



UNIVERSIDAD DE CARABOBO
ÁREA DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA
PROGRAMA DE MAESTRÍA
INGENIERÍA ELÉCTRICA



CLASIFICACIÓN DEL FOTOTIPO CUTÁNEO UTILIZANDO
APRENDIZAJE DE MÁQUINAS

AUTOR: Lic. Freddy J. Narea J.

Bárbula, Junio 2017



UNIVERSIDAD DE CARABOBO
ÁREA DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA
PROGRAMA DE MAESTRÍA
INGENIERÍA ELÉCTRICA



CLASIFICACIÓN DEL FOTOTIPO CUTÁNEO UTILIZANDO
APRENDIZAJE DE MÁQUINAS

AUTOR: Lic. Freddy J. Narea J.

TUTOR: Dr. César O. Seijas F.

Bárbula, Junio 2017



UNIVERSIDAD DE CARABOBO
ÁREA DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA
PROGRAMA DE MAESTRÍA
INGENIERÍA ELÉCTRICA



CLASIFICACIÓN DEL FOTOTIPO CUTÁNEO UTILIZANDO
APRENDIZAJE DE MÁQUINAS

AUTOR: Lic. Freddy J. Narea J.

Trabajo presentado ante el Área de Estudios de Postgrado de la Universidad de Carabobo para optar por el Título de Magíster en Ingeniería Eléctrica

Bárbula, Junio 2017



UNIVERSIDAD DE CARABOBO
FACULTAD DE INGENIERIA
DIRECCIÓN DE ESTUDIOS PARA GRADUADOS
SECCIÓN DE GRADO



ACTA DE DISCUSIÓN DE TRABAJO DE GRADO

En atención a lo dispuesto en los Artículos 137, 138 y 139 del Reglamento de Estudios de Postgrado de la Universidad de Carabobo, quienes suscribimos como Jurado designado por el Consejo de Postgrado de la Facultad de Ingeniería, de acuerdo a lo previsto en el Artículo 135 del citado Reglamento, para estudiar el Trabajo de Grado titulado:

“CLASIFICACIÓN DE FOTOTIPO CUTÁNEO UTILIZANDO APRENDIZAJE DE MÁQUINAS”

*Presentado para optar al grado de **MAGÍSTER EN INGENIERIA ELÉCTRICA** por el (la) aspirante:*

FREDDY NAREA
V.- 17.193.617

*Habiendo examinado el Trabajo presentado, decidimos que el mismo está **APROBADO**.*

En Valencia, a los doce (12) día del mes de Junio del año dos mil diecisiete.

Prof. Teddy Rojas
C.I.: 4876253
Fecha: 12/06/2017
FE: 09/06/2017 /D6.-

Prof. César Seijas
C.I.: 4567093
Fecha: 12/06/2017



Prof. Rafael Muñoz
C.I.: 4448239
Fecha: 12/06/2017

DEDICATORIA

A *Dios* que es la base de toda mi vida

A mi familia quien siempre me apoya

A las personas que han estado conmigo en este momento

A mi esposa *Caroleny Villalba* por su apoyo incondicional durante el desarrollo de
este proyecto

A mi hija *Valery Elizabeth* por alegrarme cada día

A mi madre *Mirvida de Narea* por brindarme su cariño y bendiciones

En Memoria de mi Padre *Freddy Narea Aponte*, siempre estuviste a mi lado...

AGRADECIMIENTOS

Primeramente a Dios por que sin él nada es posible.

A mi Tutor *Dr. César Seijas* por guiarme durante el desarrollo de este Trabajo de Investigación

Al Centro de Investigaciones Médicas y Biotecnológicas de la Universidad de Carabobo (**CIMBUC**) y al Servicio de Dermatología de la Ciudad Hospitalaria Dr Enrique Tejera (**CHET**) por los datos espectrales de los fototipos cutáneos suministrados.

ÍNDICE GENERAL

	Página
DEDICATORIA	VII
AGRADECIMIENTOS	VIII
ÍNDICE DE FIGURAS	XII
ÍNDICE DE TABLAS	XV
RESUMEN	XVI
INTRODUCCIÓN	1
I EL PROBLEMA DE LA INVESTIGACIÓN	3
1.1 Planteamiento del problema	3
1.2 Formulación del problema	5
1.3 Objetivos de la investigación	5
1.3.1 Objetivo general	5
1.3.2 Objetivos específicos	6
1.4 Justificación	6
1.5 Limitaciones	7
1.6 Alcance	7
II MARCO TEÓRICO	8
2.1 Antecedentes	8
2.2 Bases teóricas	9
2.3 La piel	9
	IX

2.3.1	Fototipo cutáneo	11
2.3.1.1	Determinación visual	12
2.3.2	Clasificación: escala de Fitzpatrick	13
2.4	Propagación de la luz en un tejido biológico	14
2.4.1	Absorción	15
2.4.2	Esparcimiento	17
2.4.3	Ecuación de transferencia radiativa (ETR) y aproximación de difusión	19
2.4.3.1	Consideraciones física en ETR	19
2.4.3.2	Radiancia	19
2.4.3.3	Ritmo de fluencia	20
2.4.3.4	Vector de flujo	20
2.4.4	Ecuación de transporte radiativo	21
2.5	Reconocimiento de patrones	23
2.6	Componentes principales	23
2.6.1	Calculo generalizado de las componentes principales	24
2.7	Conglomerados (clusters)	25
2.7.1	Algoritmo de k-medias	26
2.8	Modelos de redes neuronales	28
2.8.1	Perceptrón unicapa	30
2.8.2	Perceptrón multicapa	32
2.9	Máquinas de Soporte Vectorial (SVM)	34
2.9.1	Caso linealmente separable	35
2.9.2	Caso no lineal	38
2.9.3	Enfoques previos para entrenar una SVM	41
2.9.4	Librería de Máquina Soporte Vectorial (LIBSVM)	42
2.10	Curvas ROC	42
IIIMARCO METODOLÓGICO		44
3.1	Tipo de estudio y de investigación	44
3.2	Procedimiento Metodológico	45
3.3	Población y muestra	47

IV RESULTADOS	48
4.1 Distribución de datos espectrales	48
4.2 Aprendizaje de máquinas	50
4.2.1 SVM	50
4.2.2 RN	51
4.3 Experimentos	51
4.3.1 Clasificación por 31 rasgos Espectrales	53
4.3.2 Clasificación por 26 rasgos Espectrales	56
4.3.3 Clasificación por coeficiente de absorción de la Epidermis μ_a	59
4.3.4 Clasificación por PCA de 31 rasgos Espectrales	62
4.3.5 Clasificación por PCA de 26 rasgos Espectrales	66
4.3.6 Clasificación por PCA del coeficiente de absorción de la Epi- dermis μ_a	71
4.3.7 Clasificación por conglomerados	75
4.4 Comparación entre los métodos de clasificación	80
V CONCLUSIONES	83
CONCLUSIONES	83
RECOMENDACIONES	85
A Código sin reducción de Dimensión	86
B Código con PCA	93
C Código con Conglomerados-PCA	100
REFERENCIAS	109

ÍNDICE DE FIGURAS

	Página
Figura 2.1 Representación de la Estructura de la piel humana. Fuente: (Fitzpatrick and Freedberg, 2005)	11
Figura 2.2 Espectros de reflexión difusas para distintos fototipos cutaneos. Fuente: (Narea et al., 2015)	14
Figura 2.3 Representación de la interacción de la luz con un tejido biológico. Fuente: (Ishimaru, 1999)	15
Figura 2.4 Representación de la radiancia en un ángulo sólido. Fuente: (Ishimaru, 1999)	20
Figura 2.5 Representación del esparcimiento de la radiancia que incide en el volumen ρds . Fuente: (Ishimaru, 1999)	21
Figura 2.6 Esquema General de una red de McCulloch-Pitts. Fuente: (Viñuela and León, 2004)	28
Figura 2.7 Arquitectura del perceptrón unicapa. Fuente: (Wikipedia, 2017)	31
Figura 2.8 Arquitectura del perceptrón multicapa. Fuente: (Viñuela and León, 2004)	33
Figura 2.9 Maximización del margen. Fuente: (Burges, 1998)	36
Figura 2.10 Caso linealmente separable. Fuente: (Burges, 1998)	37
Figura 2.11 Mapeo del Espacio de entrada a un Espacio de características . Fuente: (Burges, 1998)	39
Figura 2.12 Representación de una Curva ROC para una buena clasificación. Fuente: (Narea, 2017)	43
Figura 4.1 Distribución de datos por Fototipo cutáneo. Fuente: (CIMBUC-CHET-2015)	49

Figura 4.2	Distribución de datos de entrenamiento por Fototipo cutáneo. Fuente:(Narea, 2017)	49
Figura 4.3	Distribución de datos de prueba por Fototipo cutáneo. Fuente: (Narea, 2017)	50
Figura 4.4	Coeficiente de absorción Melanina-Hemoglobina.	52
Figura 4.5	Datos Espectrales con 31 rasgos para cada Fototipo. Fuente: (Narea, 2017)	53
Figura 4.6	Datos Espectrales con 31 rasgos Normalizados para cada Fo- totipo. Fuente: (Narea, 2017)	53
Figura 4.7	Curva ROC SVM-Sigmaidal para 31 rasgos Fuente: (Narea, 2017)	55
Figura 4.8	Datos Espectrales con 26 rasgos para cada Fototipo. Fuente: (Narea, 2017)	56
Figura 4.9	Datos Espectrales con 26 rasgos Normalizados para cada Fo- totipo. Fuente: (Narea, 2017)	56
Figura 4.10	Curva ROC SVM-RBF para 26 rasgos Fuente: (Narea, 2017) .	58
Figura 4.11	μ_a cada Fototipo. Fuente: (Narea, 2017)	59
Figura 4.12	μ_a Normalizados para cada Fototipo. Fuente: (Narea, 2017) . .	60
Figura 4.13	Curva ROC RN para μ_a . Fuente: (Narea, 2017)	61
Figura 4.14	5 Componentes Principales de los datos espectrales. Fuente: (Narea, 2017)	62
Figura 4.15	Representación de las primeras 3 componentes principales en función de las longitudes de onda. Fuente: (Narea, 2017)	63
Figura 4.16	Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)	63
Figura 4.17	Curva ROC SVM-Lineal para PCA de 31 rasgos. Fuente: (Na- rea, 2017)	65
Figura 4.18	SVM-Lineal para PCA de 31 rasgos Fuente: (Narea, 2017) . .	66
Figura 4.19	6 Componentes Principales de 26 rasgos espectrales. Fuente: (Narea, 2017)	67

Figura 4.20 Representación de las primeras 3 componentes principales en función de las longitudes de onda para 26 rasgos espectrales. Fuente: (Narea, 2017)	67
Figura 4.21 Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)	68
Figura 4.22 Curva ROC SVM-Sigmoidal para PCA de 26 rasgos Fuente: (Narea, 2017)	69
Figura 4.23 SVM-Sigmaoidal para PCA de 26 rasgos Fuente: (Narea, 2017)	71
Figura 4.24 6 Componentes Principales de μ_a de la epidermis. Fuente: (Narea, 2017)	72
Figura 4.25 Representación de las primeras 3 componentes principales en función de las longitudes de onda para μ_a de la epidermis. Fuente: (Narea, 2017)	72
Figura 4.26 Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)	73
Figura 4.27 Curva ROC RN para PCA de μ_a de la epidermis Fuente: (Narea, 2017)	74
Figura 4.28 Distribución de datos por Fototipo cutáneo utilizando conglomerados. Fuente: (Narea, 2017)	76
Figura 4.29 Espectro de Reflectancia Difusa distribuido por conglomerados para cada fototipo cutáneo. Fuente: (Narea, 2017)	77
Figura 4.30 Componentes principales 1 y 2 para cada fototipo cutáneo . Fuente: (Narea, 2017)	78
Figura 4.31 Curva ROC para PCA de Conglomerados. Fuente: (Narea, 2017)	79
Figura 4.32 SVM-Lineal para PCA de conglomerados . Fuente: (Narea, 2017)	80
Figura 4.33 Color del fototipo cutáneo empleando entrenamiento supervisado y no supervisado. Fuente: (Narea, 2017)	81

ÍNDICE DE TABLAS

	Página
Tabla 2.1 Escala de Fitzpatrick.	13
Tabla 4.1 Máquinas de Aprendizaje para 31 rasgos espectrales.	54
Tabla 4.2 Matriz de confusión SVM-Sigmoidal 31 rasgos espectrales.	55
Tabla 4.3 Máquinas de Aprendizaje para 26 rasgos espectrales.	57
Tabla 4.4 Matriz de confusión SVM-RBF 26 rasgos espectrales.	58
Tabla 4.5 Máquinas de Aprendizaje para μ_a	60
Tabla 4.6 Matriz de confusión RN para μ_a	61
Tabla 4.7 Máquinas de Aprendizaje para PCA de los 31 rasgos espectrales.	64
Tabla 4.8 Matriz de confusión SVM-Lineal para PCA de 31 espectrales.	65
Tabla 4.9 Máquinas de Aprendizaje para PCA de los 26 rasgos espectrales.	69
Tabla 4.10 Matriz de confusión SVM-Sigmaoidal para PCA de 26 rasgos espectrales.	70
Tabla 4.11 Máquinas de Aprendizaje para PCA de μ_a de la epidermis.	73
Tabla 4.12 Matriz de confusión RN para PCA μ_a de la epidermis.	75
Tabla 4.13 Matriz de confusión conglomerados.	76
Tabla 4.14 Máquinas de Aprendizaje para PCA-Conglomerados.	78
Tabla 4.15 Comparación entre las Máquinas de Aprendizaje de cada expe- rimento.	81
Tabla 4.16 Diferencia del color para fototipos cutáneos por criterio médico y conglomerados	82



UNIVERSIDAD DE CARABOBO
ÁREA DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA
PROGRAMA DE MAESTRÍA
INGENIERÍA ELÉCTRICA



CLASIFICACIÓN DEL FOTOTIPO CUTÁNEO UTILIZANDO APRENDIZAJE
DE MÁQUINAS

AUTOR: Lic. Freddy J. Narea J.

TUTOR: Dr. César O. Seijas F.

FECHA: Junio 2017

RESUMEN

En esta investigación se clasificaron los fototipos cutáneos mediante aprendizaje de máquinas, a partir de los datos con 31 rasgos espectrales de la piel humana, suministrados por el Centro de Investigaciones Médicas y Biotecnológicas de la Universidad de Carabobo, los mismos se resumen a 26 rasgos tomando en consideración el dominio de la melanina en el rango espectral, asimismo a los datos espectrales se les determinó el coeficiente de absorción de la epidermis y adicionalmente se les aplicó un análisis de componentes principales para reducir la información espectral. Se usaron los algoritmos de redes neuronales (RN) y máquinas de soporte vectorial (SVM) para la clasificación. El desempeño fue evaluado usando el método de validación cruzada y tomando en cuenta los porcentajes de aciertos, las áreas ROC y el costo computacional. Se obtuvo, la SVM-Sigmoidal con parámetros $C=100$, $\gamma=0,001$ y $\alpha=1$, la cual acertó el 77,01 % de los fototipos. Finalmente, se evaluó la distribución de los datos espectrales utilizando conglomerados, donde el 44,18 % de los datos espectrales clasificados por criterio médico no coinciden con los clasificados por conglomerados, lo cual indica que se debe realizar una redistribución de los fototipos cutáneos del venezolano dentro de la escala de Fitzpatrick. En este sentido, se logra una clasificación del 98,81 % de aciertos empleando SVM con kernel Lineal (SVM-Lineal) de parámetros $C=1000$ y $\gamma = 0,5$, siendo esta en términos de tiempo computacionales, porcentajes de clases acertadas y relaciones de sensibilidad-especificidad, la máquina de aprendizaje más eficiente.

Palabras clave: Fototipo cutáneo, Reflexión Difusa, PCA, Conglomerados, Aprendizaje de Máquinas.

INTRODUCCIÓN

Este trabajo pretendió clasificar el fototipo cutáneo a partir de los datos espectrales de individuo venezolanos sanos utilizando aprendizaje de máquinas que puedan cambiar su comportamiento de manera autónoma basadas en su experiencia, por lo tanto, fue necesario seguir una serie de pasos desde el punto de vista metodológico que direccionaron la investigación.

De acuerdo a lo expuesto, las fases consideradas desde el punto de vista operativo para el curso de la investigación fueron las siguientes: revisión documental, Selección de atributos o rasgos de la información espectral de la piel, implementación de modelos de redes neuronales artificiales y máquina de soporte vectorial, análisis del desempeño de los distintos modelos en la clasificación y finalmente mediante la comparación de los modelos y la interpretación física se realizaron una serie de conclusiones y recomendaciones.

A continuación, conviene enunciar brevemente la estructura del presente estudio:

En el capítulo I se describe el planteamiento del problema, los objetivos, justificación de la investigación, alcance y sus limitaciones.

En el capítulo II se expone el marco teórico en el que se fundamenta el trabajo, abarcando los antecedentes, las bases teóricas de la estructura de la piel, definición de fototipo cutáneo, propagación de la luz en tejidos biológicos, reconocimiento de patrones, analisis de componentes principales, conglomerados, redes neuronales artificiales, máquina de soporte vectorial y curvas de rendimiento diagnóstico para el logro de los objetivos propuestos.

En el capítulo **III** presenta el marco metodológico describiendo el tipo y diseño de la investigación, los sujetos de estudio, procedimientos e instrumentos de recolección y análisis de datos.

En el capítulo **IV** se presentan y analizan los resultados obtenidos de los métodos de clasificación supervisados y no supervisados implementados a cada uno de los rasgos seleccionados de los datos espectrales para la clasificación del fototipo cutáneo del venezolano.

Y por último se presentan las conclusiones y recomendaciones pertinentes correspondientes a esta investigación.

CAPÍTULO I

EL PROBLEMA DE LA INVESTIGACIÓN

1.1. Planteamiento del problema

La piel es un medio biológico que se comporta perfectamente como un medio turbio multicapa, donde el principal agente de absorción es la melanina, lá cual es producida por los melanocitos en la capa basal de la epidermis; y es uno de los parámetros que determinan la coloración en la piel. En la actualidad la coloración de la piel se establece en forma indirecta por medio de una clasificación establecida en base a la capacidad de bronceado en la escala de Fitzpatrick (Fitzpatrick and Freedberg, 2005; Hunter et al., 1995) usada en dermatología.

La dermatología puede ser descrita como una rama de la medicina que se ocupa principalmente de la piel y sus enfermedades. Afortunadamente los dermatólogos se apoyan en una serie de invenciones modernas que hace que su trabajo sea más fácil y produce mejores resultados para los pacientes. La labor de un médico es hacer el diagnóstico correcto de la enfermedad así como, valorar su progreso para recordar y registrar el estado actual de los pacientes, los médicos utilizan algunos sistemas de registro escrito o iconográfico. Sin embargo, hay algunos factores que causan la falta de objetividad en estos sistemas. Un problema puede producirse cuando un paciente cambia de médico. Algunas discrepancias también se dan entre las anotaciones to-

madras por el mismo médico en diferentes circunstancias.

La clasificación Fitzpatrick fue desarrollada en 1975 por el dermatólogo Thomas Fitzpatrick. Esta escala categoriza la piel de una persona de acuerdo a su tez, color de cabello, propensión al bronceado y su tolerancia a la luz solar; lo que se conoce como fototipo cutáneo (Fitzpatrick and Freedberg, 2005). Basándose en esta información, los dermatólogos son capaces de determinar el fototipo cutáneo indicado, la tendencia de un individuo a desarrollar cáncer de piel así como qué tan bien reaccionará a los tratamientos médicos y estéticos que involucran la aplicación de algún tipo de luz como es el caso del láser y la luz intensa pulsada (IPL por sus siglas en inglés).

Actualmente existen un conjunto de técnicas ópticas no invasivas, como la Espectrofotometría de Reflectancia Difusa (ERD), que permiten realizar un registro numérico con información del fototipo cutáneo de la piel, generando así grandes matrices de información que podrían en vez de ayudar, complicar un diagnóstico médico, por lo cual tradicionalmente, el análisis de estos datos ha sido efectuado mediante técnicas estadísticas clásicas. Sin embargo, el incremento en la cantidad de datos y en el número de parámetros hace necesaria la aparición de nuevas metodologías y herramientas para un tratamiento automático de los registros depositados en las bases de datos de los espectros de reflexión difusa de la piel, que conlleva a la utilización de nuevos métodos de análisis y de clasificación como lo es el Aprendizaje de Máquinas.

Un tipo de aprendizaje de máquinas que han recibido una creciente atención son las Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés), debido a interesantes características tales como: buena generalización de la aplicación a nuevos objetos, la ausencia de mínimos locales y representación que depende de pocos parámetros a diferencia de las Redes Neuronales Artificiales (RN, por sus siglas en inglés) que presenta mínimos locales. Las Máquinas de Soporte Vectorial son nuevas estructuras fundamentadas en la teoría estadística del aprendizaje. Se basan en transformar el espacio de entrada en otro de dimensión superior (infinita) en el que el problema puede ser resuelto mediante un hiperplano óptimo (de máximo margen). Esta técnica ha probado su buen desempeño en muchas aplicaciones, tal como en la

identificación de lesiones en piel (Orozco. et al., 2010).

Las técnicas ópticas no invasivas permiten en la actualidad estimar los distintos tonos de piel ya sea por relación de las coordenadas cromáticas o por la recuperación de parámetros ópticos que cuantifican la cantidad de melanina en la epidermis a partir de los espectros de reflexión difusa(Narea et al., 2015); ofreciendo a los dermatólogos una referencia numérica que les permitan clasificar los distintos fototipos cutáneos.

En este estudio se plantea implementar técnicas de reconocimiento de patrones que permita clasificar el fototipo cutáneo a partir de los espectros de reflectancia difusa, tomando una población de 1000 individuos sanos con diferentes tipos de piel de acuerdo con la escala de Fitzpatrick.

1.2. Formulación del problema

Con la reciente llegada de las técnicas de clasificación que cada día son de uso más frecuente en el reconocimiento de patrones que permiten un diagnóstico médico, es posible implementar algoritmos que permita al computador diferenciar a partir de los datos espectrales los distintos fototipos cutáneos siguiendo como referencia la escala de Fitzpatrick. Esto con la finalidad de brindar una herramienta de clasificación de los diversos fototipos, sobre lo mencionado anteriormente se plantea la siguiente interrogante: ¿se podrá clasificar mediante aprendizaje de máquinas los espectros de reflexión difusa de cada fototipo cutáneo siguiendo la escala de Fitzpatrick?

1.3. Objetivos de la investigación

1.3.1. Objetivo general

Clasificar los fototipos cutáneos mediante aprendizaje de máquinas

1.3.2. Objetivos específicos

1. Conocer el estado del arte de los fundamentos del aprendizaje de máquinas y sus principales aplicaciones.
2. Seleccionar los principales rasgos de los espectros de reflexión difusa de los distintos fototipos cutáneos utilizando técnicas de reducción de dimensión.
3. Implementar modelos de redes neuronales artificiales y máquina de soporte vectorial para la detección de los distintos fototipos cutáneos.
4. Comparar el desempeño de los distintos modelos en la clasificación de los fototipos cutáneos a partir de los espectros de reflexión difusa.

1.4. Justificación

La presente investigación pretende beneficiar a los actores de las ciencias médicas específicamente a los especialistas en dermatología, así mismo a los pacientes que presentan riesgos de dermatosis, evitando incertidumbre diagnósticas en la práctica diaria dermatológica debido a la modificación visual de las lesiones elementales por los distintos fototipos.

En la práctica clínica la evaluación del color de la piel se realiza mediante la escala de Fitzpatrick, la cual tipifica el color de la piel de acuerdo a la capacidad de eritema y bronceado como fenómeno reaccional a la radiación electromagnética. Ahora bien cuando se maneja una cantidad grande de datos espectrales, el estudio del comportamiento de las curvas espectrales se torna subjetivo, por tal motivo la presente investigación surge como una alternativa que permite describir matemáticamente las curvas espectrales encontrando patrones únicos para diferenciar un fototipo de otro, partiendo del análisis de componentes principales, los conglomerados y el aprendizaje de máquinas como las SVM y RN. Este método de análisis se puede generalizar para describir y clasificar distintas patologías de piel (nevus, psoriasis, vitíligo, acantosis nigricans, entre otras).

la técnica de espectroscopia de reflexión difusa cuenta con una fuente de luz blanca que no afecta a los sujetos de estudio ni crea un impacto ambiental negativo, ya que, la radiación electromagnética emitida por la fuente es de baja energía y no es contaminante. Por esta razones es utilizada en diversas investigaciones como una herramienta óptica no invasiva

1.5. Limitaciones

Las SVM exhiben cierta limitaciones al sólo operar en su formulación original con muestras independientes e idénticamente distribuidas. Y las RN presentan complejidad temporal y espacial, tanto en el entrenamiento como en la evaluación, ya que los algoritmos existentes para resolver problemas con grandes conjuntos de datos tardan un tiempo que depende cuadráticamente del número de puntos.

