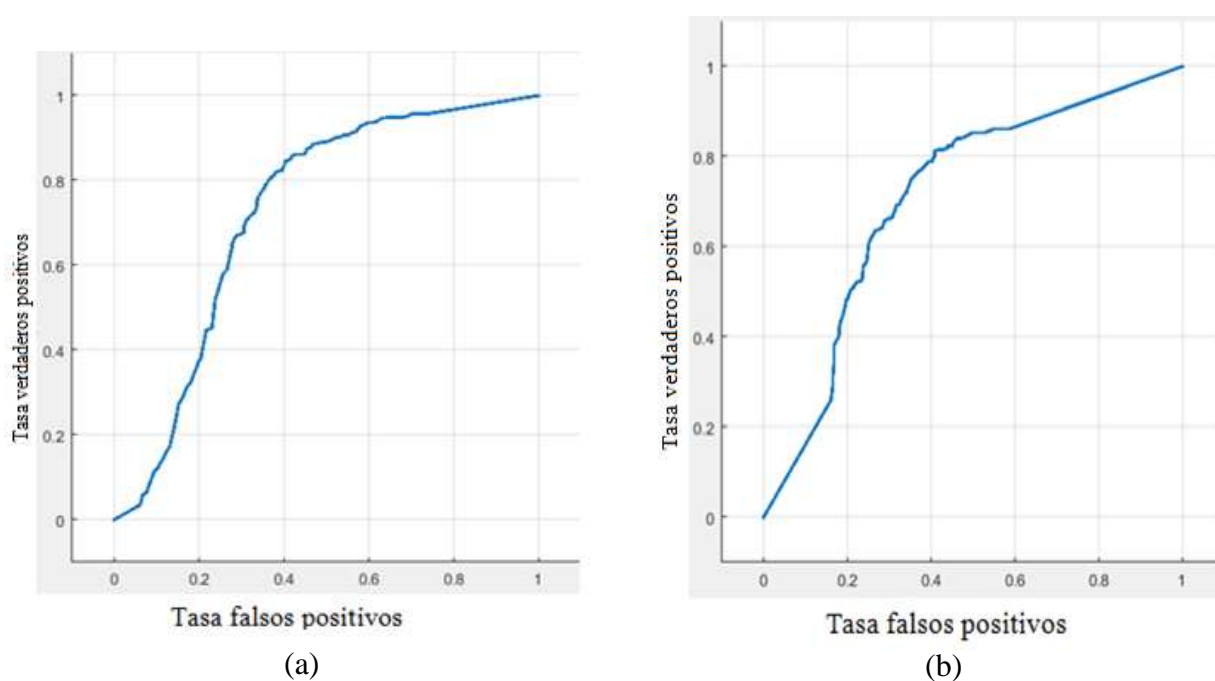


Figura 45 Curva ROC: (a) Árbol de decisiones en términos de sensibilidad, (b) Árbol de decisiones en términos de exactitud

Los resultados en términos de sensibilidad para los clasificadores SVM y árbol de decisiones se describen en la Tabla 14.

Tabla 14

Resultados del análisis del sistema de detección utilizando los clasificadores SVM y árbol de decisiones con 28 características en término de sensibilidad

Clasificador	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
SVM	98.1	65.2	63.4	74.5
Árbol de decisiones	86.1	85.5	72.7	72.7

Los resultados muestran que el sistema tiene un 98.1% de sensibilidad para el clasificador SVM, un valor que supera en un 12% al resultado con el clasificador árbol de decisiones. Pero la exactitud del clasificador árbol de decisiones supera en un 9.3% al clasificador SVM.

Los resultados en términos de exactitud para los clasificadores SVM y árbol de decisiones se describen en la Tabla 15.

Tabla 15

Resultados del análisis del sistema de detección utilizando los clasificadores SVM y árbol de decisiones con 28 características en términos de exactitud.

Clasificador	Sensibilidad (%)	Especificidad (%)	Exactitud (%)	Capacidad predictiva (%)
SVM	87.4	92.6	80.1	80.7
Árbol de decisiones	83.1	90.1	74.1	74.5

Al emplear el clasificador SVM se obtiene una exactitud del 80.1%, es decir, se tiene un error del 19.9%, equivalente a fallar en 149 de las 748 imágenes. Por otro lado, al utilizar el clasificador árbol de decisiones se tiene una exactitud del 74.1%, lo que muestra que se tiene un error del 25.9%, equivalente a fallar en 194 de las 748 imágenes.