Adicionalmente las SVM y RN se deben rentrenar al considerar nuevos datos espectrales, surgiendo distintos vectores de soporte ó parámetros de la red.

1.6. Alcance

Con la implementación del aprendizaje de máquinas se espera extraer rasgos característicos de los datos espectrales de distintos fototipos cutáneos, con la finalidad de generar una clasificación de acuerdo a la escala de Fitzpatrick, que permita al dermatólogo contar con una herramienta de análisis cuantitativo para establecer distintas patologías de piel.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes

Orozco. et al. (2010) logró discriminar lesiones benignas y malignas en la piel humana a partir de espectros de reflexión difusa empleando algoritmos de clasificación como máquinas de soporte vectorial, redes neuronales y bosques aleatorios. Evaluaron el desempeño el promedio de la k-fold cross-validation tomando en cuenta los porcentajes de instancias clasificadas correctamente, el índice kappa, el área bajo la curva ROC, la sensibilidad, y la especificidad. Encontrado que es posible clasificar correctamente al 89,89 % de los casos.

González et al. (2010) realizó una investigación para determinar los fototipos de la piel usando espectroscopia de reflexión difusa, donde encuentran que para personas hispanas, hay una diferencia clínica clara entre sujetos con fototipo I y su índice de melanina determinada por la Espectroscopia de Reflectancia Difusa; sin embargo, los sujetos con fototipos II-VI tienen un gran solapamiento del índice de melanina.

Narea et al. (2015) recuperaron los coeficientes de absorción de la melanina de la epidermis realizando un ajuste trigonométrico a las curvas de reflectancia difusa de cada fototipo cutáneo, encontrando adicionalmente que los individuos con fototipos II-III solo difieren en 1 % respecto a los parámetros a_0 entre dichos fototipos, mientras que entre las otras escala las diferencias promedios alcanza el 7 %.

Los trabajos de González et al. (2010) y Narea et al. (2015) demuestran la complejidad del análisis de los datos entre las distintas escalas intermedias de los fototipos cutáneos, contemplando la necesidad de utilizar herramientas de análisis que permitan realizar una optima clasificación.

Parikh and Shah (2016) implementó un clasificador de margen grande para diagnosticar las enfermedades de la piel más comunes en las zonas rurales, donde las personas son generalmente tratados por el personal médico, utilizando una máquina de soporte vectorial y probando diferentes núcleos para clasificación, logran una precisión del 95,39%.

2.2. Bases teóricas

2.3. La piel

Es una membrana resistente y flexible que cubre toda la superficie corporal. Es el mayor órgano del cuerpo humano, mide de 1,6 a 1,9 m^2 y posee una masa unos 2kg (Geneser, 2000). La piel es la encargada de recibir los estímulos del mundo exterior a través de las terminaciones nerviosas que se sitúan en ella y de allí se dirigen al cerebelo que indica como se debe reaccionar ante una situación; en ella se encuentra uno de los sentidos más desarrollados, denominado tacto. Cada centímetro cuadrado de piel contiene unos cinco mil receptores sensitivos.

El grosor, el color y la textura de la piel no son uniformes en todo el cuerpo; en algunas zonas como la del cuero cabelludo, se presentan más folículos pilosos que cualquier otra región del cuerpo humano; de hecho en algunas zonas como por ejemplo, la planta de los pies y las palmas de las manos carecen de algunos de estos folículos; siendo así, la piel mucho más gruesa. Por su estructura, este órgano sensorial se compone mediante tres capas:

1 Epidermis: tiene un espesor de $80\mu m$ y su función principal es proteger al cuerpo humano contra acciones perjudiciales del medio y contra la pérdida de líquidos. Esta se encuentra compuesta por:

- Estrato córneo: constituido por células aplanadas muertas que contienen el complejo proteico denominado queratina. El estrato córneo se encarga de prevenir la entrada de la mayoría de las sustancias extrañas y la pérdida de fluidos corporales.
- Queratinocitos (células escamosas): se encuentra debajo del estrato córneo, almacenando queratinocitos activos denominados también células escamosas, los cuales evolucionan progresivamente hasta formar el estrato córneo.
- Capa basal: es la más profunda de la epidermis que contiene células basales; las cuales se dividen continuamente, formando nuevos queratinocitos que reemplazan a los antiguos, los mismos se van desprendiendo de la superficie cutánea. Esta también contiene melanocitos que producen pigmentos negros o pardos en forma de gránulos denominados melanina, la cual contribuye en la coloración de la piel.

2 Dermis: es una capa gruesa de $1500\mu m$ con tejido conectivo a la cual se fija la epidermis y continúa en profundidad con el tejido subcutáneo rico en lípidos, encontrándose en ella:

- Glándulas sudoríparas: las cuales poseen forma de espiral con un tubo que se proyecta hasta el exterior, para así producir constantemente sudor, el cual sale a través de los poros; este se encarga de eliminar toxinas y de regular la temperatura corporal.
- Glándulas sebáceas: ubicada en toda la superficie externa del organismo humano, a excepción de las palmas de las manos, plantas y dorsos de los pies.
- Células adiposas: se encuentran en la parte inferior de la dermis y se encargan de proteger al organismo de golpes y proporcionan calor al mismo.
- Folículos pilosos: se extiende hacia la dermis llegando hasta la región subcutánea.

- 3 Hipodermis: tiene un espesor de $300\mu m$ y esta compuesta por tejido conjuntivo laxo, uniendo de manera poco firme la dermis con los órganos subyacentes, ésta se encuentra formada por una capa variable de tejido adiposo con una función de aislamiento, que permite que la piel se modifique y se proteja contra la pérdida de calor y traumatismos superficiales.

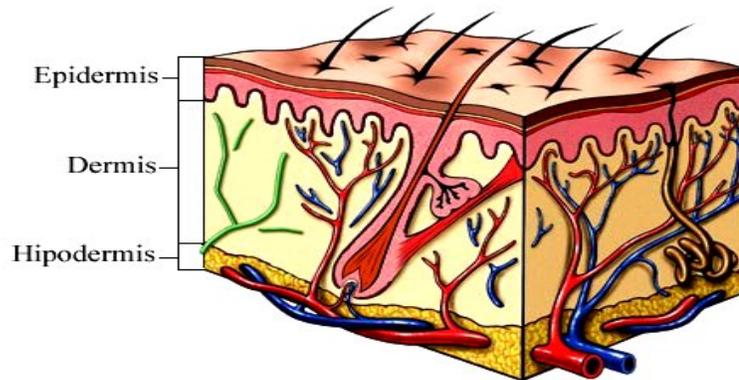


Figura 2.1: Representación de la Estructura de la piel humana. Fuente: (Fitzpatrick and Freedberg, 2005)

2.3.1. Fototipo cutáneo

El color de la piel se ha convertido en sinónimo de belleza; se considera que una persona con una piel bronceada es más atractiva, por lo que, en muchos casos, se realizan exposiciones prolongadas al sol, con el consecuente riesgo de presentar diferentes grados de eritema, melanoma, e incluso cáncer cutáneo fotoinducido. El conocimiento de los diferentes fototipos puede ayudar a tomar las medidas necesarias para evitar estos trastornos.

El fototipo cutáneo se puede definir como la capacidad de adaptación al sol que tiene cada persona desde que nace, es decir, el conjunto de características que determinan si una piel se broncea o no, y cómo y en qué grado lo hace. Cuanto más baja sea esta capacidad, menos se contrarrestarán los efectos de las radiaciones solares en la piel.

2.3.1.1. Determinación visual

Es una observación del color de la piel en las partes sin exponer o expuestas mínimamente al sol (la parte inferior de las piernas, detrás de la rodilla o la zona interna del antebrazo). No se debe utilizar la parte expuesta de la piel de la cara, del cuello o de los brazos (o pecho, en el caso de los hombres), debido a que el color de estas zonas está alterado por una fotoexposición repetida y prolongada. También son importantes otros factores, tales como:

- Aparición de pecas en verano.
- Color epidérmico en invierno: lechoso o blanco.
- Brillo: mate o ligeramente opaco.
- Color adquirido al broncearse: dorado, dorado-moreno, moreno (en sus diferentes tonalidades).
- Aparición de eritema (enrojecimiento) ante una exposición inmediata al sol (menos de 10 min).
- Presencia de eritema a los 21 días de exposición. Para determinar el grado de activación de los mecanismos fotoprotectores de la piel (formación de pigmento y engrosamiento del estrato córneo, que requieren 2-4 semanas para desarrollarse).
- Historia genética.
- Edad.
- Número de quemaduras solares graves que se ha tenido previamente.
- Posibilidad de fotosensibilización por determinados alimentos, medicamentos y cosméticos, entre otros posibles agentes causales.

Tabla 2.1: Escala de Fitzpatrick.

Fototipo de la Piel	Definición del típico fototipo de la piel	Reacción de la piel con la sobreexposición a la luz UV	Comentario sobre el fototipo de la piel
I	Cabello pelirrojo, rubio. Ojos azules, verdes. Piel muy clara	Siempre se quema. No se broncea	Pálido a veces con pecas. Generalmente admite que se quema.
II	Cabello claro a intermedio. Ojos claros a intermedios Piel clara a intermedia	Generalmente se quema. Raramente de broncea	Debería ser la primera consideración en la tipificación de la piel clara. Generalmente no reconoce que se quema pero admite, adquirir tonos rosados y necesita proteger la piel del sol
III	Cabello de color intermedio. Ojos intermedios a oscuro. Piel intermedia a oliva	Se quema moderadamente Generalmente se broncea	Generalmente no reconoce que se quema ligeramente si la exposición es moderada. Comentará que "teniendo cuidado, puede conseguir un buen bronceado"
IV	Cabello oscuro. Ojos oscuros. Piel de oliva oscura a moreno claro	Se quema levemente. Moreno moderado	Generalmente considera que se broncea con facilidad. Con exposiciones "moderadas" en climas septentrionales rara vez sufre quemaduras. Se sorprende cuando se quema "ligeramente" al visitar lugares con mayor intensidad.
V	Cabello Oscuro. Ojos oscuros Piel morena	Rara vez se quema. Moreno Oscuro	Se quema cuando no ha estado expuesto al sol durante meses o al exponerse a niveles muy altos de intensidad solar
VI	Cabello oscuro Ojos oscuros Piel muy oscura	No se quema No hay cambio en el color	Estos individuos tienen una pigmentación muy buena que les brinda una protección excepcional a la luz ultravioleta.

Fuente: (Fitzpatrick and Freedberg, 2005)

2.3.2. Clasificación: escala de Fitzpatrick

La clasificación de Fitzpatrick fue desarrollada en 1975 por el Dr. Thomas Fitzpatrick, dermatólogo graduado de la Universidad de Harvard. Esta escala categoriza la piel de una persona de acuerdo a su tez, color de cabello, propensión al bronceado y su tolerancia a la luz solar; los llamados fototipos cutáneos (ver tabla 2.1).

Por otra parte mediante la técnica de Espectroscopia de reflexión difusa se puede establecer un tipo de clasificación que se corresponda con la escala de Fitzpatrick como se propone en el trabajo de especialización en Dermatología de la Ciudad Hospitalaria Dr Enrique Tejera (Pardo et al., 2015), estos resultados se pueden evidenciar en la figura 2.2. Cabe destacar se basó en un análisis cualitativo para relacionar las

curvas espectrales con los fototipos cutáneos, pero no se logró generar una clasificación automatizada por algún factor de escala numérico.

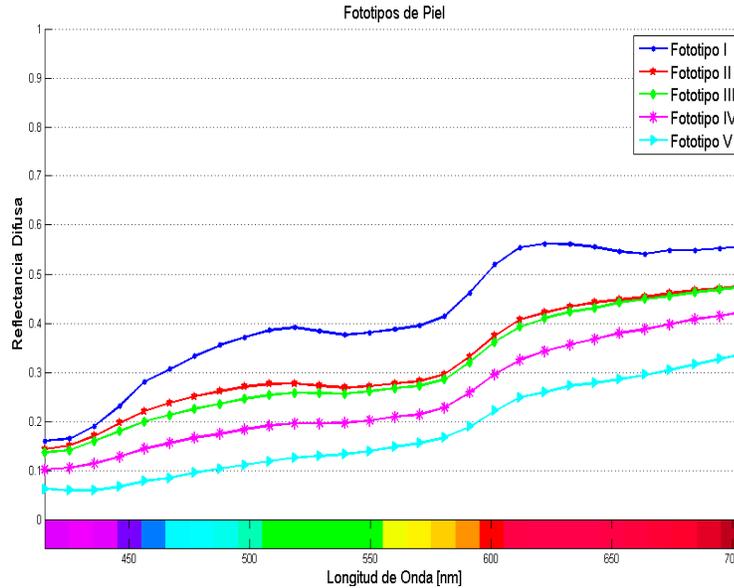


Figura 2.2: Espectros de reflexión difusa para distintos fototipos cutáneos. Fuente: (Narea et al., 2015)

2.4. Propagación de la luz en un tejido biológico

La mayoría de los tejidos biológicos son considerados turbios, es decir, estructuras heterogéneas con variaciones en sus propiedades ópticas en función de la posición. Pese a esto, el tejido biológico se describirá como un medio con partículas que absorben y esparcen luz distribuidas aleatoriamente en un volumen, el cual será isotrópico y homogéneo desde un punto de vista microscópico.

Al hacer incidir luz sobre la superficie de un tejido biológico, que es la frontera entre el tejido biológico y el medio donde la luz se propaga inicialmente, habrá reflexión especular y refracción, fenómenos que obedecen a la ley de la reflexión y la ley de la refracción de Snell respectivamente. Si la luz incidente tiene un ángulo de 0 respecto

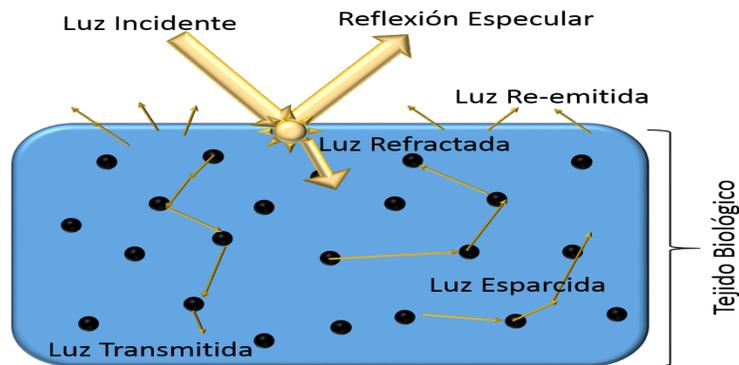


Figura 2.3: Representación de la interacción de la luz con un tejido biológico. Fuente: (Ishimaru, 1999)

a la normal de la superficie del tejido biológico, según las ecuaciones de Fresnel, la porción de luz del haz incidente que penetrará en el tejido será proporcional a $1 - r^2$, donde r es el coeficiente de reflectividad y se muestra en la ecuación 2.1.

$$r = \frac{n_r - n_i}{n_r + n_i} \quad (2.1)$$

Donde n_r es el índice de refracción del medio en el cual la luz será refractada y n_i el índice de refracción del medio donde la luz esta incidiendo.

Los fotones que logran penetrar al tejido biológico podrán ser absorbidos o esparcidos; en este ultimo caso los fotones serán desviados de su dirección de propagación inicial en un ángulo θ , el cual está fuertemente relacionado con el factor de anisotropía (Wang and Jacques, 1993).

2.4.1. Absorción

Si la energía de un fotón que incide sobre un átomo corresponde a la energía necesaria para llevar a un electrón de dicho átomo a un estado energético superior al estado base, se dará la absorción. Cuando el electrón excitado salta de un nivel energético superior a uno inferior, la energía se disipa en forma de calor o en luminiscencia del tejido.

La ley que describe cómo los efectos de absorción atenúan la radiación electromagnética, cuando esta pasa a través de un medio, es la ley de Beer-Lambert:

$$I(z) = I_0(z)e^{-\mu_a(\lambda)z} \quad (2.2)$$

donde z es la distancia desde la interfase y es paralela a la dirección de propagación, se expresa en unidades de $[m]$ de acuerdo al Sistema Internacional de Unidades (SI), para mantener la uniformidad dimensional de todo el desarrollo de este trabajo, I es la irradiancia en función de la distancia z y I_0 es la irradiancia incidente, donde ambas se expresan en unidades de $[Wm^{-2}]$; y μ_a es el coeficiente de absorción del medio el cual expresa el inverso de la distancia $[m^{-1}]$ promedio que un fotón puede recorrer antes de ser absorbido y esta expresado por la ecuación:

$$\mu_a = \rho_a \sigma_a \quad (2.3)$$

donde ρ_a es el número de partículas absorbentes contenidas en un volumen $[m^{-3}]$ y σ_a es la sección eficaz de absorción de una partícula absorbente $[m^2]$. Se asume que las partículas absorbentes con un área eficaz σ_a son idénticas y que están distribuidas de manera uniforme.

Se puede decir que la absorción es el proceso por el cual los átomos o moléculas de un material, en este caso el tejido biológico, adquieren o toman energía de una onda electromagnética incidente, provocando que esta última sea atenuada. La absorción es útil en aplicaciones médicas, tanto para procesos de diagnóstico como para procesos terapéuticos. En el caso del diagnóstico, el espectro de absorción de una molécula puede servir como identificador de ésta, pues las transiciones entre los niveles de energía, se darán a longitudes de onda específicas. Para el caso del tratamiento, donde se utiliza la luz para producir cambios estructurales en el tejido, como la terapia fotodinámica o la cirugía láser, la absorción es importante y necesaria, ya que sólo la luz absorbida puede producir cambios físicos, químicos o fisiológicos en un sistema biológico. (W. F. Cheong and Welch, 1990)

2.4.2. Esparcimiento

Cuando la frecuencia de la onda electromagnética asociada al fotón que incide sobre una partícula, no coincide con la frecuencia natural de dicha partícula, no logrará llevar a un nivel energético superior a los átomos de esa partícula, entonces habrá esparcimiento.

De forma similar a la atenuación de una onda electromagnética debida a la absorción en un medio, la atenuación debida sólo a los fenómenos de esparcimiento puede ser expresada como:

$$I(z) = I_0(z)e^{-\mu_s(\lambda)z} \quad (2.4)$$

donde μ_s es el coeficiente de esparcimiento del medio y tiene unidades de $[m^{-1}]$. El coeficiente de esparcimiento es el inverso de la distancia promedio que un fotón puede recorrer antes de enfrentarse a un evento de esparcimiento y está expresado de forma análoga a la ecuación (2.3):

$$\mu_s = \rho_s \sigma_s \quad (2.5)$$

donde ρ_s es el número de partículas que esparcen contenidas en un volumen $[m^{-3}]$ y σ_s es la sección eficaz del esparcimiento $[m^2]$. Se asume que las partículas que esparcen con un área eficaz σ_a son idénticas y que están distribuidas de manera uniforme.

Cuando un fotón que incide con una dirección descrita por vector unitario \hat{s} sufre un esparcimiento; la probabilidad de que su nueva dirección sea descrita por otro vector \hat{s}' es dada por la función de fase normalizada $f(\hat{s}, \hat{s}') = f(\theta)$. Para los tejidos biológicos se puede asumir que la probabilidad de distribución es una función que depende solamente del ángulo que forman las direcciones del fotón incidente y el fotón esparcido. Por lo tanto, la función de fase se puede expresar como una función del coseno del ángulo de esparcimiento:

$$f(\hat{s}, \hat{s}') = \cos\theta \quad (2.6)$$

Para describir el esparcimiento existen diversos modelos, dependiendo de las partículas que esparcen y de la longitud de onda del haz incidente; los modelos más conocidos y utilizados para estos estudios son el esparcimiento de Mie y de Rayleigh. Mie desarrolló una teoría que estableció que las soluciones analíticas de la función de fase son funciones derivadas y así determino el esparcimiento de una onda electromagnética plana por partículas esféricas de cualquier tamaño. Para tamaños de partículas mucho menor que la longitud de onda incidente, la teoría de Mie se puede aproximar con la teoría de Rayleigh (Meglinski and Matcher, 2003; Prahl et al., 1989; W. F. Cheong and Welch, 1990).

La anisotropía de un medio se puede caracterizar utilizando el valor medio del coseno del ángulo de esparcimiento θ . El parámetro utilizado se denomina factor de anisotropía g :

$$g = \int_0^\pi f(\theta) \cos\theta \sin\theta d\theta \quad (2.7)$$

Para $g = 0$ se tiene un esparcimiento perfectamente isotrópico, cuando $g = -1$ se tiene retro esparcimiento y cuando $g = 1$ el esparcimiento tiende hacia adelante y es anisotrópico. Los tejidos biológicos tienen factores de anisotropía comprendidos en el rango de $0,69 \leq g \leq 0,99$. El coeficiente de esparcimiento reducido μ'_s , se define como:

$$\mu'_s = \mu_s(1 - g) \quad (2.8)$$

Esta cantidad puede ser interpretada como un equivalente al coeficiente de esparcimiento y es un parámetro fundamental en la teoría de difusión de la propagación de la luz a través de medios turbios.

Combinando la absorción y el esparcimiento, se puede definir el coeficiente de atenuación total μ_t :

$$\mu_t = \mu_a + \mu'_s \quad (2.9)$$

donde el inverso de μ_t se define como el camino libre medio de un fotón ($\frac{1}{\mu_t}$). De forma analoga, el coeficiente de transporte de atenuación μ_{tr} es:

$$\mu_{tr} = \mu_a + \mu_s(1 - g) = \mu_a + \mu'_s \quad (2.10)$$

2.4.3. Ecuación de transferencia radiativa (ETR) y aproximación de difusión

La propagación de la radiación a través de un tejido biológico puede, en principio, ser descrita utilizando la teoría electromagnética fundamental. En este caso, el tejido se puede considerar un medio con una permitividad que varía espacialmente, y las variaciones del campo podrían ser descritas utilizando las ecuaciones de Maxwell. Sin embargo, este planteamiento no es factible debido a la complejidad del problema y la falta de un preciso conocimiento de la permitividad del medio. El problema puede ser simplificado ignorando fenómenos relacionados con la propagación de las ondas, como la polarización y la interferencia, y otros fenómenos relacionados con las partículas, como las colisiones inelásticas. La teoría de la transferencia radiativa básicamente describe la propagación de la radiación electromagnética en un medio, donde esta puede ser emitida, esparcida y absorbida. La ecuación básica de esta teoría y que describe los fenómenos antes mencionados, es llamada la ecuación de transporte. Dicha ecuación es deducida utilizando el principio de conservación de energía, lo que produce que la teoría de la transferencia radiativa también lo sea. Dado esto, la teoría de la transferencia radiativa es muy versátil, pues ha sido utilizada para describir la propagación de la radiación proveniente de estrellas, galaxias e incluso para modelar la propagación de la radiación en el tejido biológico.(Ishimaru, 1999)

2.4.3.1. Consideraciones física en ETR

2.4.3.2. Radiancia

La radiancia es el diferencial de potencia dP [W] que fluye dentro de un ángulo solido $d\omega$ [sr] a través de un diferencial de área dA [m^2] en el punto \vec{r} y que se propaga en la dirección del vector unitario \hat{s} . Cuando \hat{s} no es perpendicular a dA , se tiene que utilizar la proyección de dA en el plano perpendicular a \hat{s} , por lo que $dA_0 = dA \cos\theta$, donde θ es el ángulo entre la normal de dA y \hat{s} (ver figura 2.4). La radiancia L se puede expresar como:

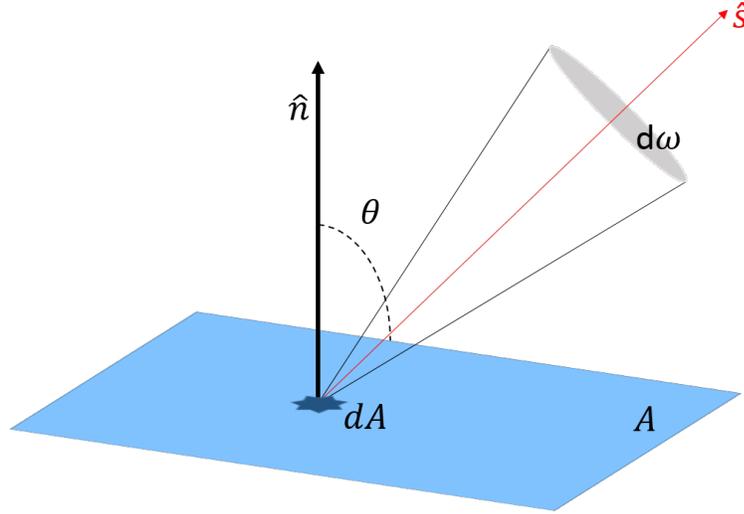


Figura 2.4: Representación de la radiancia en un ángulo sólido. Fuente: (Ishimaru, 1999)

$$L(\vec{r}, \hat{s}) = \frac{dP(\vec{r}, \hat{s})}{dA \cos\theta d\omega} \quad (2.11)$$

Donde las unidades de la radiancia es $[\frac{W}{m^2 \text{sr}d}]$.