El clasificador SVM tiene mejores resultados en términos de sensibilidad y exactitud en comparación con el clasificador árbol de decisiones. Esto es debido a que el clasificador SVM

con Kernel Gaussiano cuenta con un parámetro de escala que puede ser modificado para obtener mejores resultados. Dicho parámetro se explica en la sección 2.10.1.

Los esquemas del árbol de decisión en términos de sensibilidad y exactitud se muestran en las Figuras 46 y 47 respectivamente, donde cada columna se denotada por c_n que definen las 28 características detalladas en la Tabla 3.

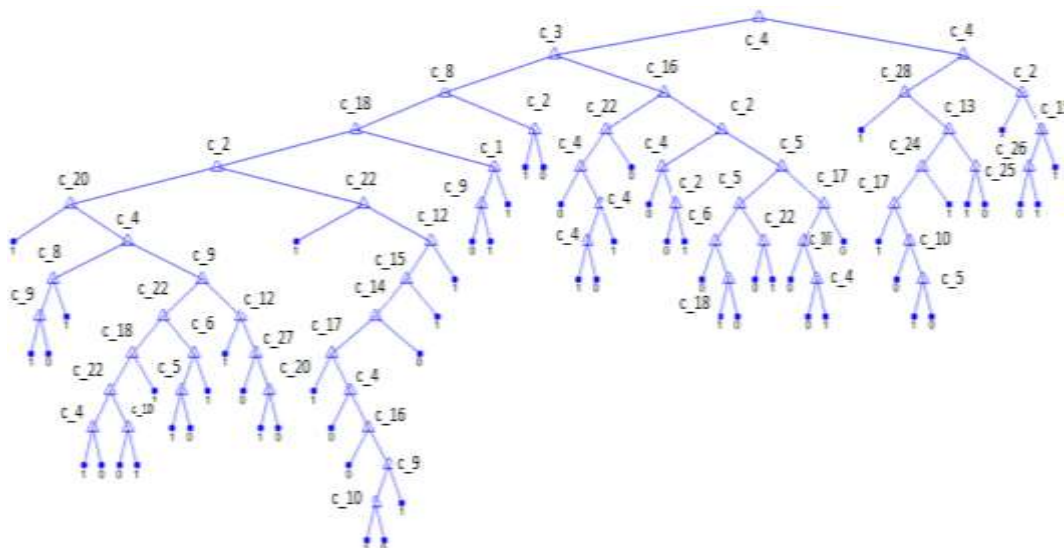


Figura 46 Representación gráfica del clasificador árbol de decisiones con 70 jerarquías

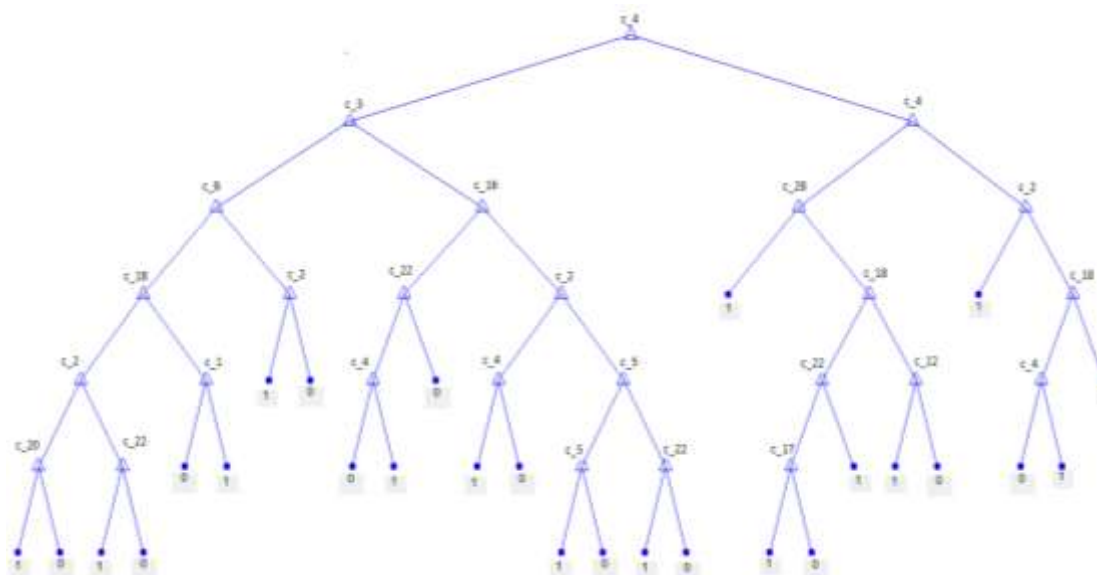


Figura 47 Representación gráfica del clasificador árbol de decisiones con 30 jerarquías