2.4.3.3. Ritmo de fluencia

El ritmo de fluencia es la potencia que incide sobre una esfera infinitesimalmente pequeña dividida entre la sección transversal de dicha esfera y se expresa:

$$\Phi(\vec{r}) = \int_0^{4\pi} L(\vec{r}, \hat{s}) d\omega \quad (2.12)$$

2.4.3.4. Vector de flujo

El vector de flujo es definido como el flujo neto de energía por unidad de área por unidad de tiempo y es expresado como:

$$\vec{F}_d(\vec{r}) = \int_0^{4\pi} L(\vec{r}, \hat{s}) \hat{s} d\omega \quad (2.13)$$

donde el vector de flujo apunta en la dirección del flujo dominante de fotones.

2.4.4. Ecuación de transporte radiativo

Considerando un volumen cilíndrico de sección transversal unitaria, una longitud ds y que contiene una cantidad ρds partículas, donde ρ es el número de partículas por unidad de volumen, como el que se muestra en la figura 2.5

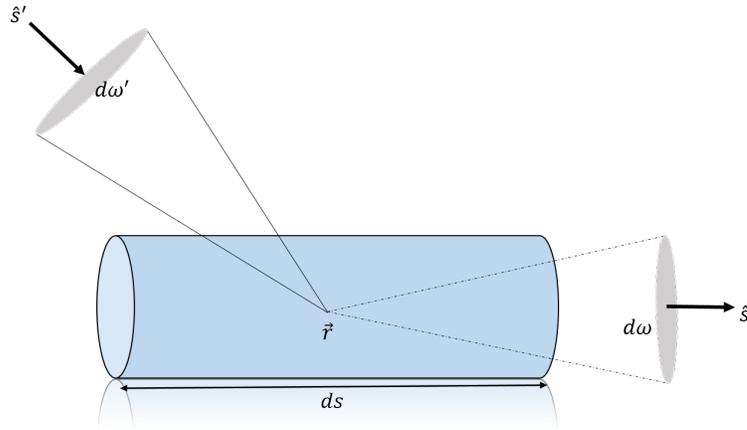


Figura 2.5: Representación del esparcimiento de la radiancia que incide en el volumen ρds . Fuente: (Ishimaru, 1999)

Cada partícula absorbe una cantidad de energía relacionada con el término $\sigma_a L(\vec{r}, \hat{s})$ y esparce una cantidad de energía relacionada con el término $\sigma_s L(\vec{r}, \hat{s})$, y por lo tanto el cambio de la radiancia debido a esto en el volumen ρds , en el punto \vec{r} y en la dirección de \hat{s} , puede ser expresado como:

$$dL(\vec{r}, \hat{s}) = -\rho(\sigma_a + \sigma_s)dsL(\vec{r}, \hat{s}) = -\mu_t L(\vec{r}, \hat{s})ds \quad (2.14)$$

A la vez, la radiancia $L(\vec{r}, \hat{s})$ puede incrementar porque los fotones que inciden sobre el volumen desde otra dirección \hat{s}' , pueden ser esparcidos en la dirección del vector unitario \hat{s} . La contribución debido a lo anterior puede ser expresada con la ecuación (2.15) y se ilustra en la figura 2.5.

$$dL(\vec{r}, \hat{s}) = -ds\mu_s \int_0^{4\pi} f(\hat{s}, \hat{s}')L(\vec{r}, \hat{s})d\omega' \quad (2.15)$$

Los fotones provenientes de una fuente de luz externa o interna al tejido y los fotones generados dentro del tejido debido a otros fenómenos físicos como la fluorescencia, pueden incrementar la radiancia $L(\vec{r}, \hat{s})$. el término fuente será expresado como:

$$L(\vec{r}, \hat{s}) = \varepsilon(\vec{r}, \hat{s})ds \quad (2.16)$$

Al dividir el resultado de la suma de las contribuciones de (\vec{r}, \hat{s}) debido a los fenómenos descritos por las ecuaciones (2.14, 2.15 y 2.16) entre ds , se tiene la ecuación de transporte:

$$\frac{dL(\vec{r}, \hat{s})}{ds} = -\mu_t(\vec{r}, \hat{s}) + \mu_s \int_0^{4\pi} f(\hat{s}, \hat{s}')L(\vec{r}, \hat{s})d\omega + \varepsilon(\vec{r}, \hat{s}) \quad (2.17)$$

El $\frac{dL(\vec{r}, \hat{s})}{ds}$ es el gradiente de la radiancia en la dirección del vector constante \hat{s} .

La radiancia en cualquier punto del tejido biológico se compone de dos partes: la radiancia esparcida $L_s(\vec{r}, \hat{s})$ y la radiancia reducida incidente $L_i(\vec{r}, \hat{s})$. La radiancia reducida incidente es la luz de una fuente externa o interna al tejido, que no ha sido esparcida ni absorbida, dicho de otra manera, los fotones que han sobrevivido a los fenómenos de absorción y de esparcimiento; la radiancia esparcida es aquella que se “forma” dentro del medio debido a los fenómenos de esparcimiento de la radiancia reducida incidente. La radiancia en cualquier punto del tejido se puede expresar como:

$$L(\vec{r}, \hat{s}) = L_s(\vec{r}, \hat{s}) + L_i(\vec{r}, \hat{s}) \quad (2.18)$$

Si se toma en cuenta el número de partículas que esparcen con una sección eficaz de esparcimiento σ_s y la probabilidad de que un fotón que se propaga en la dirección del vector unitario \hat{s}_0 (dentro del medio) sea esparcido y recogido en la dirección del vector unitario \hat{s} es $f(\hat{s}, \hat{s}_0)$, el término fuente puede ser expresado:

$$\varepsilon(\vec{r}, \hat{s}) = \mu_s f(\hat{s}, \hat{s}_0)I(\vec{r}, \hat{s}_0) \quad (2.19)$$

donde $I(\vec{r}, \hat{s})$ puede representar un haz de luz colimado o una fuente puntual

isotrópica, por lo que la ecuación de transporte en términos de $L_s(\vec{r}, \hat{s})$ queda expresado de la siguiente manera:

$$\frac{dL_s(\vec{r}, \hat{s})}{ds} = -\mu_t L_s(\vec{r}, \hat{s}) + \mu_s \int_0^{4\pi} f(\hat{s}, \hat{s}') L(\vec{r}, \hat{s}') d\omega' + \mu_s f(\hat{s}, \hat{s}_0) I(\vec{r}, \hat{s}_0) \quad (2.20)$$

La ecuación de transporte radiativo asume que las partículas que causan la absorción y el esparcimiento están uniformemente distribuidas a través del medio y que no hay interacciones entre ellas. Debido a la gran dificultad que supone resolver la ecuación de transporte radiativo, se han hecho aproximaciones con respecto a la radiancia o a la función de fase. El tipo de aproximación que se puede emplear para calcular la distribución de la luz a través de los tejidos biológicos, depende del tipo de irradiancia (difusa o colimada) y de si hay una variación del índice de refracción (Ishimaru, 1999).

2.5. Reconocimiento de patrones

Reconocimiento de Patrones es una área de la tecnología conocido como Aprendizaje de Maquinas (Machine Learning) o Aprendizaje Automático. El único propósito de este método es el clasificar un grupo de patrones conocido como conjunto de pruebas en dos o más clases de categorías. Esto es logrado al calcular las categorías del conjunto en prueba comparándolo con un conjunto de entrenamiento (previo) o training set. Un clasificador dado mide la distancia entre varios puntos dados (compara), para saber cuales puntos son más cercanos a la meta en un modelo parametrizado. En el reconocimiento de patrones se requieren extraer las características de los objetos, establecer un modelo y finalmente pre procesar y segmentar la información a través de un aprendizaje supervisado o no supervisado.

2.6. Componentes principales

Un problema central en el análisis de datos multidimensionales es la reducción de la dimensionalidad de p variables por un pequeño subconjunto $r < p$ de ellas

sin perder precisión, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información. Cabe destacar que para fines de esta investigación las p variables representan los rasgos espectrales, que contienen la información del fototipo de cada individuo en estudio.

El análisis de componentes principales tiene como objetivo representar adecuadamente n observaciones de p variables con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20 % de las originales) expliquen la mayor parte (más del 80 % de la variabilidad original). Esta técnica de componentes principales permite representar óptimamente en un espacio de dimensión pequeña observaciones de un espacio general p -dimensional. En este sentido, componentes principales es el primer paso para identificar las posibles variables latentes, o no observadas que generan los datos. A su vez se puede transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.

2.6.1. Cálculo generalizado de las componentes principales

El espacio de dimensión r que mejor representa a los puntos viene definido por los vectores propios asociados a los r mayores valores propios de la matriz \mathbf{S} . Estas direcciones se denominan direcciones principales de los datos y a las nuevas variables por ellas definidas componentes principales. En general, la matriz \mathbf{X} (y por tanto la \mathbf{S}) tiene rango p , existiendo entonces tantas componentes principales como variables que se obtendrán calculando los valores propios o raíces (Cuadras, 1991) características, $\lambda_1, \dots, \lambda_p$, de la matriz de varianzas y covarianzas de las variables \mathbf{S} mediante:

$$|\mathbf{S} - \lambda\mathbf{I}| = 0 \quad (2.21)$$

y sus vectores asociados son:

$$(\mathbf{S} - \lambda_i\mathbf{I}) a_i = 0 \quad (2.22)$$

Los términos λ_i son reales, al ser la matriz \mathbf{S} simétrica, y positivos, ya que \mathbf{S} es

definida positiva. Por ser \mathbf{S} simétrica si λ_j y λ_h son dos raíces distintas sus vectores asociados son ortogonales. Si \mathbf{S} fuese semidefinida positiva de rango $r < p$, lo que ocurriría si pr variables fuesen combinación lineal de las demás, habría solamente r raíces características positivas y el resto serían ceros.

Llamando Z a la matriz cuyas columnas son los valores de los p componentes de las n observaciones, estas nuevas variables están relacionadas con las originales mediante:

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \quad (2.23)$$

donde $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Calcular los componentes principales equivale a aplicar una transformación ortogonal \mathbf{A} a las variables \mathbf{X} (ejes originales) para obtener unas nuevas variables \mathbf{Z} incorreladas entre sí. Esta operación puede interpretarse como elegir unos nuevos ejes coordenados, que coincidan con los “ejes naturales” de los datos.

2.7. Conglomerados (clusters)

El análisis de conglomerados tiene por objeto agrupar elementos en grupos homogéneos en función de las similitudes o similaridades entre ellos. Normalmente se agrupan las observaciones, pero el análisis de conglomerados puede también aplicarse para agrupar variables (Cuadras, 1991). Estos métodos se conocen también con el nombre de métodos de clasificación automática o no supervisada, o de reconocimiento de patrones sin supervisión. El análisis de conglomerados estudia tres tipos de problemas:

- Partición de datos
- Construcción de jerarquías
- Clasificación de variables

Para el desarrollo de esta investigación se cuentan con datos heterogéneos y se desea dividirlos en un número de grupos prefijado que se corresponda por la escala de Fitzpatrick, de manera que: (1) cada elemento pertenezca a uno y solo uno de

los grupos; (2) todo elemento quede clasificado; (3) cada grupo sea internamente homogéneo. por lo tanto es conveniente implementar partición de datos utilizando algoritmos de k-medias.

2.7.1. Algoritmo de k-medias

Supongamos una muestra de n elementos con p variables. El objetivo es dividir esta muestra en un número de grupos prefijado, G . El algoritmo de k-medias requiere las cuatro etapas siguientes :

1. Seleccionar G puntos como centros de los grupos iniciales. Esto puede hacerse:
 - asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados;
 - tomando como centros los G puntos más alejados entre sí ;
 - construyendo los grupos con información a priori, o bien seleccionando los centros a priori.
2. Calcular las distancias euclídeas de cada elemento al centro de los G grupos, y asignar cada elemento al grupo más próximo. La asignación se realiza secuencialmente y al introducir un nuevo elemento en un grupo se recalculan las coordenadas de la nueva media de grupo.
3. Definir un criterio de optimalidad y comprobar si reasignando uno a uno cada elemento de un grupo a otro mejora el criterio.
4. Si no es posible mejorar el criterio de optimalidad, terminar el proceso.

El criterio de homogeneidad que se utiliza en el algoritmo de k-medias es la suma de cuadrados dentro de los grupos (SCDG) para todas las variables, que es equivalente a la suma ponderada de las varianzas de las variables en los grupos:

$$SCDG = \sum_{g=1}^G \sum_{j=1}^p \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2 \quad (2.24)$$

donde x_{ijg} es el valor de la variable j en el elemento i del grupo g y \bar{x}_{jg} la media de esta variable en el grupo. El criterio se escribe

$$\min SCDG = \min \sum_{g=1}^G \sum_{j=1}^p n_g s_{jg}^2 \quad (2.25)$$

donde n_g es el número de elementos del grupo g y s_{jg}^2 es la varianza de la variable j en dicho grupo. La varianza de cada variable en cada grupo es claramente una medida de la heterogeneidad del grupo y al minimizar las varianzas de todas las variables en los grupos obtendremos grupos más homogéneos. Un posible criterio alternativo de homogeneidad sería minimizar las distancias al cuadrado entre los centros de los grupos y los puntos que pertenecen a ese grupo. Si medimos las distancias con la norma euclídea, este criterio se escribe:

$$\min \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg})' (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg}) = \sum_{g=1}^G \sum_{i=1}^{n_g} d_{i,g}^2 \quad (2.26)$$

donde $d_{i,g}^2$ es el cuadrado de la distancia euclídea entre el elemento i del grupo g y su media de grupo. Es fácil comprobar que ambos criterios son idénticos. Como un escalar es igual a su traza, podemos escribir este último criterio como

$$\min \sum_{g=1}^G \sum_{i=1}^{n_g} \text{tr} [d_{i,g}^2] = \min \text{tr} \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg}) (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg})' \right] \quad (2.27)$$

y llamando \mathbf{W} a la matriz de suma de cuadrados dentro de los grupos,

$$\mathbf{W} = \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg}) (\mathbf{x}_{ijg} - \bar{\mathbf{x}}_{jg})' \quad (2.28)$$

tenemos que

$$\min \text{tr}(\mathbf{W}) = \min SCDG \quad (2.29)$$

Como la traza es la suma de los elementos de la diagonal principal ambos criterios coinciden. El algoritmo de kmedias busca la partición óptima con la restricción de que en cada iteración sólo se permite mover un elemento de un grupo a otro. El algoritmo funciona como sigue:

- (1) Partir de una asignación inicial
- (2) Comprobar si moviendo algún elemento se reduce \mathbf{W} .
- (3) Si es posible reducir \mathbf{W} mover el elemento, recalcular las medias de los dos grupos afectados por el cambio y volver a (2). Si no es posible reducir \mathbf{W} terminar.

En consecuencia, el resultado del algoritmo puede depender de la asignación inicial y del orden de los elementos. Conviene siempre repetir el algoritmo desde distintos valores iniciales y permutando los elementos de la muestra. El efecto del orden de las observaciones suele ser pequeño, pero conviene asegurarse en cada caso de que no está afectando.

2.8. Modelos de redes neuronales

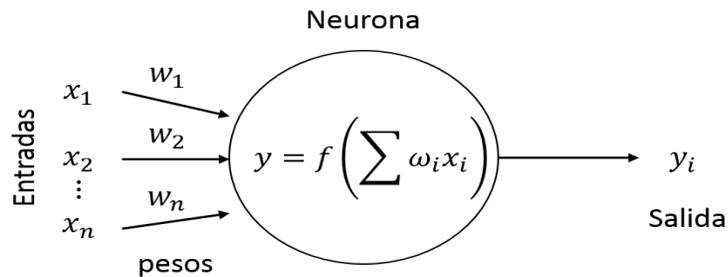


Figura 2.6: Esquema General de una red de McCulloch-Pitts. Fuente: (Viñuela and León, 2004)

Los modelos de Redes Neuronales provienen originalmente de los modelos en McCulloch and Pitts (1943), las neuronas son representadas por un nodo, las unidades de entradas generalmente se denotan por: x_1, x_2, \dots, x_n . Las y_i toman el valor de 1 o 0 dependiendo si la i -ésima neurona es alimentada o no. Las neuronas forman una suma pesada de las entradas (ver figura 2.6) y si esta suma es más grande que un umbral las neuronas han sido alimentadas, es decir:

$$w_1x_1 + w_2x_2 + \dots + w_nx_n > T$$

entonces la salida y de las neuronas, toma el valor 1, además estas pueden ser transmitidas como nuevas entradas a neuronas de capas posteriores de la red. Por otro lado sí:

$$w_1x_1 + w_2x_2 + \dots + w_nx_n \leq T$$

las neuronas no han sido alimentadas, la salida y toma el valor 0. La cantidad de la intensidad de la sinapsis entre la neurona i y la neurona j . Generalmente en los modelos de redes neuronales, las neuronas son arregladas en capas basadas en una estructura conocida con entradas en la primera capa y más capas ocultas y una capa de salida, entonces cada capa se comunica con la próxima hasta que la capa de salida sea alcanzada.

El umbral T puede ser ajustado en la mayoría de los procedimientos de entrenamiento de redes neuronales y por lo tanto el énfasis es sobre los pesos. Además T puede escribirse como: $T = w_0$ y representa el sesgo de la neurona.

x_0 toma siempre el valor de 1. La condición para que la salida de la neurona sea alimentada es:

$$w_0x_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n > 0$$

El proceso de ajustar los pesos en una red es considerado importante, ya que permite ejecutar diversas tareas que son llamadas entrenamiento de la red. Para este propósito, un conjunto de datos es usado, los cuales consisten en una lista de entradas conjuntamente con las salidas deseadas. Durante el proceso de entrenamiento las entradas son presentadas en algún orden, usualmente en forma aleatoria. Después de cada presentación, los pesos de la red son ajustados usando un algoritmo especificado, por lo que la ejecución es mejorada para la entrada presentada. Después de suficientes presentaciones, la red es entrenada y esta puede ser usada para predecir nuevos casos.

En Searle (1994) se define las redes neurales como una clase de modelos flexibles de regresión no lineal y modelos de discriminación, modelos de reducción de datos y sistemas dinámicos no lineal; que consisten en un gran número de neuronas interconectadas en formas complejas y organizadas en capas. Por otra parte, establece que las redes neurales son utilizadas en tres formas principales:

1. Como modelos de sistemas biológicos y de inteligencia.
2. Como procesadores de señales adaptativa en tiempo real o como controladores.
3. Como métodos analíticos de datos.

Hilera and Martínez (1995) se señala que el algoritmo de aprendizaje del perceptron es de tipo supervisado, lo cual requiere que los resultados sean evaluados y por lo tanto se realicen las modificaciones del sistema si fuera necesario. Los valores de los pesos se pueden fijar o adaptar utilizando diferentes algoritmos de entrenamiento de la red. El perceptron se caracteriza por separar las regiones de hiperplano cuya ecuación queda determinado por los pesos de las cone-xiones y el valor del sesgo de la función activación de la neurona. En el caso de dos entradas, la separación es llevada a cabo por un plano y así sucesivamente hasta considerar el caso de n entradas la división es realizada por un hiperplano.

Searle (1994) establece que el perceptron puede tener una o más salidas. Cada salida tiene un sesgo separado y un conjunto de pesos. Usualmente la misma función de activación es usada para cada salida, aunque es posible usar diferentes funciones de activación.

2.8.1. Perceptrón unicapa

Como su nombre lo indica está constituido por una capa de neuronas, llamada capa de entrada. En las redes neuronales las conexiones entre neuronas se realizan de una capa a la capa siguiente, por lo tanto un Perceptrón Unicapa es un conjunto de neuronas no unidas entre sí, de manera que cada una de las entradas del sistema se conectan a cada neurona, produciendo cada una de ellas su salida individual (ver figura 2.7).

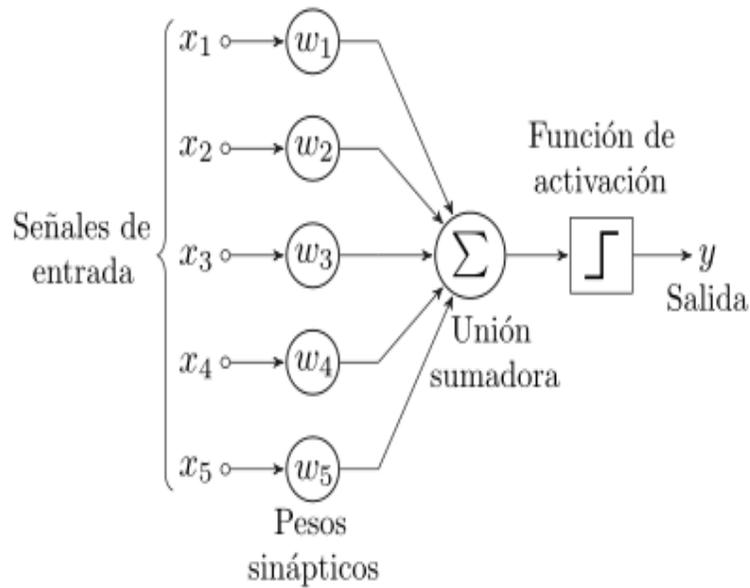


Figura 2.7: Arquitectura del perceptrón unicapa. Fuente: (Wikipedia, 2017)

El modelo para el Perceptrón Unicapa es dado por:

$$p_j = f_j(a_i + \sum_{j=1}^{n_x} w_{ij}x_i)$$

donde:

n_x = número de variables independientes (cantidad de entradas).

x_i = variables independientes (entradas).

a_i = bias o sesgos de la capa de salida.

w_{ij} = pesos de la capa de entrada a la capa de salida.

q_j = es la entrada a la capa de salida = $a_i + \sum_{j=1}^{n_x} (w_{ij}x_i)$.

p_j = son los valores calculados (Salidas calculadas) = $f(q_j)$.

y_j = son los valores esperados (Salidas esperadas).

r_j = error residual = $(y_j - p_j)$.

Existen varias funciones de activación o transferencia f típicas que determinan distintos tipos de neuronas:

1. Umbral:

$$f(x) = \begin{cases} -1 & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$$

2. Lineal o Identidad:

$$f(x) = x$$

3. Tangente Hiperbólica:

$$f(x) = \tanh(x).$$

4. Logística:

$$f(x) = \frac{\exp(-x)}{1 + \exp(-x)} = \frac{\tanh(\frac{x}{2}) + 1}{2}.$$

5. Gaussiana:

$$f(x) = \exp\left(\frac{-x^2}{2}\right).$$

6. Sigmoidal:

$$f(x) = \frac{1}{1 + \exp(-x)}.$$

2.8.2. Perceptrón multicapa

Es un conjunto de neuronas agrupadas en varios niveles o capas. Una forma para distinguir la capa a la que pertenece una neurona, consiste en fijarse en el origen de las señales que recibe a la entrada y el destino de la señal de salida. Esta red permite establecer regiones de decisión mucho más complejas que el perceptrón de una sola capa. Puede tener múltiples entradas y salidas, el número de capas ocultas pueden ser diferentes en lo que se refiere al número de entradas con respecto al número de

salidas. Conjuntamente se puede hacer la conexión directa de la capa de entrada a la capa de salida (ver figura 2.8)

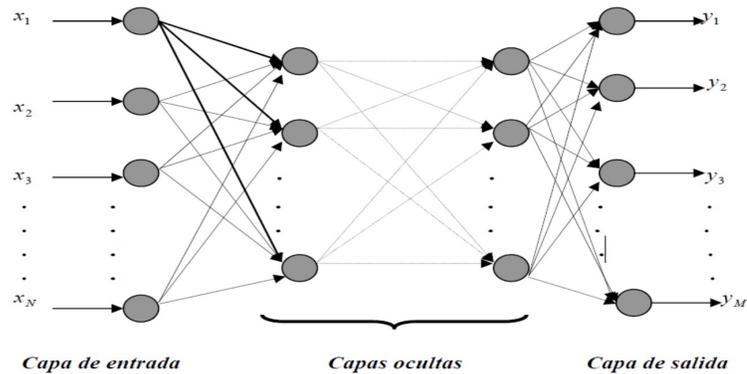


Figura 2.8: Arquitectura del perceptrón multicapa. Fuente: (Viñuela and León, 2004)

En consecuencia existen dos tipos de conexiones: hacia adelante y hacia atrás. En la conexión hacia adelante, todas las neuronas de una capa reciben señales de entrada de otra capa anterior, más cercanas a la entrada de la red y envían las señales de salida a una capa posterior, más cercana a la salida de la red hacia atrás, se conectan las salidas de las neuronas de capas posteriores.

Los perceptrones multicapas son modelos no lineales que dadas suficientes neuronas y datos, pueden aproximar una función con el grado de exactitud deseado. Además pueden constituirse en una herramienta de mucha utilidad cuando se tiene poco conocimiento a cerca de la relación entre variables independientes y las variables dependientes.

Searle (1994) señala que la complejidad de los modelos se obtienen variando el número de capas ocultas y el número de neuronas en cada capa. Cuando se tiene un número pequeño de neuronas ocultas, el perceptron multicapas es un modelo paramétrico. Con un número moderado de neuronas, el perceptrón multicapas puede ser considerado un modelo semiparamétrico. Si el número de neuronas ocultas aumentan conjuntamente con el tamaño de la muestra, el perceptron multicapas puede ser considerado como un modelo no paramétrico.

El modelo con 1 capa oculta es dado por:

$$p_j = f_0\left[c_k + \sum_{j=1}^{n_h} w_{kj} f_j(a_j + \sum_{i=1}^{n_x} b_{ij} x_i)\right]$$

donde:

- n_x = número de variables independientes (Cantidad de entradas).
- n_h = número de neuronas ocultas.
- x_i = variables independientes (Entradas).
- a_i = bias o sesgos de la capa oculta.
- b_{ij} = pesos de la capa de entrada a la capa oculta.
- g_j = entrada a la capa oculta = $a_i + \sum_{j=1}^{n_x} b_{ij} x_i$.
- h_j = valores de la capa oculta = $f_h(g_j)$.
- c_k = bias o sesgos de la capa de salida.
- w_{ij} = pesos de la capa oculta a la capa de salida.
- q_j = entrada a la capa de salida = $c_k + \sum_{j=1}^{n_h} (w_{kj} h_k)$.
- p_j = valores calculados (Valores predichos) = $f_0(q_j)$.
- y_j = variables dependientes (Valores entrenados).
- r_j = residual = error = $(y_j - p_j)$.

Los perceptrones multicapas son utilizados en la práctica, debido a que la función de activación es continua y diferenciable y además permite un criterio de ajuste del error con respecto a los pesos y estos son estimados recursivamente de una entrada a una salida usando el algoritmo propacación hacia atrás.

2.9. Máquinas de Soporte Vectorial (SVM)

Las Máquinas de Soporte Vectorial o SVM (por sus siglas en inglés) son algoritmos (máquinas) que aprenden a discriminar entre miembros positivos y negativos de una clase de vectores de m -dimensional dada. Fue desarrollada por Vapnik and Lerner (1963), pero se generaliza en los años noventa por Boser. (1992) y Cortes and Vapnik (1995). Fue ideada originalmente para la resolución de problemas de clasificación binarios en los que las clases eran linealmente separables. Una SVM construye

un hiperplano que puede ser utilizado para clasificar datos, por esta razón también se le conoce como Hiperplanos Óptimo de Separación o OSH (“Optimal Separating Hyperplane”), ya que la solución obtenida clasificaban de manera correcta todas las muestras en el espacio de entrada, colocando el hiperplano de separación lo más lejos posible de todas ellas (Burges, 1998). Para lograr esto, se debe realizar un entrenamiento con los datos dispuestos para ese fin, a partir del cual se define el hiperplano óptimo, es decir, el plano que maximiza el margen de separación entre las muestras de cada clase.

El objetivo es que después de este entrenamiento, la máquina generalice bien para datos nuevos que no han participado en el entrenamiento, es decir, clasifique correctamente cada una de las muestras a la clase a la que pertenecen.

2.9.1. Caso linealmente separable

Para solucionar un problema de clasificación binaria, la SVM debe establecer una superficie de decisión adecuada, basándose en el conjunto de datos de entrenamiento. La superficie de decisión es un hiperplano que separa los datos de entrenamiento en dos clases.

Supongamos se dispone de un conjunto de n datos de entrenamiento, de la forma $(z_1, y_1), (z_2, y_2), \dots, (z_n, y_n)$ donde $z \in R^m$. Cada escalar y_i (llamado etiqueta) corresponderá a una de las dos clases que se identificarán como $+1$ y -1 . Se llamará vector de etiquetas al vector $y = (y_1, y_2, \dots, y_n)$.

Un conjunto de vectores $\{(z_1, y_1), (z_2, y_2), \dots, (z_n, y_n)\}$ donde $z_i \in R^m$ e $y_i \in \{-1, +1\}$ para $i = 1, \dots, n$ se dice linealmente separable si existe algún hiperplano en R^m que separa los vectores $Z = \{z_1, z_2, \dots, z_n\}$ con etiquetas $y_i = 1$ de aquellos con etiqueta $y_i = -1$.

En un problema linealmente separable existen muchos hiperplanos que pueden clasificar los datos. Pero las SVM no hallan uno cualquiera de estos hiperplanos, sino

el único que maximiza la distancia entre él y el dato mas cercano de cada clase. Esta distancia es llamada margen, y el hiperplano que la maximiza se le llama Hiperplano de Máximo Margen o Separación Óptima (HSO).

En la Figura 3-1 mostrada a continuación se puede observar un caso en R^2 de datos con dos clases separadas por dos diferentes márgenes de separación. Como se ha dicho, el hiperplano óptimo será aquel que maximice el margen de separación entre las dos clases, y como se aprecia en este ejemplo, el margen 1 pareciera ser el que mejor separa las dos clases. También se puede observar que solo tres vectores (z_i) se encuentran justo a la distancia del margen 1 y 2.

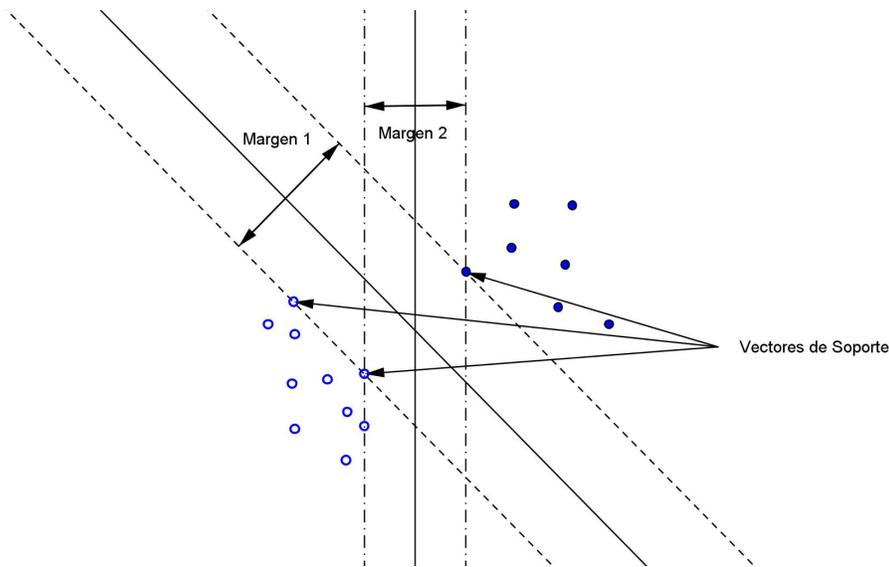


Figura 2.9: Maximización del margen. Fuente: (Burges, 1998)

El hiperplano separador está dado de manera general por

$$H : z^t w + b = 0 \quad (2.30)$$

siendo b una constante que indica la posición del plano respecto al origen de coordenadas. Esta constante recibe el nombre de sesgo y define el umbral de decisión, w es el vector normal al hiperplano. El trabajo consiste en hallar el vector w que es normal

al hiperplano, pero en el fondo es un vector de pesos que contiene la ponderación de cada dato, indicando qué tanto aportan en el proceso de clasificación. La función discriminante (distancia) d será la función:

$$d(z, w, b) = \frac{|z^t w + b|}{\|w\|} \text{ con } \|w\| = \sqrt{w^t w} \quad (2.31)$$

donde $\|w\|$ es la norma asociada al producto escalar en R^m . Representando H_1 y H_2 los hiperplanos canónico tal como se aprecia en la Figura

refimagen2, la distancia perpendicular del origen al hiperplano H_1 ($z_i^t w + b = +1$) es $\frac{|1-b|}{\|w\|}$ y del origen al hiperplano H_2 ($z_i^t w + b = -1$) es $\frac{|-1-b|}{\|w\|}$. Aquellos objetos que se encuentran sobre los hiperplanos H_1 y H_2 son llamados Vectores de Soporte.

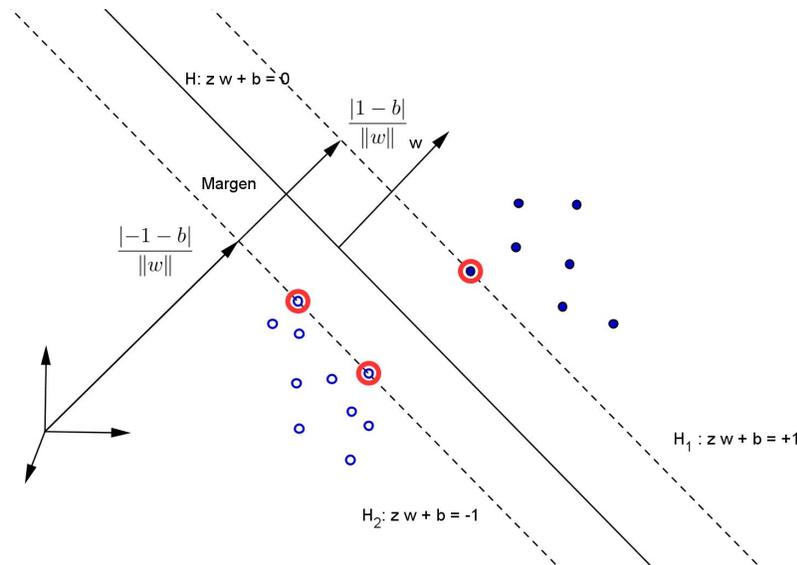


Figura 2.10: Caso linealmente separable. Fuente: (Burges, 1998)

Eliminar cualquiera de los puntos que no se encuentran sobre H_1 y H_2 no afecta el resultado de la clasificación, pero eliminar alguno de los vectores de soporte si. La distancia entre los dos hiperplanos H_1 y H_2 es $\frac{2}{\|w\|}$, donde $\frac{1}{\|w\|}$ es la distancia entre

el hiperplano definido por H respecto a cada uno de los hiperplanos canónicos. El objetivo para lograr una correcta clasificación es maximizar $\frac{2}{\|w\|}$ lo cual es equivalente a minimizar $\frac{\|w\|^2}{2}$. De esta manera, el problema de optimización a tratar es el siguiente:

$$(P) \underset{w,b}{\text{minimizar}} \quad \frac{\|w\|^2}{2} \quad (2.32)$$

$$\text{sujeto a} \quad y_i(z_i^t w + b) - 1 \geq 0 \quad \text{para } i = 1, \dots, n$$

donde n es el número de muestras de entrenamiento.

Este problema puede resolverse usando las técnicas de programación cuadrática (“Quadratic Programming” o QP por sus siglas en inglés). Esta solución se puede caracterizar a través de las condiciones de optimalidad de primer orden, conocidas como las condiciones de Karush, Kuhn y Tucker (KKT) (Kuhn and Tucker, 1951). Una vez obtenidos el vector w y la constante b la clasificación de nuevos datos se realiza mediante la función:

$$h(z) = \text{sgn} \left(\sum_{i \in N_{SV}} y_i x_i \langle z, z_i \rangle + b \right) = \text{sgn}(H) \quad (2.33)$$

donde un nuevo dato es clasificado como $y_i = h(z_i) = +1$ si $H > 0$ y $y_i = h(z_i) = -1$ si $H < 0$.

2.9.2. Caso no lineal

En la mayoría de las aplicaciones del mundo real, el espacio de entradas Z por lo general, es un conjunto que no puede ser separado linealmente por un hiperplano gracias a la naturaleza de los datos, y debido a esta limitación, su tratamiento consiste en utilizar una función $\phi : R^m \rightarrow R^f$, con $m, f \in N$, $f \geq m$ ó $f = \infty$. Dicha función mapea la información de los datos de entrada a una de mayor dimensión llamado espacio de características y tiene como objetivo hacer que el conjunto obtenido $\phi(Z)$ sea un conjunto linealmente separable o linealmente no separable donde se minimice el error de aquellos datos mal clasificados, es decir que el número de

vectores clasificados incorrectamente sea mínimo.

$$Z = \{z_1, z_2, \dots, z_n\} \rightarrow \phi(Z) = \{\phi(z_1), \phi(z_2), \dots, \phi(z_n)\} \in R^f$$

Pero como no es posible determinar a priori, dado el conjunto de entrada Z , una función ϕ que cumpla los objetivos descritos previamente, la representación por medio de funciones Kernel ofrece una solución al producto interno de las transformaciones de cada dato consideradas en su problema dual:

$$\begin{aligned} (D) \text{ maximizar}_x \quad & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i x_j y_i y_j \phi(z_i)^t \phi(z_j) + \sum_{i=1}^n x_i \\ \text{sujeto a} \quad & \sum_{i=1}^n x_i y_i = 0 \\ & 0 \leq x_i \leq C, \quad i = 1, \dots, n. \end{aligned} \tag{2.34}$$

En la Figura 2.11 se muestra un mapeo de un espacio de entradas de dos dimensiones a un espacio de características de tres dimensiones, donde la información no puede ser separada por una máquina lineal en el espacio de entrada mientras que en el espacio de características esto resulta muy sencillo.

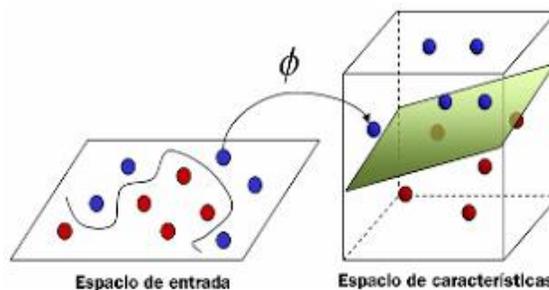


Figura 2.11: Mapeo del Espacio de entrada a un Espacio de características . Fuente: (Burges, 1998)

Existen distintas funciones Kernel que permiten adaptar la SVM a cada conjunto de muestras, con el fin de obtener mejores resultados. Algunos ejemplos de funciones Kernel (Schölkopf and Smola, 2002) que han sido sugeridas o empleadas en proble-

mas de clasificación son las siguientes:

- Kernel Lineal: $K(z_i, z_j) = z_i^t z_j$
- Kernel Polinomial: $K(z_i, z_j) = (r + \gamma z_i^t z_j)^d$, $\gamma > 0$, $r, d \in R$
- Kernel Gaussiano: $K(z_i, z_j) = c \cdot \exp(-\gamma \|z_i - z_j\|_2^2)$, $\gamma > 0$, $c \in R$
- Kernel Neural (Sigmoid, Tanh): $K(z_i, z_j) = \tanh(a z_i^* z_j + b)$
- Kernel Anova: $K(z_i, z_j) = \sum_i \exp(-\gamma z_i - z_j)^d$
- Kernel series de Fourier: $K(z_i, z_j) = \frac{\sin(N + \frac{1}{2})(z_i - z_j)}{\sin(\frac{1}{2}(z_i - z_j))}$

Una propiedad de las SVM en el caso lineal es que éstas pueden ser expresadas en una representación dual, esto significa que la ecuación 2.33 puede ser expresada como una combinación lineal de los puntos de entrenamiento. Por lo tanto, la regla de decisión puede ser evaluada usando productos punto

$$h(z) = \operatorname{sgn} \left(\sum_{i \in N_{SV}} y_i x_i \phi(z)^t \phi(z_i) + b \right) \quad (2.35)$$

Pero gracias a los Kernels, (2.35) se convierte en:

$$h(z) = \operatorname{sgn} \left(\sum_{i \in N_{SV}} y_i x_i K(z, z_i) + b \right) \quad (2.36)$$

Así mismo el problema dual de optimización para el caso no lineal esta dado por:

$$\begin{aligned} (D) \text{ maximizar } & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i x_j y_i y_j K(z_i, z_j) + \sum_{i=1}^n x_i \\ \text{sujeto a } & \sum_{i=1}^n x_i y_i = 0 \\ & 0 \leq x_i \leq C, \quad i = 1, \dots, n. \end{aligned} \quad (2.37)$$

Ahora 2.37 en términos de minimización y en forma matricial el problema de optimización cuadrática convexa que surge de las SVM es:

$$\begin{aligned}
 (D) \text{ minimizar } & \frac{1}{2}x^tQx - e^tx \\
 \text{sujeto a} & \quad y^tx = 0 \\
 & \quad 0 \leq x_i \leq C, \quad i = 1, \dots, n
 \end{aligned} \tag{2.38}$$

donde $Q_{i,j} = y_i y_j K(z_i, z_j)$.

Para hallar el umbral b se promedia

$$b = y_j - \sum_{i \in N_{SV}} x_i y_i K(z_i, z_j) \tag{2.39}$$

para todos los puntos con $x_j > 0$.

2.9.3. Enfoques previos para entrenar una SVM

Entrenar una SVM es equivalente a resolver un problema de QP n-dimensional, donde n es el número de muestras en la matriz de datos de entrenamiento. Resolver este problema usando las técnicas tradicionales para QP involucra un número grande de operaciones, lo cual es muy costoso en tiempo y recursos computacionales; y en la mayoría de los casos esto no resulta práctico para problemas grandes. Producto a esta limitante se han creado varios algoritmos para solucionar el problema del entrenamiento de las SVM. Un pseudoalgoritmo para calcular SVM sería:

1. Seleccionar el parámetro C (que representa un compromiso entre minimizar el error de entrenamiento y la maximización del margen de separación entre los elementos de las clases), la función Kernel y cualquier otro parámetro requerido por la función Kernel empleada.
2. Resolver el problema de programación cuadrática o la formulación alternativa usando el algoritmo apropiado de programación cuadrática para obtener los

vectores de soporte.

3. Obtener el umbral b usando los vectores de soporte ya calculados.
4. Obtener la función de decisión.

Existen varios algoritmos para el entrenamiento de SVM. Algunos de ellos usan algoritmos de descomposición que son procedimientos iterativos que optimizan solamente un pequeño subconjunto de las variables en cada iteración. Por los resultados reportados se destacan:

2.9.4. Librería de Máquina Soporte Vectorial (LIBSVM)

LIBSVM es una librería para SVM creada en 2001 por Chih-Chung Chang y Chih-Jen Lin (Hsu et al., 2010). Se trata de un programa que soporta estimación de distribución, clasificación y regresión de soporte de vectores. La mayoría de los métodos de descomposición obtienen un conjunto de datos inicial a partir del conjunto de datos entero, este conjunto de datos es optimizado en cada iteración, mejorando el valor de la función objetivo en cada iteración. El proceso iterativo llega a su fin cuando un criterio de parada derivado de las condiciones de Karush-Kuhn-Tucker es satisfecho o una precisión requerida es alcanzada. El algoritmo LIBSVM está basado en el algoritmo SMO, sin embargo, posee una estrategia de descomposición llamada selección de conjunto de trabajo (Keerthi et al., 2001) el cual es mucho más avanzado. LIBSVM emplea un algoritmo de dirección de búsqueda que minimiza la función objetivo en cada iteración.

2.10. Curvas ROC

El acrónimo ROC del inglés Receiving Operating Characteristic (Características de funcionamiento de recepción) también conocida como curva de rendimiento

diagnóstico (ver figura 2.12), es una representación gráfica de la tasa de éxito (probabilidad de detectar correctamente una señal cuando dicha señal está efectivamente presente) frente a la tasa de falsa alarma (probabilidad de detectar una señal cuando efectivamente no esta presente) para tareas de detección con solo dos resultados posibles (sí/no, presente/ausente, 1/0) lo que se conoce como clasificación binaria, sin embargo cuando se desea realizar una curva ROC para una clasificación multi-clase, se transforma los datos en una clasificación binaria utilizando el criterio de uno contra todos, es decir, se asume la primera clase como 1 y el resto de las clases como 0 y se determina la ROC para este sistema, posteriormente se toma la segunda clase y se asume 1 y las otras como 0, levantando la curva ROC; este procedimiento se realiza tantas clases como tenga el sistema y posteriormente se estima una curva ROC multiclase a partir del promedio estadístico.

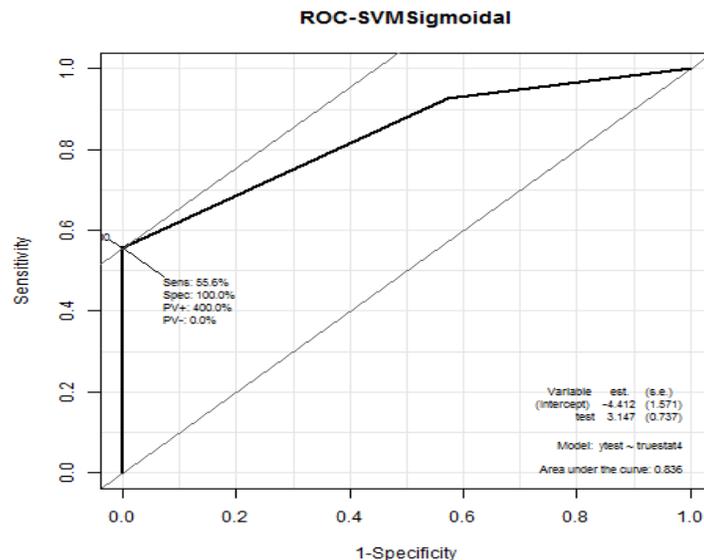


Figura 2.12: Representación de una Curva ROC para una buena clasificación. Fuente: (Narea, 2017)

En medicina se utilizan de forma cotidiana los conceptos de sensibilidad y especificidad de una prueba diagnóstica, donde la sensibilidad es la proporción de casos diagnosticados como afirmativos, a partir del criterio o regla de decisión establecido,

en los que se comprueba que efectivamente sucede el estado que se pretende detectar o diagnosticar. Y la especificidad es la proporción de casos diagnosticados como negativos, a partir de la regla de decisión establecida, en los que se comprueba que efectivamente no sucede el estado que se pretende detectar o diagnosticar.