Las columnas empleadas son especificadas a continuación:

c_1 - Velo blanco azul	c_16 - Entropía de G
c_2 - Bordes	c_17 - Media de B
c_3 - Asimetría Proceso 1	c_18 - Desviación estándar de B
c_4 - Asimetría Proceso 2	c_20 - Tercer momento de B
c_5 - Media de R	c_22 - Entropía de B
c_8 - Tercer momento de R	c_24 - Desviación estándar de Gris
c_9 - Uniformidad de R	c_25 - Suavidad de Gris
c_10 - Entropía de R	c_26 - Tercer momento de Gris
c_12 - Desviación estándar de G	c_27 - Uniformidad de Gris
c_14 - Tercer momento de G	c_28 - Entropía de Gris
c_15 - Uniformidad de G	

4.5 Análisis de desempeño del sistema

Para medir el tiempo que el CPU ocupa en un proceso se emplea el comando *cputime* de Matlab. Este comando se basa en las funciones *tic* y *toc*, que indican los marcadores de inicio y de fin de la cuenta, respectivamente.

Para cada imagen el proceso de segmentación tiene una duración promedio de 30 segundos, tiempo que puede incrementarse al tener imágenes no compatibles con el sistema.

El proceso de clasificación, que incluye el procesamiento digital de imágenes y el entrenamiento del sistema, tiene una duración promedio de 5186 segundos para la extracción de 28 características en las 2000 imágenes, y 2581 segundos en 748 imágenes. Este tiempo puede aumentar o disminuir en función de los recursos del computador empleado para el proceso.

CAPÍTULO 5

CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS

5.1 Conclusiones

En el sistema de detección de cáncer de piel tipo melanoma se emplea un ordenador como elemento estructural para el procesamiento de datos, ya que no se requiere tener un hardware dedicado o de última tendencia para su funcionamiento. La principal ventaja de este sistema es que es un método no invasivo.

Se determinó que, tanto para un sistema médico asistido por ordenador como para un diagnóstico médico realizado por un especialista, se deben tener mayores valores en los parámetros de sensibilidad y capacidad predictiva, sin descuidar los parámetros de especificidad y exactitud. Se hace referencia a la especificidad ya que es preferible que el paciente con probabilidad de padecer melanoma, sea derivado a un centro de especialización donde se le realice un examen de confirmación o biopsia.

Se empleó validación cruzada como método de validación de resultados, ya que esta permite particionar el conjunto total de imágenes de la base de datos, obteniendo como resultado dos subconjuntos de imágenes uno para entrenamiento y otra para evaluación.

Los Kernels permiten trabajar con conjuntos que no son linealmente separables a partir del mapeo y transformación del espacio original de características.

Si la frontera del clasificador SVM está en sobreajuste se conseguirá que más imágenes negativas queden correctamente clasificadas, sin embargo, esto provoca que las imágenes positivas se dejen de clasificar correctamente. Por otro lado, si la frontera está en subajuste se clasifican correctamente las imágenes positivas, pero se dejará de clasificar las imágenes negativas.

Se determinó que algunas imágenes contenidas en la base de datos no son compatibles con el sistema. Las razones que se encontraron en la mayoría de imágenes inconsistentes son: falta de claridad, enmarcado oscuro o circular, ángulo de captura no constante, imágenes que no contienen lunares, técnicas de captura no adecuadas, y la segmentación realizada en la base de datos no sigue el contorno con precisión. Sin embargo, esto no influye en el desarrollo del sistema, ya que una gran cantidad de imágenes de la base de datos si son compatibles.

PCA permite dar un valor único a los lunares por ser un método apropiado para aplicaciones de clasificación. Es un método utilizado para disminuir la dimensión de las características asociadas a los lunares, reduciendo así la carga computacional pero no asegura un incremento en los valores de los parámetros medidos.

5.2 Recomendaciones

Uno de los factores indispensables para la investigación es la base de datos, se recomienda que esta sea estandarizada haciendo referencia a las imágenes, con esto se evita que las mismas sean incompatibles con el proceso o sistema. Principalmente se hace referencia a la estandarización en la captura de las imágenes, la cantidad, la igualdad y que no se encuentren enmarcadas.

Al capturar las imágenes de los lunares es recomendable que se utilice una misma lente y sea a una misma distancia, para evitar incompatibilidad con el sistema y se tenga mejores resultados.

Es recomendable que la base de datos tenga la suficiente cantidad de lunares y sea equilibrada, es decir que contenga la misma cantidad de imágenes con lunares malignos que con benignos.

Se recomienda establecer las características que se requieran extraer en un lunar, en función de los equipos empleados de captura de imágenes, para posteriormente buscar los procesos y algoritmos que permitan conseguir el objetivo deseado.

Se recomienda elegir el tipo de clasificador en función de sus características, y validar cada uno de sus parámetros. Además, probar cada uno de los tipos de sub-clasificadores, ya que dentro de cada uno se tiene un subconjunto con sus propias características o definiciones.

5.3 Trabajos Futuros

Se propone detectar otras estructuras que se pueden encontrar en un lunar mediante el método de *Menzies*, además se puede utilizar la extracción de características en el dominio de la frecuencia, permitiendo extraer mayor información útil para ser ingresada al clasificador y obtener un mayor rendimiento del sistema de detección en términos de sensibilidad, especificidad, exactitud y capacidad predictiva.

Se propone mejorar la segmentación de lesiones cutáneas, mediante la inserción de algoritmos que permitan eliminar el error producido por vellosidades, manchas que se forman en la captura de la imagen y el enmarcado de las mismas. Además, emplear la segmentación basada en energía, para lo cual se utiliza el algoritmo *Active Contour*.

En la investigación se empleó el método de clasificación supervisada con SVM y árbol de decisiones, se propone la inserción de otros algoritmos de clasificación que permitan realizar una comparación y un análisis con mayor cantidad de parámetros tales como redes neuronales, regresión logística y clasificador bayesiano.

REFERENCIA

- Abbasi, N., Shaw, H., & Rigel, D. (2004). Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. *The Journal of the American Medical Association*, 292(22), 2771–2776.
- Abril de Mur, A. (2017). *Detección de personas en secuencias de imágenes en entornos interiores*. Universidad de Alcalá.
- Alcántara, P., & Menárguez, J. (2013). Dermatoscopia en Atención Primaria. *AMF*, 9(6):331-339. Recuperado el 10 de Mayo de 2017, de Lunares cancerígenos: Cómo reconocerlos?: <http://www.saludysuerte.net/2013/05/lunares-cancerigenos-como-reconocerlos.html>
- Álvarez, D. (2012). *Detección precoz de la enfermedad de Alzheimer mediante clasificación automática de imágenes estructurales de resonancia magnética con SVM*. España.
- Amorim, S., Corominas, J., & Zarzosa, L. (2009). Análisis comparativo de técnicas estadísticas y de aprendizaje para evaluar la susceptibilidad del terreno a los deslizamientos superficiales en el pirineo catalán. *In VII Simposio Nacional sobre Taludes y Laderas Inestables*, 737-748.
- Argenziano, G., & Soyer, H. (2001). Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma. *The Lancet Oncology*, 2(7), 443–449.
- Barra, C. (2008). *Estudio comparativo de diferentes arquitecturas neuronales para tratamiento de imágenes hiperespectrales*. Universidad de Extremadura.
- Benitez, A. (2008). *Tema 9: Clasificación*. Recuperado el 15 de Mayo de 2017, de Universidad de Murcia: <http://www.um.es/geograf/sigmur/teledet/tema09.pdf>
- Berrocal, J., Figuerola, C., & Ángel, R. (2006). *Mejoras en la recuperación de información en la Web mediante el tratamiento de la información de los enlaces*. España: Universidad de Salamanca.
- Betancourt, G. (2005). Las máquinas de soporte vectorial (svms). *Scientia et technica*, 1(27).