La curva ROC se puede interpretar como un gráfico de las tasas de verdaderos positivos frente a falsos positivos asociadas con todos los posibles puntos de corte o umbrales para clasificar un evento u objeto como positivo. Siempre se encuentra en el espacio cuadrado entre $(0, 0)$ y $(1, 1)$, al que algunas veces se denomina *espacio ROC*, y tradicionalmente se dibuja también la diagonal positiva para indicar el punto de nula capacidad predictiva. Efectivamente, si la sensibilidad iguala a 1-especificidad a lo largo de todos los puntos de corte es que nuestra prueba simplemente selecciona unos casos de otros por puro azar.(Prati et al., 2008)

El área bajo la curva ROC, o AUC (siglas del nombre inglés Area Under Curve), es el estadístico que proporciona una medida completa de la capacidad predictiva de un instrumento o sistema diagnóstico. Cuando el criterio de medida o indicador diagnóstico no puede distinguir entre los dos grupos o resultados, esto es, cuando no hay diferencia entre las dos distribuciones, el área será 0,5, y la curva ROC coincidirá con la diagonal positiva del *espacio ROC*. Y en el otro caso extremo, cuando la capacidad predictiva sea perfecta, esto es, haya perfecta separación, ningún solapamiento entre las dos distribuciones, el área bajo la ROC es igual a 1, y alcanzará perfectamente la esquina superior izquierda del *espacio ROC*. (Prati et al., 2008)

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Tipo de estudio y de investigación

El tipo de estudio según el origen de los datos es de laboratorio, de acuerdo a [Hurtado \(2010\)](#) en este tipo de diseño “se recogen datos a partir de fuentes vivas o materiales, pero cuando éstas se encuentran en un ambiente creado artificialmente para efectos de la investigación, o en un ambiente que no suele ser el contexto habitual para las unidades de estudio” (p.694). En este sentido los datos espectrales fueron recabados con un espectrofotometro de reflexión difusa MiniScan XEPLUS, que consta de una fuente de luz artificial controlada y estandarizada.

Ahora bien, según la temporalidad y la secuencialidad de las mediciones el diseño es transeccional según [Hurtado \(2010\)](#) “el interés del investigador en este tipo de diseño se centra en estudiar el evento en un momento único en el tiempo” (p.694). Por tal razón los datos espectrales representan una medida única del fototipo cutáneo de los distintos pacientes en estudio, ya que el fototipo cutáneo es una condición de la piel invariante en el tiempo en regiones de la piel que son pocas expuestas a la luz solar.

La investigación presentada es de tipología predictiva, según [Hurtado \(2010\)](#): “Una investigación predictiva es aquella que se vale de los pronósticos que representan juicios razonados sobre un asunto importante que se tomará como base de algún

programa de acción, este tipo de investigación se encarga de anticipar el futuro probable basándose en probabilidades y sustentándose en la identificación de procesos explicativos”

Con base en ello, este trabajo se clasificaron los fototipos cutáneos mediante aprendizaje de máquinas, lo cual permitirá a los dermatólogos establecer de forma cuantitativa las distintas tipologías de piel, sin depender del espacio, ni de las condiciones ambientales donde se realicen los estudios.

3.2. Procedimiento Metodológico

Conocer el estado del arte de los fundamentos del aprendizaje de máquinas y sus principales aplicaciones

En esta etapa se inició con la recopilación bibliográfica de documentos o artículos nacionales como internacionales, publicados en diferentes medios como revistas científicas e Internet, acerca del aprendizaje de máquina, minería de datos, parámetros físicos y médicos que permitan la clasificación dermatológica de la piel asistida por computadora .

En este sentido los documentos científicos de interés se buscaron en revistas en OSA Publishing's Digital Library, SPIE Digital Library, Journal of Dermatological Science - Elsevier y en IEEE Xplore Digital Library que son bases de datos de revistas y libros arbitrados internacionalmente.

Selección de los principales rasgos de los espectros de reflexión difusa de distintos fototipos cutáneos utilizando técnicas de reducción de dimensión

Los datos espectrales contienen información de la interacción de la componente visible del espectro electromagnético con la piel humana en cada una de sus longitudes de ondas. La información espectral se obtiene usando un espectrofotómetro de esfera integradora con una resolución de 10nm, lo cual permite establecer 31 rasgos espectrales que contienen la información sobre el contenido de hemoglobina, bilirru-

bina, melanina, entre otros parámetros de la piel. Biológicamente el fototipo cutáneo es definido por la melanina contenida en la piel se puede reducir el rango espectral a la región en la cual la melanina predomina en las curvas espectrales, por otro lado utilizando análisis de componentes principales se podrá reducir estadísticamente las dimensiones de los rasgos espectrales. Adicionalmente se pudo reducir las dimensiones espectrales recuperando un parámetro óptico conocido como coeficiente de absorción de la epidermis (Narea et al., 2015); todo esto con la finalidad de definir cuales son los mejores rasgos para el entrenamiento de las redes neuronales y la máquina de soporte vectorial.

Implementación de redes neuronales artificiales y máquina de soporte vectorial para la detección de los distintos fototipos cutáneos

En esta fase de la investigación se implementaron algoritmos de redes neuronales y máquina de soporte vectorial en lenguaje R (R Core Team, 2016) que permitieron realizar una clasificación multiclase de los distintos fototipos cutáneos. Se utilizó el 70 % uniformemente distribuido de la muestra en estudio para entrenar el sistema y el 30 % restante para el proceso de validación. Los datos de entrada para el entrenamiento tanto de la red neuronal como de la máquina de soporte vectorial, dependieron de la selección de los principales rasgos de las curvas espectrales.

Comparación del desempeño de los distintos modelos en la clasificación de los fototipos implementados

En este paso se analizó el desempeño entre un perceptrón multicapa, así como el de las máquinas de soporte vectorial usando diferentes kernel con validación cruzada (cross validation). El análisis se realizó estudiando las curvas de características de funcionamiento del receptor (ROC, por sus siglas en inglés, Receiver Operating Characteristic). Con esto se estableció cual máquina de aprendizaje presenta un mejor desempeño para clasificar los distintos fototipos cutáneos.

3.3. Población y muestra

En el desarrollo de esta fase se expone el contexto en el que se realizó la investigación en función a una población y a una muestra respectivamente, en consecuencia, Hurtado (2010) expresa “la población es el conjunto de unidades de estudio que forman parte del contexto donde se quiere investigar el evento, los cuales comparten características comunes llamadas también criterios de inclusión” (p.268). En este sentido de acuerdo al objetivo planteado inicialmente en el trabajo, la población está constituida aproximadamente por 10000 pacientes atendidos en la Ciudad Hospitalaria Dr Enrique Tejera, en el Servicio de Dermatología entre los años 2014-2015. De los cuales se considero una muestra de 1005 pacientes a los cuales se les determino la firma espectral (datos espectrales) de los distintos fototipos cutáneos, de los mismo se utilizaron el 70 % de la muestra para entrenar las maquinas de aprendizajes y el 30 % restante de la muestra seleccionada para el proceso de validación del entrenamiento.

CAPÍTULO IV

RESULTADOS

En este capítulo se presentaran los resultados de la clasificación de los fototipos cutáneos utilizando Máquina de Soporte Vectorial con kernel lineal, función de base radial y sigmoïdal y una Red Neuronal multicapa con función de activación logística, utilizando criterios supervisados (basados en el criterio de clasificación del médico especialista) y no supervisados (basados en reagrupamientos por conglomerados) que permitirán establecer a priori una re-clasificación de los datos espectrales para cada fototipo.

4.1. Distribución de datos espectrales

Los datos espectrales contienen la información del fototipo cutáneo de una muestra de 1005 individuos, los cuales de acuerdo al criterio médico se distribuyen como se muestra en la gráfica 4.1, en la cual se evidencia que el fototipo I representa el 2,5%, fototipo II el 9.5% fototipo III 36,7%, fototipo IV 39,8%, fototipo V 8,4% y el fototipo VI 3,2% de los datos totales. Se puede observar de la gráfica 4.1 que la mayor distribución de datos se encuentra entre los fototipos III y IV sumando el 76,5% de los datos totales, de lo cual se puede establecer que los fototipos cutáneos de los venezolanos evaluados en este estudio son principalmente tipo III y IV de acuerdo a la escala de Fitzpatrick.

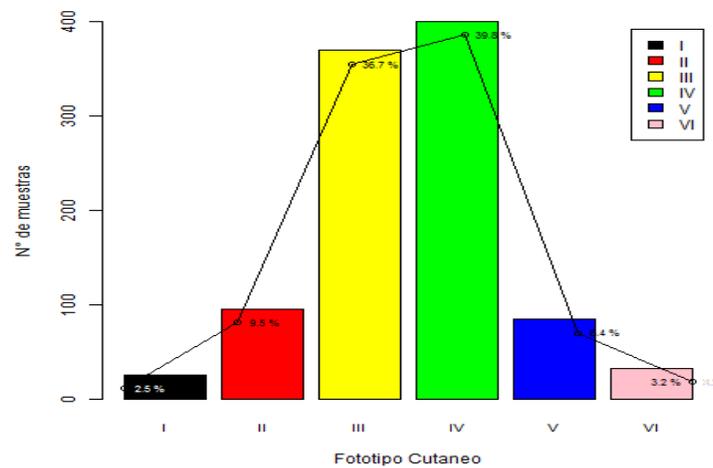


Figura 4.1: Distribución de datos por Fototipo cutáneo. Fuente: (CIMBUC-CHET-2015)

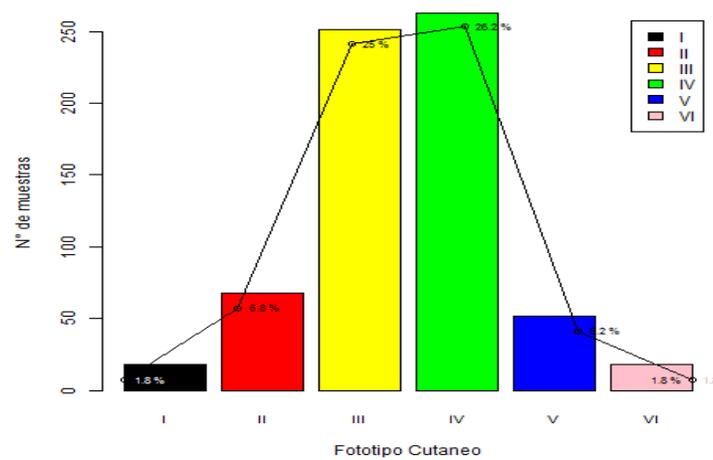


Figura 4.2: Distribución de datos de entrenamiento por Fototipo cutáneo. Fuente:(Narea, 2017)

Asimismo los datos totales se particionan en dos grandes grupos (entrenamiento

y prueba) en los cuales se garantiza información de todos los fototipos. La distribución de los datos de entrenamientos (670 datos) y de prueba (335 datos) se pueden observar en las gráficas 4.2 y 4.3 respectivamente. Esta distribución de datos se mantiene igual para el entrenamiento de las máquinas de aprendizajes utilizadas en cada experimento.

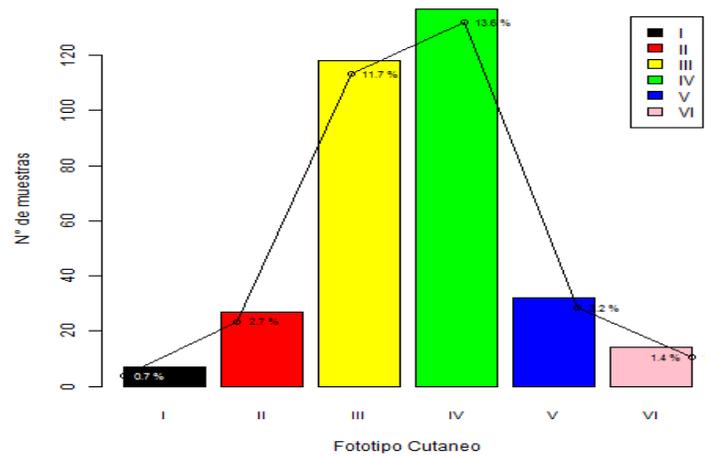


Figura 4.3: Distribución de datos de prueba por Fototipo cutáneo. Fuente: (Narea, 2017)

4.2. Aprendizaje de máquinas

Para encontrar el mejor clasificador del fototipo se utilizan como máquinas de aprendizajes las SVM y RN

4.2.1. SVM

Para implementar una SVM se utiliza el paquete *e1071* en R, que se basa en la librería LIBSVM (Meyer et al., 2015) que se hace mención en el capítulo II sección 2.9.4, del cual se utilizaron tres distintos kernels, ajustando los parámetros del kernel utilizando correlación cruzada de 10-fold.

Los parámetros para ajustar los kernels son los siguientes:

- $\gamma = 0,001 \rightarrow 2,5$
- $C = 0,001 \rightarrow 1000$
- $\alpha = 0,001 \rightarrow 2,5$

donde C es el parámetro de regularización en termino de la formulación de Lagrange para todos los kernels, $gamma$ define el ancho de la función los kernels RBF y sigmoidal, mientras que α es un coeficiente para el kernel sigmoidal (ver los apéndices A, B y C).

4.2.2. RN

Para generar la clasificación con Redes Neuronales se utilizo el paquete *nnet* en R que se basa en un percentrón multicapa con una sola capa oculta y función de activación logística (Venables and Ripley, 2002). En el cual se optimiza el numero de neuronas en la capa oculta (N) y la velocidad de aprendizaje (η) (ver los apéndices A, B y C).

4.3. Experimentos

Los datos espectrales tienen 31 rasgos que definen el comportamiento espectral de la piel de los individuos en estudio, partiendo de esto se puede establecer un primer experimento en el cual se utilicen los 31 rasgos para generar un clasificar y evaluar cual es el más eficiente.

El fototipo cutáneo esta definido fuertemente por la cantidad de melanina de la piel, que la principal causante de absorción de la luz en la piel. Sin embargo en los datos espectrales existe información de otros agente que absorben, como es el caso de la hemoglobina, que presenta un alto pico de absorción en el rango espectral de

400 a 450 nm respecto a la absorción de la melanina (ver gráfico 4.4).

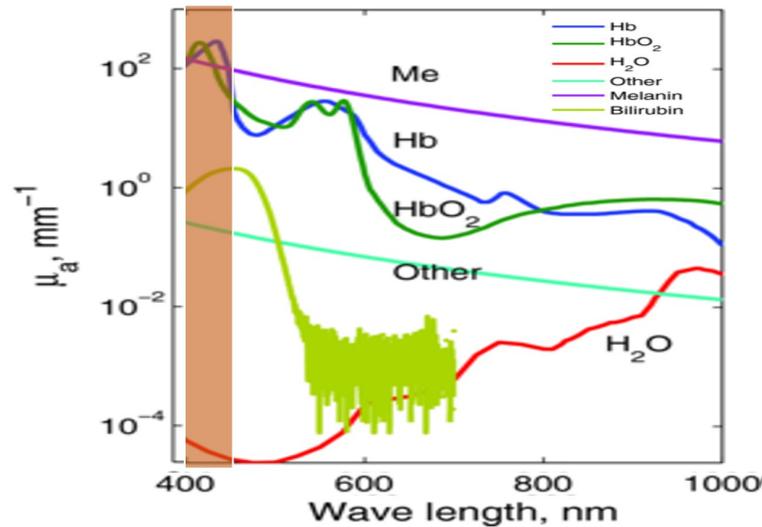


Figura 4.4: Coeficiente de absorción Melanina-Hemoglobina.

Debido a esto una región espectral de los datos se ve influenciado por este pico de hemoglobina, por tal motivo se plantea un segundo experimento donde se eliminan de los datos espectrales los primeros 5 rasgos que contienen en mayor medida información de la hemoglobina, resumiendo la información espectral en 26 rasgos.

En este sentido ya que la melanina es el principal agente de absorción en la piel, la cual se produce por lo melanosomas en la capa basal de la epidermis, se plantea un tercer experimento donde se determina la curva del coeficiente de absorción de la melanina en la epidermis a partir de los datos espectrales (Narea et al., 2015).

Tomando en consideración estos tres experimentos se plantea una serie de tres (03) experimentos más en el cual se le aplica reducción de dimensión utilizando PCA (Análisis de Componentes Principales) que permiten resumir toda la información en un par de componentes.

4.3.1. Clasificación por 31 rasgos Espectrales

La distribución de los datos espectrales por fototipo cutáneo se representa en el gráfico 4.5, donde se observa que los espectros con mayor reflectancia hacen referencia a los fototipos cutáneos I, mientras que los fototipos VI por las curvas espectrales de menor reflectancia.

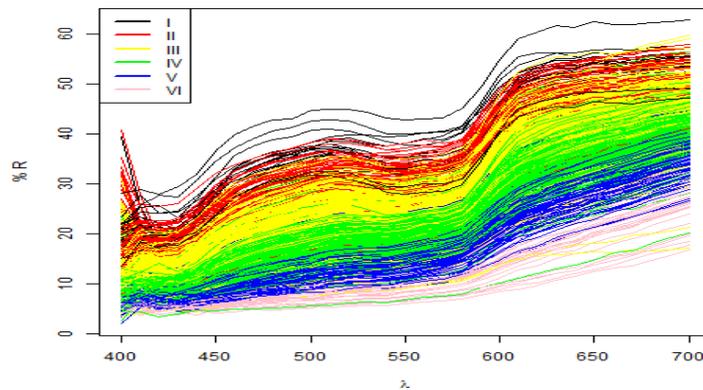


Figura 4.5: Datos Espectrales con 31 rasgos para cada Fototipo. Fuente: (Narea, 2017)

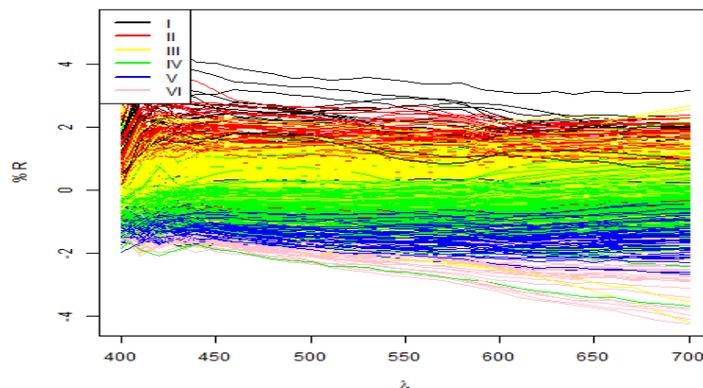


Figura 4.6: Datos Espectrales con 31 rasgos Normalizados para cada Fototipo. Fuente: (Narea, 2017)

Antes de comenzar el proceso de clasificación, es necesario normalizar los datos espectrales (ver gráfico 4.6) para evitar problemas de escalamientos.

Ahora una vez normalizados los datos se establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.1, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador. En la cual se puede observar que el mejor clasificador fue la SVM con kernel Sigmoidal con parámetros $C=100$, $\gamma=0,001$ y $\alpha=1$ con 473 vectores de soportes que obtuvo un 77,01 % de aciertos y una área ROC de 0,84, el cual indica que es un buen clasificador en cuanto a la relaciones de de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.7)

Tabla 4.1: Máquinas de Aprendizaje para 31 rasgos espectrales.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	$C=10$ $\gamma=0,03$	441	76,42	0,80	0,94
SVM-RBF	$C=100$ $\gamma=0,001$	445	75,52	0,86	0,58
SVM-Sigmoidal	$C=100$ $\gamma=0,001$ $\alpha=1$	473	77,01	0,84	0,41
RN	$N=1$ $\eta=0,0001$		75,82	0,89	0,11

Fuente: (Narea, 2017)

En la figura 4.7 se observa una especificidad del 100 % y una sensibilidad del 55 %, que indica que el sistema es capaz que discernir entre los verdaderos negativos, de este gráfico se obtuvo el área ROC que muestra en la tabla 4.1.

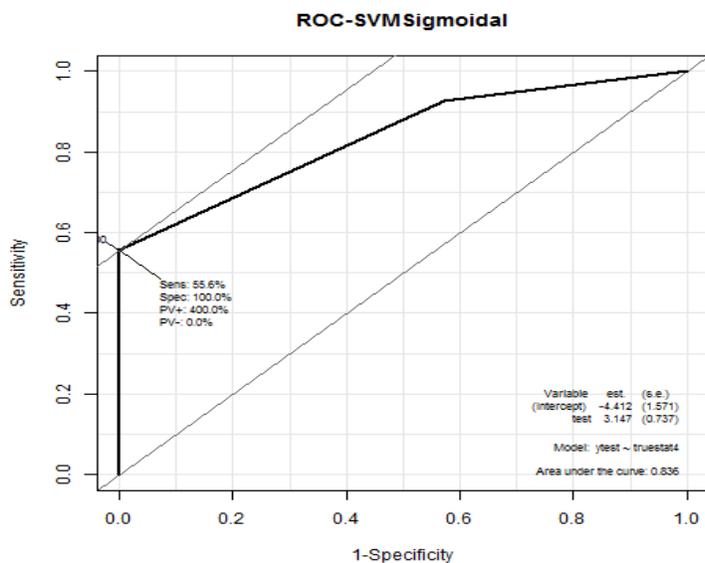


Figura 4.7: Curva ROC SVM-Sigmoidal para 31 rasgos Fuente: (Narea, 2017)

El porcentaje de acierto se determinó a partir de la matriz de confusión generada por el clasificador SVM-Sigmoidal (ver tabla 4.2).

Tabla 4.2: Matriz de confusión SVM-Sigmoidal 31 rasgos espectrales.

		Real					
		I	II	III	IV	V	VI
Asignado	I	3	2	0	0	0	0
	II	4	10	3	0	0	0
	III	0	14	97	13	1	0
	IV	0	1	17	122	13	0
	V	0	0	0	2	18	6
	VI	0	0	1	0	0	8

Fuente: (Narea, 2017)

4.3.2. Clasificación por 26 rasgos Espectrales

La distribución de los datos espectrales con 26 rasgos por fototipo cutáneo se representa en el gráfico 4.8 y en el gráfico 4.9 la distribución de datos normalizados

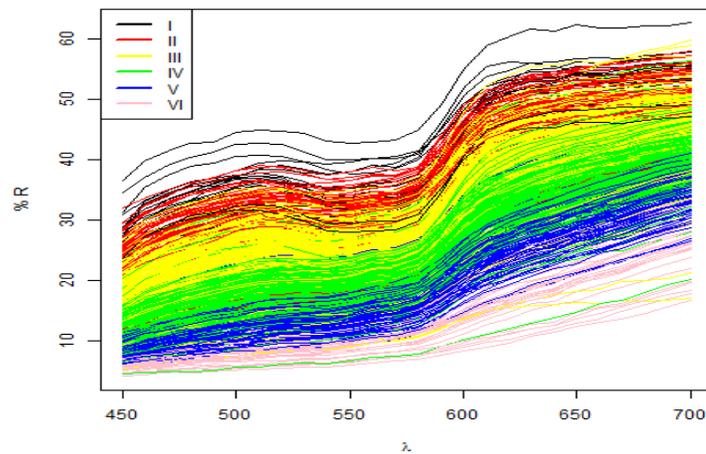


Figura 4.8: Datos Espectrales con 26 rasgos para cada Fototipo. Fuente: (Narea, 2017)

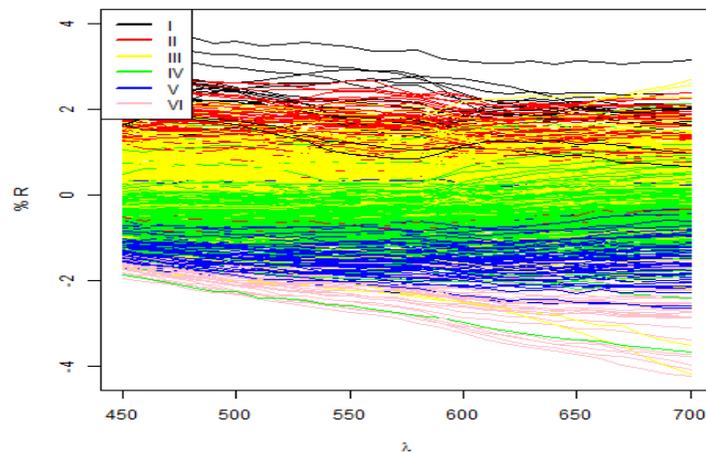


Figura 4.9: Datos Espectrales con 26 rasgos Normalizados para cada Fototipo. Fuente: (Narea, 2017)

Una vez normalizados los datos se establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.3, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador.