- Binder, M., Schwarz, M., & Winkler, A. (1995). Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Archives of Dermatology*, 131(3), 286–291.
- Burroni, M., Corona, R., & Dell'Eva, G. (2004). Melanoma computer-aided diagnosis: reliability and feasibility study. *Clinical Cancer Research*, 10(6), 1881–1886.
- Carbon, P. (2017). *Aprendizaje Automatizado y Data Mining*. Obtenido de Universidad Miguel Hernández: <http://player.slideplayer.es/3/1075498/#>
- Cinfasalud. (2014). *Cuidado de la piel*. Recuperado el 08 de Agosto de 2017, de https://www.cinfasalud.com/wp-content/uploads/2014/09/Capas-de-la-piel_Cinfa.jpg
- Corso, C. (2009). Aplicación de algoritmos de clasificación supervisada usando Weka. *Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba*.
- Cover, T. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE transactions on electronic computers*, (3), 326–334.
- Crichton, N. J. (1997). Models for diagnosing chest pain: is CART helpful? *Statistics in medicine*, 16(7), 16(7), 717–727.
- Cuello, C. (21 de Febrero de 2016). *Entendiendo las Curvas ROC*. Recuperado el 17 de Febrero de 2018, de <https://www.youtube.com/watch?v=Y1XAP6omGzo>
- Cueva, P., & Yépez, J. (2014). *Epidemiología Del Cáncer en Quito*. Quito.
- Devijver, P. A. (1982). *Pattern recognition: A statistical approach (Vol. 761)*. London: Prentice-Hall.
- Dhawan, A., Gordon, R., & Rangayyan, M. (1984). Nevoscopy: three-dimensional computed tomography of nevi and melanomas in situ by transillumination. *IEEE Transactions on Medical Imaging*, 3(2), 54–61.

- Farrús, M., Anguita, J., & Hernando, J. (2007). *Fusión de sistemas de reconocimiento basados en características de alto y bajo nivel*. España: Universitat Politècnica de Catalunya, Universitat de Barcelona.
- Ferlay, J. (Marzo de 22 de 2017). *Estadísticas del cáncer*. Recuperado el 10 de Mayo de 2017, de Instituto Nacional de Cáncer: <https://www.cancer.gov/espanol/cancer/naturaleza/estadisticas>
- Gonzalez, R. C., & Woods, R. (2002). *Digital Image Processing Using Matlab*. PrenticeHall.
- Hamilton, H. G., Findlater, L., & Olive, W. (2013). *Computer science 831: knowledge discovery in databases*. Recuperado el 15 de Mayo de 2017, de http://www2.cs.uregina.ca/~hamilton/courses/831/notes/confusion_matrix/confusion_matrix.html
- HealSmith, M., Bourke, J., Osborne, J., & Graham-Brown, R. (1994). An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma. *British Journal of Dermatology*, 130(1), 48–50.
- Hedblad, M., & Lindelöf, B. (1994). Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic. *The Journal of Dermatology*, 21(7), 461–464.
- Johr, R. (2002). Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. *Clinics in Dermatology*, 20(3), 240–247.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In Ijcai*, 14(2), 1137-1145.
- Mackie, C., & Morton, A. (1998). Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *British Journal of Dermatology*, 138(2), 283–287.
- Magliano, J. (2015). Cáncer de piel. *Tendencia en medicinas*, 46, 49-54.
- Martínez, C. (2006). *Detector multiusuario para DS-CDMA basado en SVM*. Escuela Técnica Superior de Ingenieros.

- Matlab. (2017). *MathWorks*. Recuperado el 01 de Diciembre de 2017, de PCA: <https://www.mathworks.com/help/stats/pca.html>
- Mayer, J. (1997). Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *Medical Journal of Australia*, 167(4), 206–210.
- Menzies, M., & Vestergaard, S. (2008). Automated diagnostic instruments for cutaneous melanoma. *Seminars in Cutaneous Medicine and Surgery*, 27(1), 32–36.
- Menzies, S., Ingvar, C., Crotty, K., & McCarthy, W. (1996). Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. *Archives of Dermatology*, 132(10), 1178–1182.
- Menzies, W., & Talbot, H. (2005). The performance of SolarScan: an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma. *Archives of Dermatology*, 141(11), 1388–1396.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill International.
- Moscovitz, L., & Rengifo, P. (2010). Al interior de una máquina de soporte vectorial. *Revista de Ciencias*, 14, 73-85.
- Nachbar, F., Stolz, W., & Merkle, T. (1994). The ABCD rule of dermatoscopy. *Journal of the American Academy of Dermatology*, 30(4), 551–559.
- Ortiz, M. (2013). *Procesamiento digital de imágenes*. Recuperado el 10 de Mayo de 2017, de Benemérita universidad Autónoma de Puebla: <http://www.cs.buap.mx/mmartin/pdi>.
- Padilla, F. (2010). *Implementación eficiente de clasificadores Prior-SVM para Matlab (Bachelor's thesis)*. Madrid.
- Palacios, J. (2007). *Tejidos. Membranas. Piel. Derivados de la piel*. Recuperado el 10 de Mayo de 2017, de Infermera Virtual: <https://www.infermeravirtual.com/files/media/file/95/Tejidos%2C%20membranas%2C%20piel%20y%20derivados.pdf?1358605871>

- Pehamberger, H., Binder, M., Steiner, A., & Wolff K. (1993). In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. *Journal of Investigative Dermatology*, 100(3).
- Pérez, G. (2017). *Cancer Sintomas*. Recuperado el 10 de Mayo de 2017, de Síntomas de cáncer de piel: <http://cancersintomas.com/cancer-de-piel>
- Perrier, A. N. (2003). Cost-effectiveness analysis of diagnostic strategies for suspected pulmonary embolism including helical computed tomography. *American journal of respiratory and critical care medicine*, 167(1), 167(1),39-44.
- Pescador, M. I. (19 de Marzo de 2012). *Onmeda.es*. Recuperado el 10 de Mayo de 2017, de La piel: <http://www.onmeda.es/anatomia/piel.html>
- Pozzo, V., Benelli, C., & Roscetti, E. (1999). The seven features for melanoma: a new dermoscopic algorithm for the diagnosis of malignant melanoma. *European Journal of Dermatology*, 9(4),303–308.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 532-538.
- Reyes, P., & Torres, J. (2009). Diversidad, distribución, riqueza y abundancia de condrictios de aguas profundas a través del archipiélago patagónico austral, Cabo de Hornos, Islas Diego Ramírez y el sector norte del paso Drake. *Revista de biología marina y oceanografía*, 44(1), 243-251.
- Romón, B., & Enriquez, J. (2011). Trnasformaciones basada en el algoritmo Local Binary Pattern de imágenes capturadas con la Kinect para clasificación facial. *Universidad de Las Palmas de Gran Canaria*, 102.
- Saegerman, C. S. (2004). Decision support tools for clinical diagnosis of disease in cows with suspected bovine spongiform encephalopathy. *Journal of clinical microbiology*, 42(1), 42(1),172-178.

- Steiner, A. P., & Wolff, K. (1987). Improvement of the diagnostic accuracy in pigmented skin lesions by epiluminescent light microscopy. *Anticancer Research*, 7(3B), 433–434.
- Trujillano, J., & Sarria, A. (2008). Aproximacion a la metodologia basada en arboles de decision. Mortalidad hospitalaria del infarto agudo de miocardio. *Gaceta Sanitaria*, 22(1), 22(1),65-57.
- Trujillano, J., March, J., & Sorribas, A. (2004). Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina. *Med Clin (Barc)*, 122(s1).
- Turrisi, J., & Robinson, K. (2006). Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. *Archives of Dermatology*, 142(4),447–452.
- Vallejos, S. (1 de Enero de 2016). *American Academy of Dermatology*. Recuperado el 10 de Mayo de 2017, de <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/acerca/que-es-melanoma.html>
- Vidrio, R. (2003). Cáncer de piel. *Revista de la Facultad de Medicina UNAM*, 46(4), 166-171.
- Vila, A., Sedano, M., & Anna, L. (2003). *Contrastes de hipótesis de 1 población*. Catalunya: Universitat Oberta de Catalunya.
- Vilet, J. (2005). *Procesamiento digital de Imágenes*. Recuperado el 10 de Mayo de 2017, de http://read.pudn.com/downloads159/ebook/711796/Procesamiento_Digital_de_Imagenes.pdf
- Zaballos, P., Carrera, C., Puig, S., & Malvehy, J. (2004). Criterios dermatoscópicos para el diagnóstico del melanoma. *Med Cutan Iber Lat Am*, 32(1), 3-17.
- Zaccaro, J., Cárdenas, J., & Salazar, L. (2011). *Clasificación de señales cerebrales relacionadas con la imaginación de movimientos para aplicaciones de BCI*. Bogotá: Pontificia Universidad Javeriana.
- Zalaudek, I., Argenziano, G., & Soyer, H. (2006). Three-point checklist of dermoscopy: an open internet study. *British Journal of Dermatology*, 154(3), 431–437.

Zouridakis, G., Doshi, M., Duvic, M., & Mullani, N. (2005). Transillumination imaging for early skin cancer detection. Biomedical Imaging Lab. *Department of Computer Science, University of Houston*, 18.