Tabla 4.3: Máquinas de Aprendizaje para 26 rasgos espectrales.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	C=1 $\gamma=0,04$	453	74,63	0,80	0,25
SVM-RBF	C=100 $\gamma=0,01$	407	76,12	0,84	0,41
SVM-Sigmoidal	C=1000 $\gamma=0,001$ $\alpha=0,1$	453	75,52	0,86	0,28
RN	N=1 $\eta=0,0001$		74,63	0,80	0,07

Fuente: (Narea, 2017)

Asimismo en tabla 4.3 se puede observar que el mejor clasificador fue la SVM con kernel RBF con parámetros $C=100$ y $\gamma=0,01$ con 407 vectores de soportes que obtuvo un 76,12% de aciertos y una área ROC de 0,84, el cual indica que es un buen clasificador en cuanto a la relaciones de de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.10)

En la figura 4.10 se observa una alta una especificidad del 71,4% y una sensibilidad del 92,6%, que indica que el sistema es capaz que discernir entre los verdaderos positivos; de este gráfico se obtuvo el área ROC que muestra en la tabla 4.3.

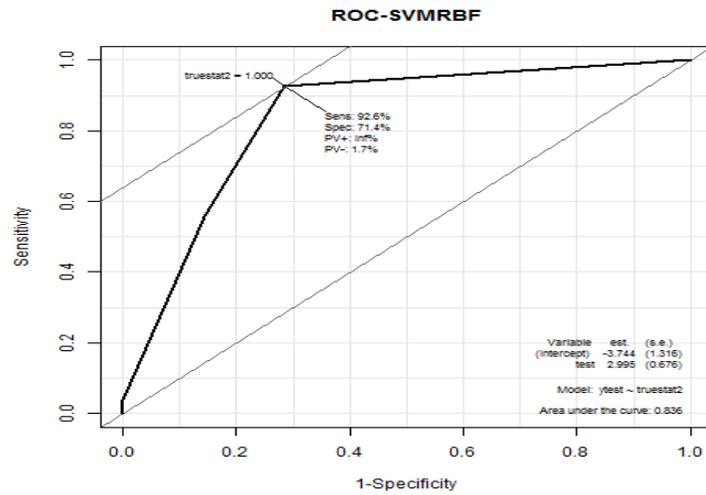


Figura 4.10: Curva ROC SVM-RBF para 26 rasgos Fuente: (Narea, 2017)

El porcentaje de acierto se determinó a partir de la matriz de confusión generada por el clasificador SVM-RBF (ver tabla 4.4).

Tabla 4.4: Matriz de confusión SVM-RBF 26 rasgos espectrales.

		Real					
		I	II	III	IV	V	VI
Asignado	I	5	2	0	0	0	0
	II	1	10	5	0	0	0
	III	1	14	96	16	1	0
	IV	0	1	17	117	11	0
	V	0	0	0	4	20	7
	VI	0	0	0	0	0	7

Fuente: (Narea, 2017)

4.3.3. Clasificación por coeficiente de absorción de la Epidermis μ_a

El coeficiente de absorción de la epidermis μ_a se determino ajustando los datos espectrales con series de Fourier, extrayendo el parametro a_0 y sutituyendo en la expresión 4.1 (Narea et al., 2015).

$$\mu_{a,epi}(a_0, \lambda) = Z e^{ka_0} 6,6 \cdot 10^{11} \lambda^{-3,3} \quad (4.1)$$

donde $Z = 0,2796$ y $k = -7,174$ son las contaste adimensionales del ajuste exponencial

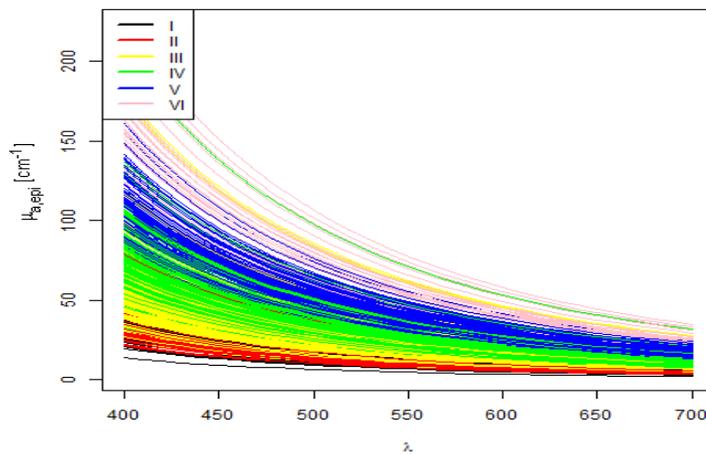


Figura 4.11: μ_a cada Fototipo. Fuente: (Narea, 2017)

La distribución de los coeficientes de absorción por fototipo cutáneo se representa en el gráfico 4.11 y en el gráfico 4.12 la distribución de datos normalizados

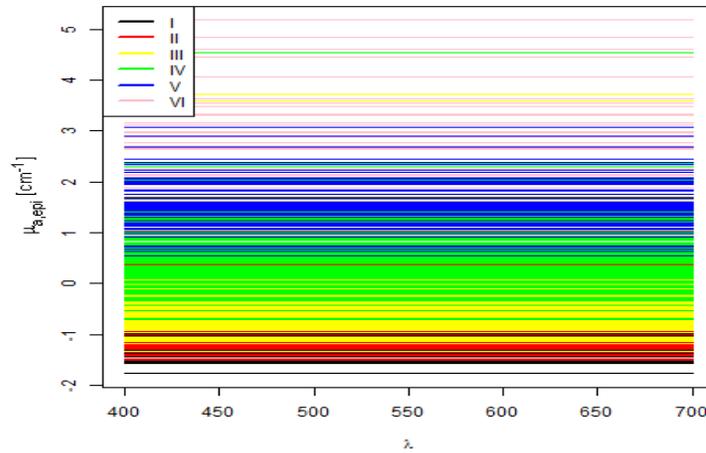


Figura 4.12: μ_a Normalizados para cada Fototipo. Fuente: (Narea, 2017)

Tabla 4.5: Máquinas de Aprendizaje para μ_a .

Máquina de Aprendizaje	Parámetros	nVS	%Acertos	Área ROC	Tiempo [s]
SVM-Lineal	C=1 $\gamma=0,03$	480	73,43	0,78	0,28
SVM-RBF	C=10 $\gamma=0,1$	432	74,33	0,80	0,52
SVM-Sigmoidal	C=100 $\gamma=0,001$ $\alpha=1$	493	74,33	0,78	0,39
RN	N=1 $\eta=0,0001$		76,12	0,84	1,61

Fuente: (Narea, 2017)

Con los datos normalizados establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.5, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador.

En la tabla 4.5 se puede observar que el mejor clasificador fue la RN con parámetros $N=1$ y $\eta=0,0001$ que obtuvo un 76,12% de aciertos y una área ROC de 0,84, el

cual indica que es un buen clasificador en cuanto a la relaciones de de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.13)

En la figura 4.13 se observa una especificidad del 100% y una sensibilidad del 55,6%, que indica que el sistema es capaz que discernir entre los verdaderos negativo por encima de los verdaderos positivos; de este gráfico se obtuvo el área ROC que se muestra en la tabla 4.5.

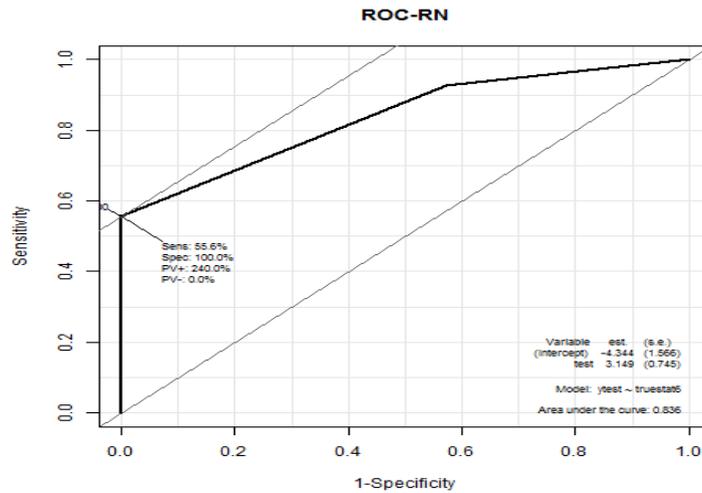


Figura 4.13: Curva ROC RN para μ_a . Fuente: (Narea, 2017)

El porcentaje de acierto se determino a partir de la matriz de confusión generada por el clasificador RN (ver tabla 4.6).

Tabla 4.6: Matriz de confusión RN para μ_a .

		Real					
		I	II	III	IV	V	VI
Asignado	I	3	2	0	0	0	0
	II	4	10	5	0	0	0
	III	0	14	96	15	1	0
	IV	0	1	16	117	10	0
	V	0	0	0	5	20	5
	VI	0	0	1	0	1	9

Fuente: (Narea, 2017)

4.3.4. Clasificación por PCA de 31 rasgos Espectrales

Realizando un análisis de PCA como se indica en el capítulo II sección 2.6 se puede establecer una reducción de dimensión de los datos espectrales resumidos en componentes, en la figura 4.14 se presenta una distribución porcentual de la información que resume las primeras cinco componentes del análisis con PCA, donde se puede apreciar que las dos primeras componentes resumen aproximadamente el 72% de la información espectral que permita un posible clasificación por fototipos cutáneos.

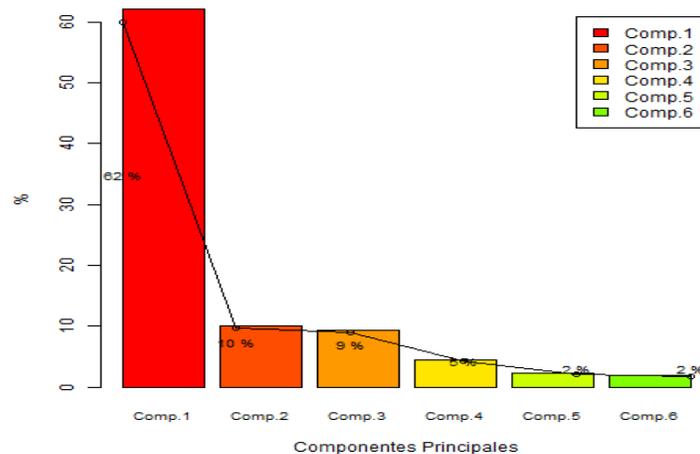


Figura 4.14: 5 Componentes Principales de los datos espectrales. Fuente: (Narea, 2017)

Cabe destacar que la primera componente guarda relación directa con la morfología de los datos espectrales, es decir contiene la información del comportamiento que describe las curvas espectrales, como se evidencia en la figura 4.15, mientras que las otras dos componentes pudieran estar asociados a elementos que influyen en los máximos de absorción evidenciados en los datos espectrales.

Ahora los datos que se usaran para generar la clasificación por cada máquina de aprendizaje, serán las dos primeras componentes del análisis de PCA.

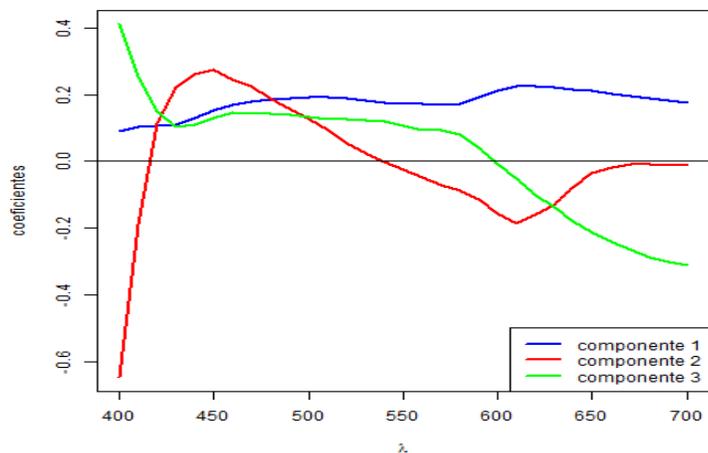


Figura 4.15: Representación de las primeras 3 componentes principales en función de las longitudes de onda. Fuente: (Narea, 2017)

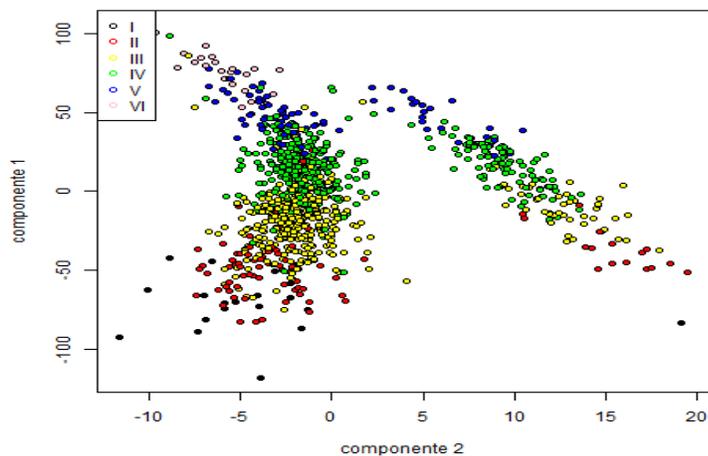


Figura 4.16: Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)

En este sentido se puede representar los fototipos con solo dos (02) componentes principales (ver figura 4.16), permitiendo reducir de $R^{31} \rightarrow R^2$ lo cual facilita una representación bidimensional de los hiperplanos de máxima separación entre clases.

Ahora se establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.7, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador.

Tabla 4.7: Máquinas de Aprendizaje para PCA de los 31 rasgos espectrales.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	C=10 $\gamma=0,05$	460	75,22	0,79	0,13
SVM-RBF	C=1 $\gamma=1$	430	74,93	0,80	0,27
SVM-Sigmoidal	C=1000 $\gamma=0,001$ $\alpha=0,001$	476	74,93	0,78	0,11
RN	N=5 $\eta=0,0001$		74,93	0,74	0,26

Fuente: (Narea, 2017)

En la tabla 4.7 el mejor clasificador presento un 75,22% de aciertos para el clasificador SVM con kernel Lineal y parámetros C=10 y $\gamma=0,5$ con 460 vectores de soportes que obtuvo una área ROC de 0,79, el cual indica que es un buen clasificador en cuanto a la relaciones de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.17)

En la figura 4.17 se observa una alta una especificidad del 85,7% y una sensibilidad del 59,3%, que indica que el sistema es capaz que discernir entre los verdaderos negativos; de este gráfico se obtuvo el área ROC que muestra en la tabla 4.7.

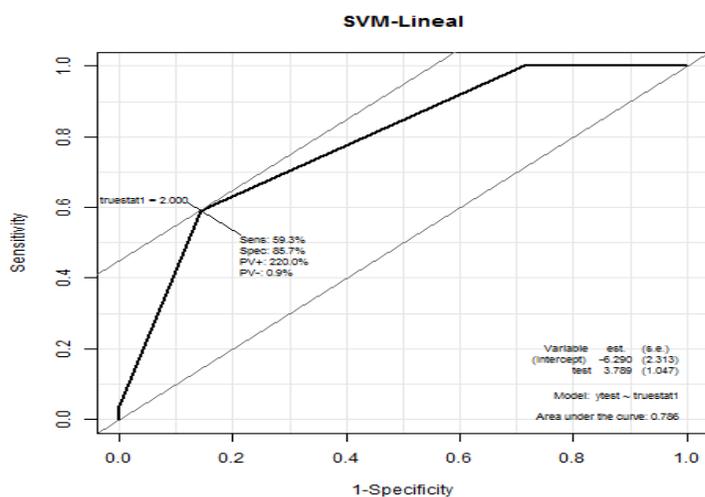


Figura 4.17: Curva ROC SVM-Linear para PCA de 31 rasgos. Fuente: (Narea, 2017)

El porcentaje de acierto se determinó a partir de la matriz de confusión generada por el clasificador SVM-Linear (ver tabla 4.8).

Tabla 4.8: Matriz de confusión SVM-Linear para PCA de 31 espectrales.

		Real					
		I	II	III	IV	V	VI
Asignado	I	2	0	0	0	0	0
	II	4	11	5	0	0	0
	III	1	15	96	17	1	0
	IV	0	1	16	118	14	0
	V	0	0	0	2	17	6
	VI	0	0	1	0	0	8

Fuente: (Narea, 2017)

Dado que la clasificación se realizó para una estructura de datos de dimensión dos (02), se puede realizar una representación bidimensional de la SVM (ver figura 4.18, indicando con una equis (X) el lugar donde son clasificados los datos que se

utilizaron para probar el clasificador y los distintos tonos de gris indican la región que ocupan los hiperplanos de separación entre las clases.

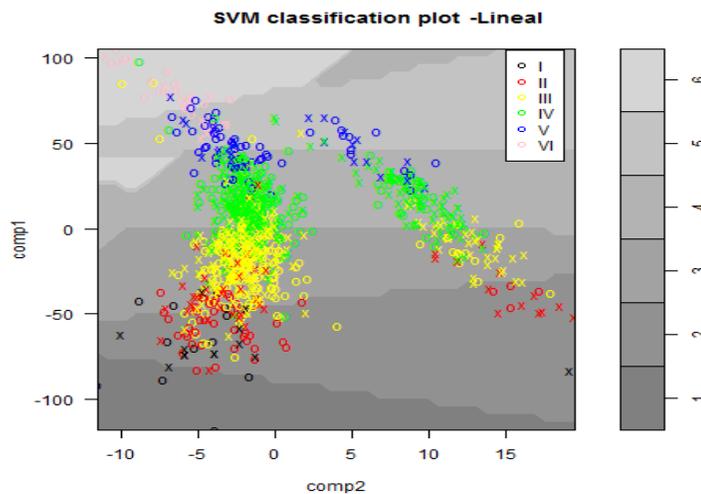


Figura 4.18: SVM-Lineal para PCA de 31 rasgos Fuente: (Narea, 2017)

4.3.5. Clasificación por PCA de 26 rasgos Espectrales

Al igual que la sección anterior se establece una reducción de dimensión de los 26 rasgos de los datos espectrales resumidos en componentes, en la figura ?? se presenta una distribución porcentual de la información que resume las primeras seis (06) componentes del análisis con PCA, donde las dos primeras componentes resumen aproximadamente el 81 % de la información espectral que permite una clasificación por fototipos cutáneos.

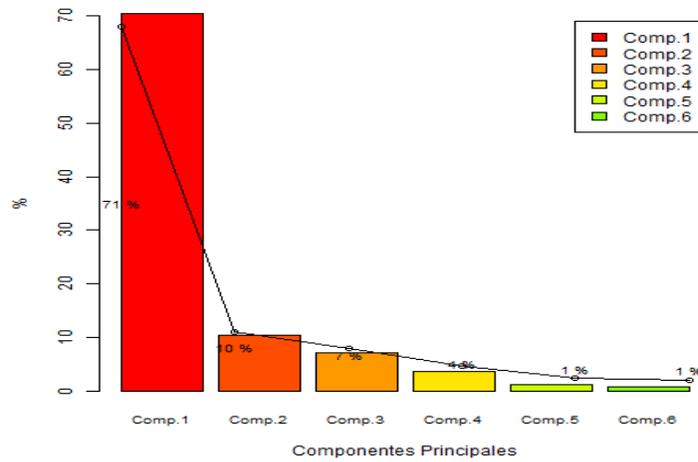


Figura 4.19: 6 Componentes Principales de 26 rasgos espectrales. Fuente: (Narea, 2017)

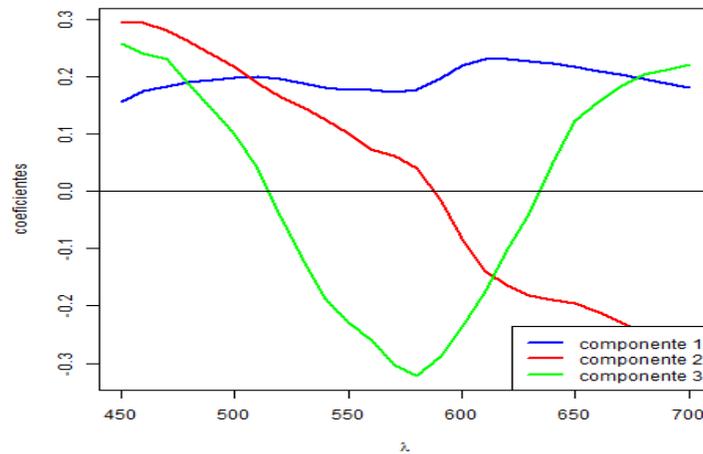


Figura 4.20: Representación de las primeras 3 componentes principales en función de las longitudes de onda para 26 rasgos espectrales. Fuente: (Narea, 2017)

Se observa en la figura 4.20 que la primera componente principal, posee el mismo comportamiento que la primera componente principal de los 31 rasgos espectrales, confirmando de esta manera que existe una relación directa con la morfología de

los datos espectrales que definen la curva espectral de cada fototipo.

Ahora los datos que se usaran para generar la clasificación para cada máquina de aprendizaje, serán las dos primeras componentes del análisis de PCA.

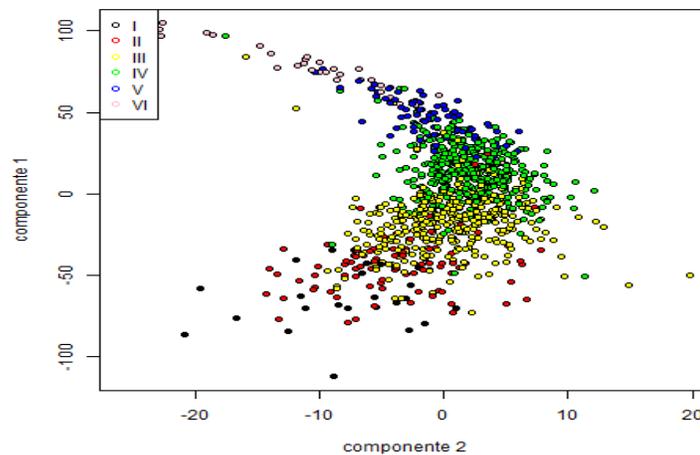


Figura 4.21: Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)

En consecuencia se representan los fototipos con solo dos (02) componentes principales (ver figura 4.21), permitiendo reducir de $R^{26} \rightarrow R^2$ lo cual facilita una representación bidimensional de los hiperplanos de máxima separación entre clases.

Ahora se establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.9, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador.

Tabla 4.9: Máquinas de Aprendizaje para PCA de los 26 rasgos espectrales.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	C=10 $\gamma=0,5$	466	74,33	0,82	0,12
SVM-RBF	C=1000 $\gamma=1$	474	75,52	0,84	0,30
SVM-Sigmoidal	C=100 $\gamma=0,01$ $\alpha=0,1$	482	75,52	0,84	0,11
RN	N=1 $\eta=0$		74,03	0,87	0,10

Fuente: (Narea, 2017)

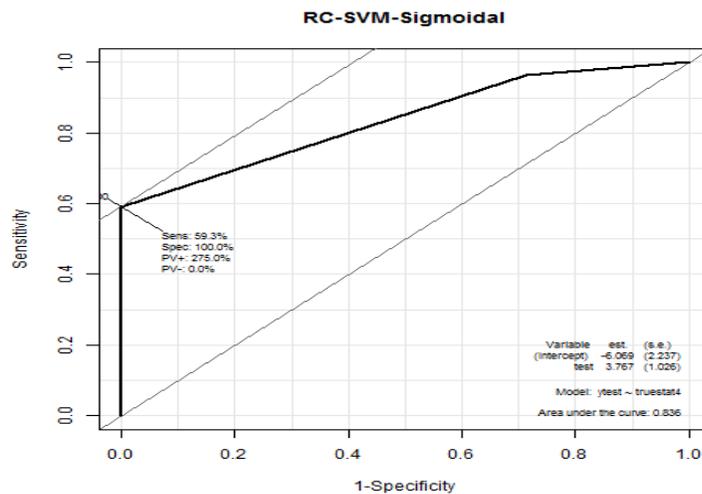


Figura 4.22: Curva ROC SVM-Sigmoidal para PCA de 26 rasgos Fuente: (Narea, 2017)

De la tabla 4.9 se visualizan que SVM-RBF y SVM-Sigmoidal obtienen el mismo porcentaje de aciertos, siendo este del 75,52%, así bien el criterio que permite decidir por el mejor entre los dos es el rendimiento en el entrenamiento de cada uno, donde el tiempo de entrenamiento para el SVM-Sigmoidal fue de 0,11s, mientras que las

SVM-RBF fue de 0,30s.

Los parámetros óptimos de entrenamiento fueron: $C=100$, $\gamma=0,01$ y $\alpha=0,1$ con 482 vectores de soportes y un área ROC de 0,84 el cual indica que es un buen clasificador en cuanto a la relaciones de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.22)

En la figura 4.22 se observa una alta una especificidad del 100% y una sensibilidad del 59,3%, que indica que el sistema es capaz que discernir entre los verdaderos negativos; de este gráfico se obtuvo el área ROC que muestra en la tabla 4.9. El porcentaje de acierto se determino a partir de la matriz de confusión generada por el clasificador SVM-Sigmaoidal (ver tabla 4.10).

Tabla 4.10: Matriz de confusión SVM-Sigmaoidal para PCA de 26 rasgos espectrales.

		Real					
		I	II	III	IV	V	VI
Asignado	I	2	1	0	0	0	0
	II	5	10	5	0	0	0
	III	0	15	96	15	1	0
	IV	0	1	16	120	13	0
	V	0	0	0	2	18	7
	VI	0	0	1	0	0	7

Fuente: (Narea, 2017)

En la figura 4.23 se indica con una equis (X) el lugar donde son clasificados los datos que se utilizaron para probar el clasificador y los distintos tonos de gris indican la región que ocupan los hiperplanos de separación entre las clases.

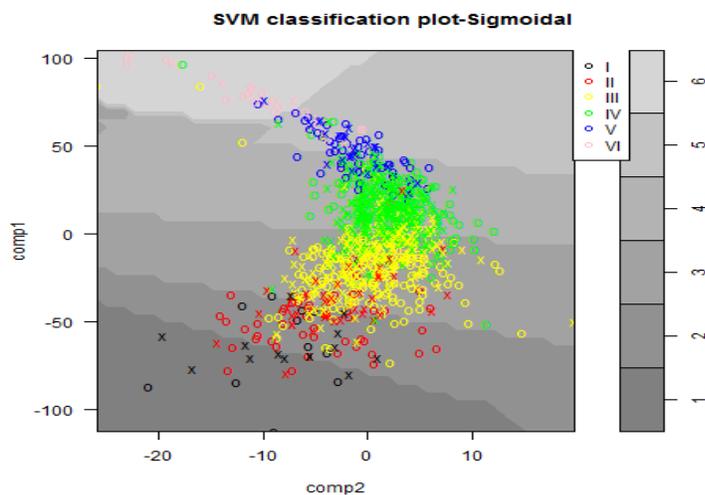


Figura 4.23: SVM-Sigmaoidal para PCA de 26 rasgos Fuente: (Narea, 2017)

4.3.6. Clasificación por PCA del coeficiente de absorción de la Epidermis μ_a

Se establece una reducción de dimensión de los μ_a de la epidermis resumidos en componentes principales, en la figura 4.24 se presenta una distribución porcentual de la información que resume las primeras seis (06) componentes del análisis con PCA, donde solo la primera componente resume aproximadamente el 10% de la información espectral de la absorción que permite una clasificación por fototipos cutáneos.

En la figura 4.25 la primera componente principal describe con exactitud el comportamiento del coeficiente de absorción de la melanina en la epidermis.

Ahora los datos que se usaran para generar la clasificación para cada máquina de aprendizaje, serán las dos primeras componentes del análisis de PCA.

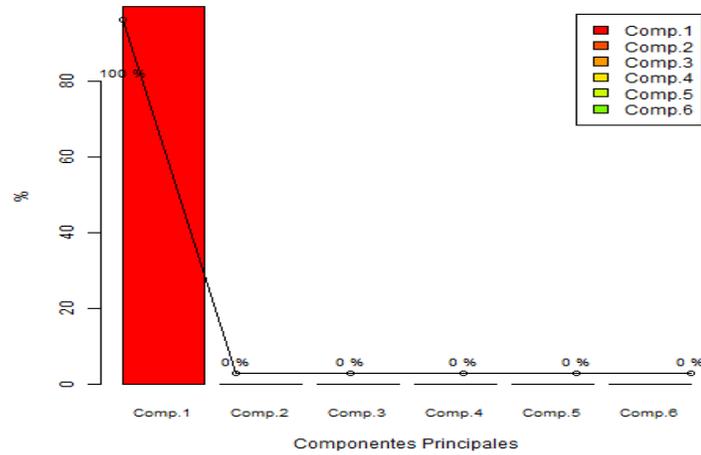


Figura 4.24: 6 Componentes Principales de μ_a de la epidermis. Fuente: (Narea, 2017)

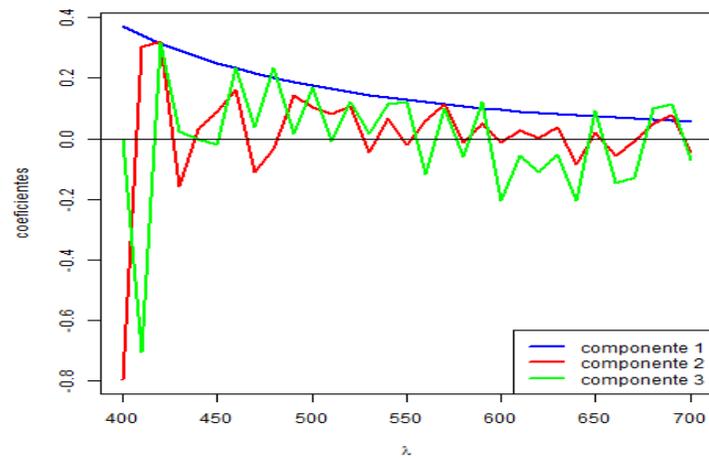


Figura 4.25: Representación de las primeras 3 componentes principales en función de las longitudes de onda para μ_a de la epidermis. Fuente: (Narea, 2017)

Como las componentes principales permiten una reducción de dimensión de $R^{31} \rightarrow R^2$, se representan los fototipos con solo dos (02) componentes principales (ver figura 4.21), lo cual facilita una representación bidimensional de los hiperplanos de máxima separación entre clases.

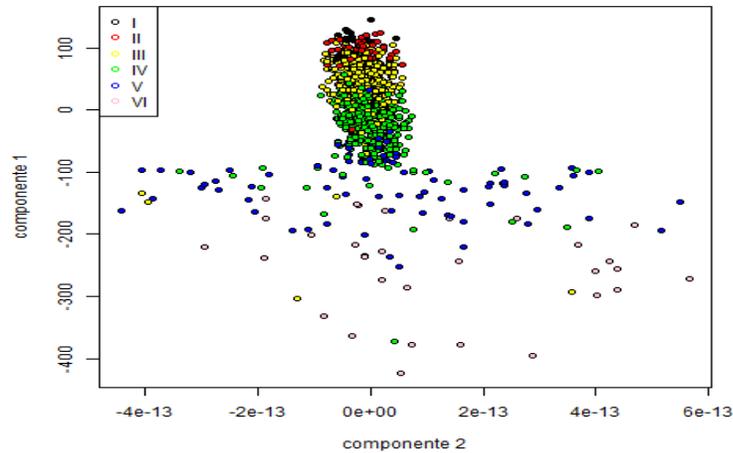


Figura 4.26: Componentes principales 1 y 2 para cada fototipo cutáneo. Fuente: (Narea, 2017)

Ahora se establecen los parámetros óptimos de clasificación para cada máquina de aprendizaje utilizando correlación cruzada, reportados en la tabla 4.11, junto con el porcentaje de aciertos, el área de la curva ROC y el tiempo de entrenamiento de cada clasificador.

Tabla 4.11: Máquinas de Aprendizaje para PCA de μ_a de la epidermis.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	C=10 $\gamma=0,5$	481	74,03	0,73	0,12
SVM-RBF	C=100 $\gamma=0,01$	467	74,33	0,78	0,28
SVM-Sigmoidal	C=1000 $\gamma=0,001$ $\alpha=0,001$	497	74,33	0,75	0,13
RN	N=3 $\eta=0,0001$		75,82	0,80	0,53

Fuente: (Narea, 2017)

De la tabla 4.11 el 75,82 % de los aciertos se logran utilizando una RN de una capa oculta con tres neuronas ($N=3$) y velocidad de aprendizaje de $\eta=0,0001$, con un área ROC de 0,80 el cual indica que es un buen clasificador en cuanto a la relaciones de sensibilidad y especificidad del sistema de clasificación. (ver figura 4.27)

En la figura 4.27 se observa una alta especificidad del 100 % y una sensibilidad del 59,3%, que establecen que el sistema es capaz que discernir entre los verdaderos negativos; de este gráfico se obtuvo el área ROC que muestra en la tabla 4.11.

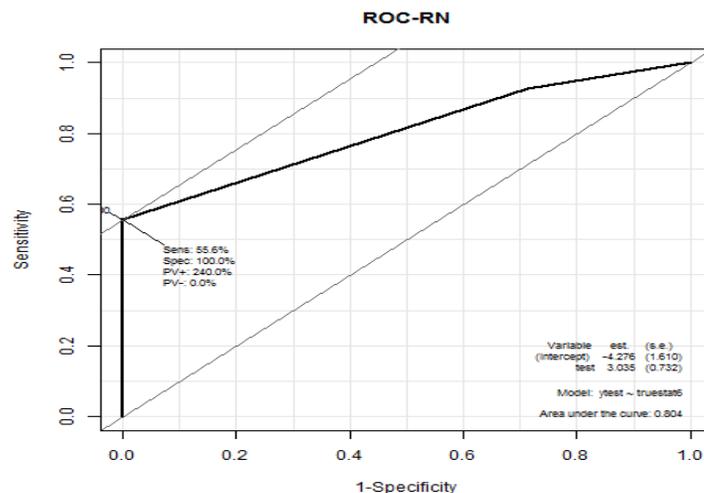


Figura 4.27: Curva ROC RN para PCA de μ_a de la epidermis Fuente: (Narea, 2017)

El porcentaje de acierto se determino a partir de la matriz de confusión generada por el clasificador SVM-Sigmaoidal (ver tabla 4.12).

Tabla 4.12: Matriz de confusión RN para PCA μ_a de la epidermis.

		Real					
		I	II	III	IV	V	VI
Asignado	I	2	2	0	0	0	0
	II	5	10	5	0	0	0
	III	0	14	95	15	1	0
	IV	0	1	17	117	9	0
	V	0	0	0	5	22	6
	VI	0	0	1	0	0	8

Fuente: (Narea, 2017)

4.3.7. Clasificación por conglomerados

Las máquinas de aprendizajes utilizadas para las distintas estructuras de datos permitieron obtener porcentajes de aciertos que van desde 75,22% a 77,01% para clasificar los fototipos cutáneos de acuerdo con la escala de Fitzpatrick, como se evidencia en los resultados de las secciones anteriores. Asimismo al utilizar los 31 rasgos espectrales se logro el mayor porcentaje de aciertos con una SVM-Sigmaoidal.

Es evidente que los porcentajes se mantuvieron cercanos unos de otros sin importar la forma de la estructura de los datos que se utilizaron para entrenar las máquinas de aprendizaje, esto se debe principalmente a que el proceso de entrenamiento fue supervisado por el criterio de médicos dermatólogos que insisten en clasificar el fototipo del venezolano en una escala establecida para etnias donde el mestizaje no es tomado en cuenta.

De acuerdo a esto si se desea eliminar el sesgo en la clasificación, se realizo una pre clasificación de los datos espectrales bajo un criterio no supervisado utilizando el reagrupamiento por conglomerados bajo el criterio de k-media, donde el único parámetro establecido es el numero de clases, que es igual a 6 numero de clases contemplada por Fitzpatrick. Esto permite generar una redistribución de las clases (ver figura 4.28).

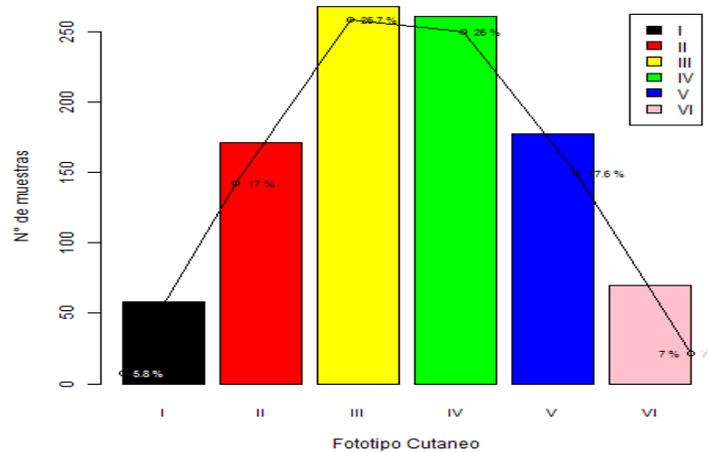


Figura 4.28: Distribución de datos por Fototipo cutáneo utilizando conglomerados.
Fuente: (Narea, 2017)

En la figura 4.28 se pone en manifiesto que la redistribución de los datos se da hacia los fototipos centrales (III,IV y V), indicando la tendencia de los fototipos del venezolano.

Tabla 4.13: Matriz de confusión conglomerados.

		Real					
		I	II	III	IV	V	VI
Asignado	I	18	30	10	0	0	0
	II	7	51	109	4	0	0
	III	0	12	196	59	1	0
	IV	0	1	46	212	1	1
	V	0	1	3	120	53	0
	VI	0	0	5	5	29	31

Fuente: (Narea, 2017)

Tomando los datos supervisado y los no supervisados se construye una matriz de confusión (ver tabla 4.13) del conglomerado, encontrando solo un 55,82% de aciertos

respectos al criterio médico, lo cual indica que 44,18% de las no coincidencias se deben principalmente a sesgos por parte del dermatólogo al confundir color de piel con fototipo cutáneo y por lo tanto establecer una clasificación errónea o forzada dentro de la escala de Fitzpatrick. Este sesgo se produce por factores de iluminación a la hora de evaluar un paciente, así como a la percepción óptica entre distintos especialistas.

Una premisa importante para validar el método por conglomerado es que las personas con bajo fototipo cutáneo (I) poseen una alta reflexión en el espectro visible, mientras que los de fototipo cutáneo alto (VI) tienen curvas espectrales de baja reflectancia; esta relación se mantiene para los fototipos intermedios (II-V). De esta forma al calcular los conglomerados por el algoritmo de k-medias, se puede establecer un orden decreciente en las distancias euclidianas entre los centroides de cada conglomerado, permitiendo así realizar una reclasificación de los fototipos cutáneos a partir de estas distancias (ver figura 4.29).

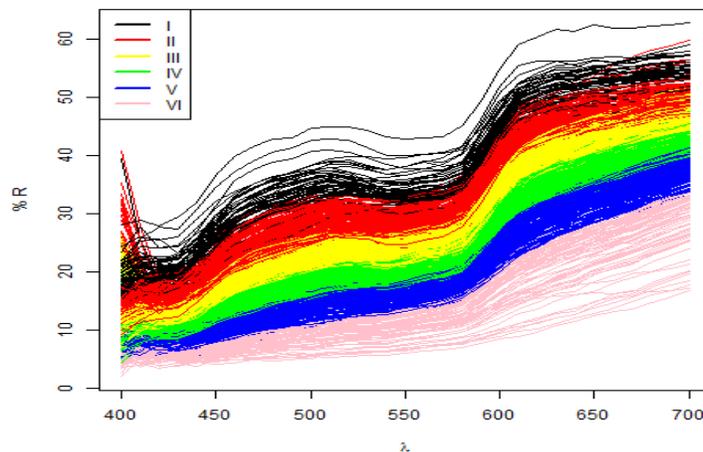


Figura 4.29: Espectro de Reflectancia Difusa distribuido por conglomerados para cada fototipo cutáneo. Fuente: (Narea, 2017)

Tomando los datos espectrales por conglomerados de la figura 4.29, y realizando un PCA, se puede determinar las componentes principales, donde de acuerdo a los expuesto en las en los experimentos de las secciones anteriores, las dos (02) primeras

componentes resumen la mayor parte de la información espectral para poder realizar una clasificación de la piel.

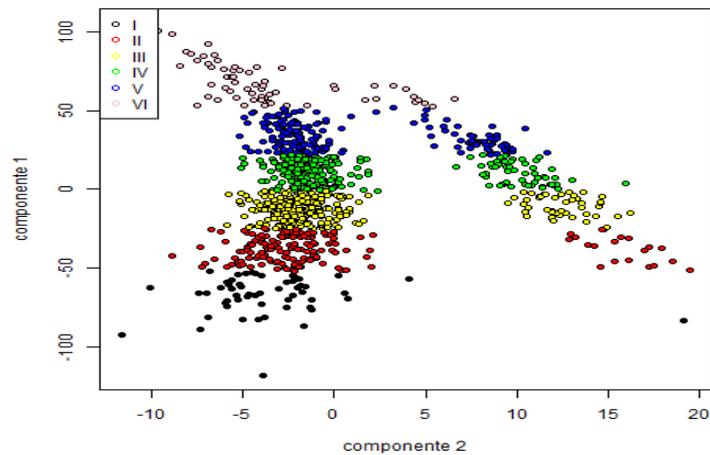


Figura 4.30: Componentes principales 1 y 2 para cada fototipo cutáneo . Fuente: (Narea, 2017)

Tabla 4.14: Máquinas de Aprendizaje para PCA-Conglomerados.

Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
SVM-Lineal	$C=1000$ $\gamma=0,5$	33	98,81	1,00	0,05
SVM-RBF	$C=1000$ $\gamma=0,01$	107	98,21	1,00	0,08
SVM-Sigmoidal	$C=1000$ $\gamma=0,01$ $\alpha=0,001$	135	97,61	1,00	0,05
RN	$N=1$ $\eta=0$		99,40	0,97	1,00

Fuente: (Narea, 2017)

En la figura 4.30 se observa una distribución bien diferenciada entre las clases (fototipos), permitiendo así establecer clasificadores más eficiente. Una vez realizada

esta preclasificación, se implementaron los algoritmos de aprendizajes de SVM y RN, y los resultados que se pueden observar en la tabla 4.14.

En la tabla 4.14 se evidencia la eficiencia para máquina de aprendizaje, ya que los porcentajes de aciertos se encuentran entre 97,61 % y 99,40 %, con áreas ROC iguales a 1, lo cual indica que todos son buenos clasificadores en cuanto a la relaciones de de sensibilidad y especificidad del sistema de clasificación. (ver figuras 4.31(a),4.31(b), 4.31(c) y 4.31(d)) y tiempos de entrenamientos $t \leq 1s$.

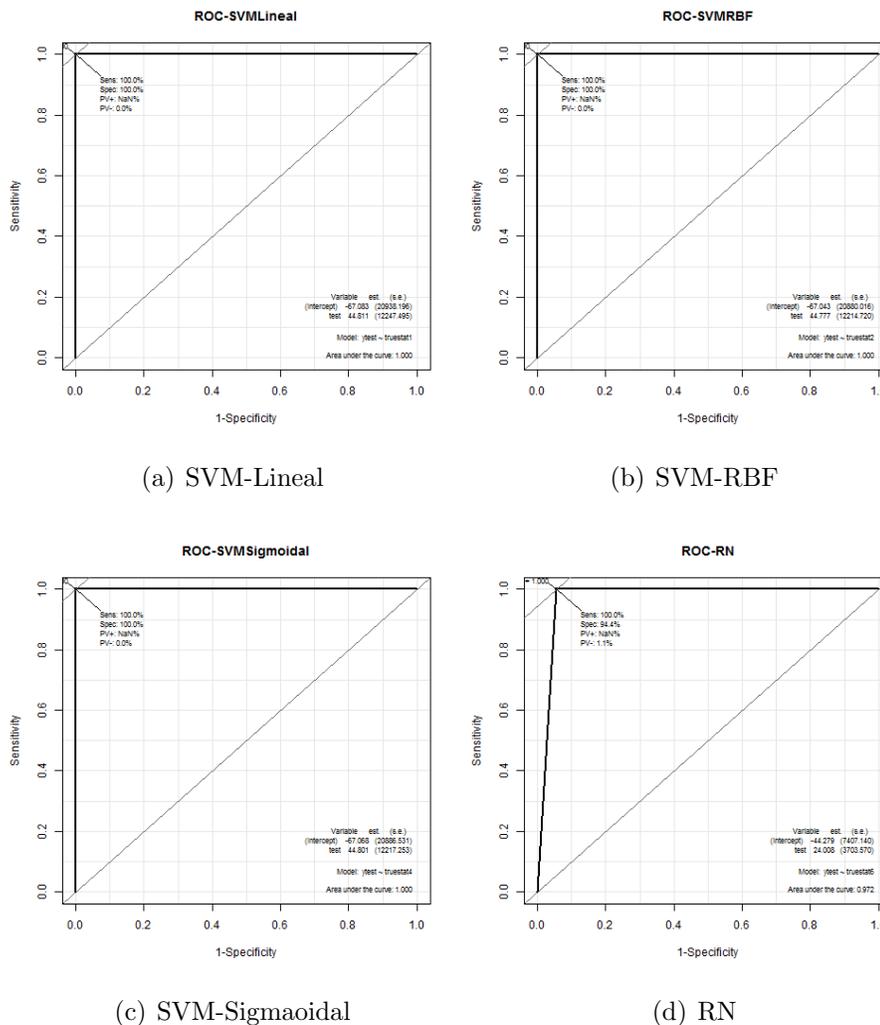


Figura 4.31: Curva ROC para PCA de Conglomerados. Fuente: (Narea, 2017)

Como los porcentajes de aciertos son altos y el área ROC son superiores a 0,90, se toma como criterio para seleccionar al mejor clasificador el tiempo de entrenamiento, lo cual se descarta las RN, dado a que el tiempo de entrenamiento fue de $t = 1s$, dejando para decidir solo entre las SVM y tomando la que posee en menor numero de vectores de soporte ($nV = 33$), se toma como el mejor clasificador las SVM con kernel Lineal (SVM-Lineal), con parámetros $C=1000$ y $\gamma = 0,5$. Estos datos se representa en la figura 4.32 donde se indica con una equis (X) el lugar donde son clasificados los datos que se utilizaron para probar el clasificador y los distintos tonos de gris indican la región que ocupan los hiperplanos de separación entre las clases.

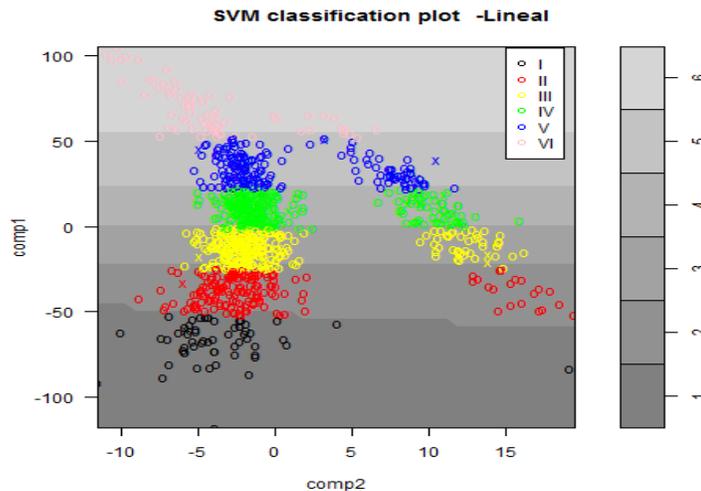


Figura 4.32: SVM-Lineal para PCA de conglomerados . Fuente: (Narea, 2017)

4.4. Comparación entre los métodos de clasificación

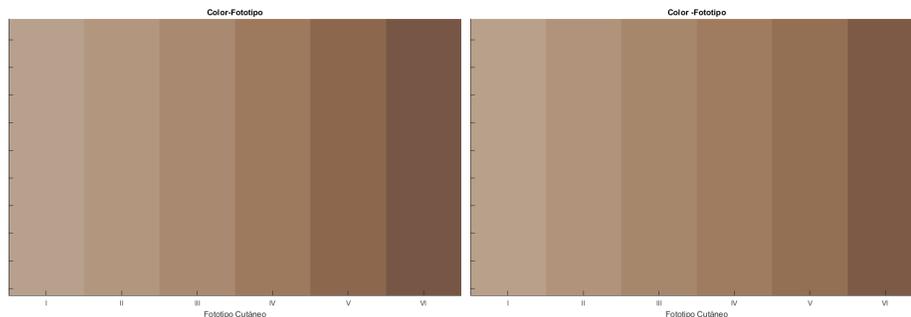
En la tabla 4.15 se presenta una síntesis con los resultados de las máquinas de aprendizaje que poseen mejor desempeño en cuanto al porcentaje de aciertos para clasificar los fototipos cutáneos. En este sentido el mejor clasificador para un sistema de entrenamiento supervisado (criterio médico) se logró con una SVM-Sigmoidal para la estructura de datos con los 31 rasgos espectrales, logrando un 77,01 % de aciertos.

Ahora para un entrenamiento no supervisado utilizando conglomerados y PCA se alcanza un porcentaje de clasificación de 98,81 % empleando una SVM-Lineal.

Tabla 4.15: Comparación entre las Máquinas de Aprendizaje de cada experimento.

Datos	Máquina de Aprendizaje	Parámetros	nVS	%Aciertos	Área ROC	Tiempo [s]
31 rasgos	SVM-Sigmoidal	C=100 $\gamma=0,001$ $\alpha=1$	473	77,01	0,84	0,41
26 rasgos	SVM-RBF	C=100 $\gamma=0,01$	407	76,12	0,84	0,41
μ_a	RN	N=1 $\eta=0,0001$		76,12	0,84	1,61
PCA-31 rasgos	SVM-Lineal	C=10 $\gamma=0,05$	460	75,22	0,79	0,13
PCA-26 rasgos	SVM-Sigmoidal	C=100 $\gamma=0,01$ $\alpha=0,1$	482	75,52	0,84	0,11
PCA- μ_a	RN	N=3 $\eta=0,0001$		75,82	0,80	0,53
PCA-Conglomerados	SVM-Lineal	C=1000 $\gamma=0,5$	33	98,81	1,00	0,05

Fuente: (Narea, 2017)



(a) Criterio médico

(b) Conglomerados

Figura 4.33: Color del fototipo cutáneo empleando entrenamiento supervisado y no supervisado. Fuente: (Narea, 2017)

Utilizando el criterio de entrenamiento supervisado y no supervisado se realizó una escala del fototipo cutáneo en términos del color característico de la piel del

Venezolano (ver figura 4.33). En la figura 4.33(a) se evidencia los distintos tonos del color de piel de acuerdo al criterio del médico dermatólogo y en la figura 4.33(b) se establece una distribución del color del fototipo a partir de la pre-clasificación por conglomerados.

Existe una diferencia del color entre los fototipos que se presentan en las figuras 4.33(a) y 4.33(b), y se determinó utilizando la ecuación 4.2 establecida por la Comisión Internacional de Iluminación CIE76 por sus siglas en inglés (International Commission on Illumination) que establece para el espacio del color CieLab1965 (Westland et al., 2012) como diferencia apenas perceptible que se produce entre dos niveles de intensidad de un estímulo sensorial cuando $\Delta E^* < 3$.

$$\Delta E^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \quad (4.2)$$

Las diferencias del color se presentan en la tabla 4.16, donde se observa una disparidad colorimétrica poco perceptible para el ojo humano, la cual hace referencia al 44,18% de los datos sesgado.

Tabla 4.16: Diferencia del color para fototipos cutáneos por criterio médico y conglomerados

		Criterio médico			Conglomerados			ΔE^*
		L_1^*	a_1^*	b_1^*	L_2^*	a_2^*	b_2^*	
Fototipo	I	67,31	5,54	13,64	67,36	5,79	14,59	0,99
	II	64,07	6,56	16,53	62,95	7,04	17,08	1,34
	III	59,57	8,09	18,36	58,73	8,40	19,04	1,13
	IV	53,74	9,78	20,30	54,53	9,70	20,33	0,79
	V	46,81	11,18	19,96	49,69	10,76	20,83	3,04
	VI	39,67	11,00	15,54	41,33	11,15	17,32	2,45

Fuente: (Narea, 2017)

CAPÍTULO V

CONCLUSIONES

Después de realizar el análisis de los datos y de aplicar las distintas máquinas de aprendizaje para clasificar los fototipos cutáneos se concluye:

Los espectros de reflexión difusa de la piel sana, en conjunto con los sistemas clasificadores por reconocimientos de patrones bajos criterios supervisados y no supervisados son una combinación valiosa que permitiría a los médicos dermatólogos aumentar la precisión en la evaluación de los fototipos cutáneos, principalmente en Venezuela, donde existe una diversidad de pieles producto del mestizaje que no se logran clasificar por los criterios de Fitzpatrick establecidos en 1975.

En cuanto a los resultados obtenidos, una vez seleccionados los principales rasgos de los espectros de reflexión difusa de los distintos fototipos cutáneos, utilizando técnicas de reducción de dimensión e implementando los modelos de redes neuronales artificiales y máquinas de soporte vectorial para los distintos experimentos planteados se obtuvo que, utilizando los 31 rasgos espectrales, es decir, toda la información espectral bajo un entrenamiento supervisado, la mejor máquina de aprendizaje fue la SVM de kernel Sigmoidal con los parámetros $C=100$, $\gamma=0,001$ y $\alpha=1$ que obtuvo un 77,01% de aciertos y una área ROC de 0,84 indicando en cuanto a la relaciones de sensibilidad y especificidad que es un sistema de clasificación confiable. De forma general todos los algoritmos implementados tienen un buen desempeño.

Sin embargo, cuando se evalúa la eficiencia en términos del tiempo de entrenamiento, el mejor clasificador se alcanzó cuando se le aplicó análisis de componentes principales a los datos con 26 rasgos espectrales, utilizando una SVM de kernel Sigmoïdal con parámetros $C=100$, $\gamma=0,01$ y $\alpha=0,1$, se obtuvo un 75,52 % de aciertos y un área ROC de 0,84, empleando $\frac{1}{4}$ del tiempo de entrenamiento que el utilizado con los 31 rasgos espectrales, teniendo así un coste de entrenamiento más bajo. En este sentido, el desempeño de ambas SVM son comparables en términos del número de aciertos y los costos de entrenamientos.

Estos porcentajes de aciertos se deben principalmente al sesgo que poseen los dermatólogos al intentar ubicar a los pacientes venezolanos en la escala de Fitzpatrick, basándose solo en el color de la piel, llegando a confundir este concepto con el del fototipo, esto se debe principalmente a una disparidad colorimétrica poco perceptible para el ojo humano, producto del mestizaje entre las etnias. Lo cual se evidencia, cuando se realiza un entrenamiento no supervisado, donde el 44,18 % de los datos espectrales clasificados por criterio médico no coincide con los clasificados por conglomerados, lo cual indica que se debe realizar una redistribución de los fototipos cutáneos del venezolano dentro de la escala de Fitzpatrick.

De esta nueva distribución de los datos se tiene que la mejor máquina de aprendizaje en término de costos computacionales, porcentajes de aciertos y relaciones de sensibilidad-especificidad se logró utilizando una SVM con kernel Lineal (SVM-Lineal) de parámetros $C=1000$ y $\gamma = 0,5$ alcanzando un 98,81 % de aciertos y un área ROC de 1; empleando un mínimo 33 vectores de soporte para generar los hiperplanos de separación en comparación a las otras SVM implementadas.

RECOMENDACIONES

Se recomienda para investigaciones futuras extender la experimentación usando un mayor número de espectros de entrenamiento validados a juicios de expertos con el objetivo de obtener mayores porcentajes de acierto para los sistemas supervisados.

Se propone controlar las variables de temperatura e iluminación en los espacios donde se realiza la recolección de los datos con el fin de estandarizar la evaluación médica disminuyendo la subjetividad, para aumentar la confiabilidad del diagnóstico.

Se sugiere realizar investigaciones sobre una escala para los fototipos cutáneos del venezolano, ya que biológicamente existe una mezcla que redefine la escala de Fitzpatrick para el hispano hablante, producto del mestizaje entre las etnias.

APÉNDICE A

Código sin reducción de Dimensión

```
1  rm(list=ls()) ## limpia las variables
2  ## Se cargan las Librerias ##
3  library(xlsx) # leer datos xlsx #
4  library(e1071) # SVM, basado en LIBSVM
5  library(Epi) # curvas ROC
6  library(caret) #entrenamiento de modelos de clasificaci n
7  library(nnet) # Red Neuronal (percentron multicapa)
8  #Leer datos Excel
9  file="./DatosEspectros.xlsx" ## Datos espectrales para 31, 26
   ragos y coeficiente de absorcion
10 res1 <- t(read.xlsx(file, 1,endRow=32))[-1,] # declaraci n de
   variables
11 res2 <- t(read.xlsx(file, 2,endRow=32))[-1,]
12 res3 <- t(read.xlsx(file, 3,endRow=32))[-1,]
13 res4 <- t(read.xlsx(file, 4,endRow=32))[-1,]
14
15 res <- rbind(res1,res2,res3,res4) #matriz de datos espectrales
16 indice<-as.logical(apply(!is.na(res),1,prod))
17 res <- res[indice,-(1:5)]
18 res<-scale(res) # normalizaci n de los datos
19 res<-as.data.frame(res)
20
```

```
21 #definición de la etiqueta de los datos (clases)
22 fototipo<-t(cbind(
23   as.vector(read.xlsx(file,1,startRow=35,endRow=36))[-1]
24 ,as.vector(read.xlsx(file,2,startRow=35,endRow=36))[-1]
25 ,as.vector(read.xlsx(file,3,startRow=35,endRow=36))[-1]
26 ,as.vector(read.xlsx(file,4,startRow=35,endRow=36))[-1]
27   ) )
28 rownames(fototipo )<-NULL
29 fototipo[fototipo=="I", ]<-1
30 fototipo[fototipo=="II", ]<-2
31 fototipo[fototipo=="III", ]<-3
32 fototipo[fototipo=="IV", ]<-4
33 fototipo[fototipo=="V", ]<-5
34 fototipo[fototipo=="VI", ]<-6
35
36 colores<-c("black","red","yellow","green","blue","pink") [
   unclass(as.factor(fototipo))] # se fijan los colores
   asociados para cada clase
37 png("./imagenes/Nombre de la imagen.png") # guarda imagen
38 matplot(seq(450,700,by=10),t(res),type="l",col=colores,xlab =
   expression(lambda),ylab="% R",lty=1) #grafica de los datos
   espectrales
39 legend("topleft", legend = c("I","II","III","IV","V","VI"),
   col =c("black","red","yellow","green","blue","pink"),lwd=2)
40 dev.off()
41
42 ## SVM
43 index <- 1:nrow(res)
44 lambda<-seq(400,700,by=10)
45 testindex <- sample(index, trunc(length(index)/3))
46 testset <- res[testindex,]
47 trainset <- res[-testindex,]
48 fototipo1<-fototipo[-testindex]
49 resTrain<-data.frame(res[-testindex,],fototipo1) #datos de
   entrenamiento
```

```
50 colnames(resTrain)<-c(lambda,"fototipo")
51 fototipo2<-fototipo[testindex]
52 resTest<-data.frame(res[testindex,],fototipo2) #datos de
   prueba
53 colnames(resTest)<-c(lambda,"fototipo")
54
55 ## Parametros de las SVM
56     gamma1<-c(0.001,0.01,0.1,1,1.5,2,2.5)
57     c1<-c(0.001,0.01,0.1,1,10,100,1000)
58     r1<- c(0.001,0.01,0.1,1,1.5,2,2.5)
59     d<-seq(1,5,by=1)
60 ### Sintonizaci n de parametros de las SVM
61 tiempo1<-system.time(obj1 <- tune.svm(fototipo ~ ., data =
   resTrain, cost = c1, kernel = "linear"))
62 tiempo2<-system.time(obj2 <- tune.svm(fototipo ~ ., data =
   resTrain, gamma = gamma1, cost = c1, kernel = "radial"))
63 tiempo4<-system.time(obj4 <- tune.svm(fototipo ~ ., data =
   resTrain, coef0=r1,gamma=gamma1, cost =c1,,kernel = "
   sigmoid"))
64 #####
65
66 ##### Entrenamiento de las SVM con los mejores parametros de
   la sintonizaci n #####
67
68 ### SVM Lineal ###
69 tiempo01<-system.time(svm.model1 <- svm(fototipo ~ ., data =
   resTrain, cost = as.numeric(obj1$best.parameters[1]),kernel
   = "linear",cross=10))
70 svm.pred1 <- predict(svm.model1, resTest) # predicci n
71 Mconfusion1<-table(pred = svm.pred1, true = fototipo2, dnn = c
   ("Asignado","Real")) # Matriz de confusi n
72 rownames(Mconfusion1)=c("I","II","III","IV","V","VI")
73 colnames(Mconfusion1)=c("I","II","III","IV","V","VI")
74 Resumen1<-summary(svm.model1) # muestra resumen de la
   clasificaci n
```

```
75 TotalAccuracy01<-Resumen1$tot.accuracy
76 aciertos1<-sum(diag(Mconfusion1)) / nrow(resTest) * 100 #
    porcentaje de aciertos
77
78 ### SVM RBF ###
79 tiempo02<-system.time(svm.model2 <- svm(fototipo ~ ., data =
    resTrain, cost = as.numeric(obj2$best.parameters[2]), gamma
    = as.numeric(obj2$best.parameters[1]),kernel = "radial",
    cross=10))
80 svm.pred2 <- predict(svm.model2, resTest)# predicci n
81 Mconfusion2<-table(pred = svm.pred2, true = fototipo2, dnn = c
    ("Asignado","Real")) # Matriz de confusi n
82 rownames(Mconfusion2)=c("I","II","III","IV","V","VI")
83 colnames(Mconfusion2)=c("I","II","III","IV","V","VI")
84 Resumen2<-summary(svm.model2)# muestra resumen de la
    clasificaci n
85 TotalAccuracy02<-Resumen2$tot.accuracy
86 aciertos2<-sum(diag(Mconfusion2)) / nrow(resTest) * 100 #
    porcentaje de aciertos
87
88 ### SVM Sigmaoidal ###
89 tiempo04<-system.time(svm.model4 <- svm(fototipo ~ ., data =
    resTrain, coef0=as.numeric(obj4$best.parameters[2]),gamma=
    as.numeric(obj4$best.parameters[1]), cost =as.numeric(obj4$
    best.parameters[3]),kernel = "sigmoid",cross=10))
90 svm.pred4 <- predict(svm.model4, resTest)# predicci n
91 Mconfusion4<-table(pred = svm.pred4, true = fototipo2, dnn = c
    ("Asignado","Real")) # Matriz de confusi n
92 rownames(Mconfusion4)=c("I","II","III","IV","V","VI")
93 colnames(Mconfusion4)=c("I","II","III","IV","V","VI")
94 Resumen4<-summary(svm.model4)# muestra resumen de la
    clasificaci n
95 TotalAccuracy04<-Resumen4$tot.accuracy
96 aciertos4<-sum(diag(Mconfusion4)) / nrow(resTest) * 100 #
    porcentaje de aciertos
```

```
97
98 ## Redes Neuronales ##
99
100 # Sintonizaci n de los datos de la RN
101 tiempo6<-system.time(parametrosRN <- train(fototipo ~., data=
  resTrain, method="nnet", trace=F,trControl = trainControl(
  method = "cv"))))
102 size <- parametrosRN$bestTune$size
103 decay <- parametrosRN$bestTune$decay
104 parametrosRN$bestTune
105
106
107 # Entrenamiento de la red neuronal con los valores de train
108 tiempo06<-system.time(modeloRN <- nnet(fototipo ~., size=size,
  decay=decay, trace=F,data=resTrain,maxit = 20000))
109 #modeloRN
110 # Predicci n. Se crea un data frame con las probabilidades y
  los nombres de las
111 # especies
112 prediccionesRN <- data.frame(predict(modeloRN, resTest),
  FOTOTIPOCU=predict(modeloRN,resTest, type="class"))
113 #prediccionesRN
114 mcRN <- table(prediccionesRN$FOTOTIPOCU,resTest$fototipo, dnn
  = c("Asignado","Real")) # Matriz de confusi n
115 rownames(mcRN)=c("I","II","III","IV","V","VI")
116 colnames(mcRN)=c("I","II","III","IV","V","VI")
117 aciertosRN <- sum(diag(mcRN)) / nrow(resTest) * 100 # Aciertos
  en %
118
119 # graficas de las curvas ROC
120 ytest<-resTest[,27]
121
122 truestat1=as.numeric(svm.pred1)
123 png("./imagenes/ROCSVMLLineal26rasgosNormalizados.png")
124 curvaROC1<-ROC(truestat1, ytest)
```

```
125 title("ROC-SVMLineal")
126 dev.off()
127
128 truestat2=as.numeric(svm.pred2)
129 png("./imagenes/ROCSVMBRF26rasgosNormalizados.png")
130 curvaROC2<-ROC(truestat2, ytest)
131 title("ROC-SVMRBF")
132 dev.off()
133
134
135 truestat4=as.numeric(svm.pred4)
136 png("./imagenes/ROCSVMSigmoidal26rasgosNormalizados.png")
137 curvaROC4<-ROC(truestat4, ytest)
138 title("ROC-SVMSigmoidal")
139 dev.off()
140
141
142 truestat6=as.numeric(prediccionesRN$FOTOTIPOCU)
143 png("./imagenes/ROCRN26rasgosNormalizados.png")
144 curvaROC6<-ROC(truestat6, ytest)
145 title("ROC-RN")
146 dev.off()
147
148
149 ##### Resultados #####
150
151 SVMLineal<-matrix(list(Resumen1$cost, Resumen1$gamma, Resumen1
    $tot.nSV, aciertos1, curvaROC1$AUC, tiempo01[1]))
152 colnames(SVMLineal)<-c("SVM-Lineal")
153 rownames(SVMLineal)<-c("C", "gamma", "nSV", "%Aciertos", "AreaROC",
    "Tiempo")
154 t(SVMLineal)
155
156 SVMBRF<-matrix(list(Resumen2$cost, Resumen2$gamma, Resumen2$
    tot.nSV, aciertos2, curvaROC2$AUC, tiempo02[1]))
```

```
157 colnames(SVMBRF) <-c("SVM-RBF")
158 rownames(SVMBRF) <-c("C", "gamma", "nSV", "%Aciertos", "AreaROC", "
    Tiempo")
159 t(SVMBRF)
160
161 SVMsigmoidal <-matrix(list(Resumen4$cost, Resumen4$gamma,
    Resumen4$coef0, Resumen4$tot.nSV, aciertos4, curvaROC4$AUC,
    tiempo04[1])
162 colnames(SVMsigmoidal) <-c("SVM-Sigmoidal")
163 rownames(SVMsigmoidal) <-c("C", "gamma", "coef", "nSV", "%Aciertos"
    , "AreaROC", "Tiempo")
164 t(SVMsigmoidal)
165
166 RN <-matrix(list(size, decay, aciertosRN, curvaROC6$AUC,
    tiempo06[1]))
167 colnames(RN) <-c("RN")
168 rownames(RN) <-c("Neuronas", "decay", "%Aciertos", "AreaROC", "
    Tiempo")
169 t(RN)
```

APÉNDICE B

Código con PCA

```
1  rm(list=ls()) ## limpia las variables
2  ## Se cargan las Librerias ##
3  library(xlsx) # leer datos xlsx #
4  library(e1071) # SVM, basado en LIBSVM
5  library(Epi) # curvas ROC
6  library(caret) #entrenamiento de modelos de clasificaci n
7  library(nnet) # Red Neuronal (percentron multicapa)
8  #Leer datos Excel
9  file="./DatosEspectros.xlsx" ## Datos espectrales para 31, 26
   ragos y coeficiente de absorcion
10 res1 <- t(read.xlsx(file, 1,endRow=32))[-1,] # declaraci n de
   variables
11 res2 <- t(read.xlsx(file, 2,endRow=32))[-1,]
12 res3 <- t(read.xlsx(file, 3,endRow=32))[-1,]
13 res4 <- t(read.xlsx(file, 4,endRow=32))[-1,]
14
15 res <- rbind(res1,res2,res3,res4) #matriz de datos espectrales
16 indice<-as.logical(apply(!is.na(res),1,prod))
17 res <- res[indice,-(1:5)]
18 res<-scale(res) # normalizaci n de los datos
19 res<-as.data.frame(res)
20
```

```
21 #definici n de la etiqueta de los datos (clases)
22 fototipo<-t(cbind(
23   as.vector(read.xlsx(file,1,startRow=35,endRow=36))[-1]
24   ,as.vector(read.xlsx(file,2,startRow=35,endRow=36))[-1]
25   ,as.vector(read.xlsx(file,3,startRow=35,endRow=36))[-1]
26   ,as.vector(read.xlsx(file,4,startRow=35,endRow=36))[-1]
27   ) )
28 rownames(fototipo )<-NULL
29 fototipo[fototipo=="I", ]<-1
30 fototipo[fototipo=="II", ]<-2
31 fototipo[fototipo=="III", ]<-3
32 fototipo[fototipo=="IV", ]<-4
33 fototipo[fototipo=="V", ]<-5
34 fototipo[fototipo=="VI", ]<-6
35
36 colores<-c("black","red","yellow","green","blue","pink") [
37   unclass(as.factor(fototipo))] # se fijan los colores
38   asociados para cada clase
39 png("./imagenes/Nombre de la imagen.png") # guarda imagen
40 matplot(seq(450,700,by=10),t(res),type="l",col=colores,xlab =
41   expression(lambda),ylab="% R",lty=1) #grafica de los datos
42   espectrales
43 legend("topleft", legend = c("I","II","III","IV","V","VI"),
44   col =c("black","red","yellow","green","blue","pink"),lwd=2)
45 dev.off()
46
47 ##### Pruebas de Hipotesis #####
48 bartlett.sphere(res)
49 KMOS(res, use = "pairwise.complete.obs")
50
51 ##### Componentes Principales #####
52
53 pca<-princomp(res)
54 componentes<-loadings(pca)
```

```

51
52 ##### SVM
53 index <- 1:nrow(pca$score)
54 testindex <- sample(index, trunc(length(index)/3))
55 testset <- pca$score[testindex,]
56 trainset <- pca$score[-testindex,]
57 fototipo1<-fototipo[-testindex]
58 resTrain<-data.frame(pca$score[-testindex,1],pca$score[-
    testindex,2],fototipo1) #datos de prueba
59 colnames(resTrain)<-c("comp1","comp2","fototipo")
60 fototipo2<-fototipo[testindex]
61 resTest<-data.frame(pca$score[testindex,1],pca$score[testindex
    ,2],fototipo2) #datos de entranamiento
62 colnames(resTest)<-c("comp1","comp2","fototipo")
63 datos<-data.frame(pca$score[,1],pca$score[,2],fototipo)
64 colnames(datos)<-c("comp1","comp2","fototipo")
65
66
67 ## Parametros de las SVM
68     gamma1<-c(0.001,0.01,0.1,1,1.5,2,2.5)
69     c1<-c(0.001,0.01,0.1,1,10,100,1000)
70     r1<- c(0.001,0.01,0.1,1,1.5,2,2.5)
71     d<-seq(1,5,by=1)
72 ### Sintonizaci n de parametros de las SVM
73 tiempo1<-system.time(obj1 <- tune.svm(fototipo ~ ., data =
    resTrain, cost = c1,kernel = "linear"))
74 tiempo2<-system.time(obj2 <- tune.svm(fototipo ~ ., data =
    resTrain, gamma = gamma1, cost = c1,kernel = "radial"))
75 tiempo4<-system.time(obj4 <- tune.svm(fototipo ~ ., data =
    resTrain, coef0=r1,gamma=gamma1, cost =c1,,kernel = "
    sigmoid"))
76 #####
77
78 ##### Entrenamiento de las SVM con los mejores parametros de
    la sintonizaci n #####

```

```
79
80 ### SVM Lineal ###
81 tiempo01<-system.time(svm.model1 <- svm(fototipo ~ ., data =
      resTrain, cost = as.numeric(obj1$best.parameters[1]),kernel
      = "linear",cross=10))
82 svm.pred1 <- predict(svm.model1, resTest) # predicci n
83 Mconfusion1<-table(pred = svm.pred1, true = fototipo2, dnn = c
      ("Asignado","Real")) # Matriz de confusi n
84 rownames(Mconfusion1)=c("I","II","III","IV","V","VI")
85 colnames(Mconfusion1)=c("I","II","III","IV","V","VI")
86 Resumen1<-summary(svm.model1) # muestra resumen de la
      clasificaci n
87 TotalAccuracy01<-Resumen1$tot.accuracy
88 aciertos1<-sum(diag(Mconfusion1)) / nrow(resTest) * 100 #
      porcentaje de aciertos
89
90 ### SVM RBF ###
91 tiempo02<-system.time(svm.model2 <- svm(fototipo ~ ., data =
      resTrain, cost = as.numeric(obj2$best.parameters[2]), gamma
      = as.numeric(obj2$best.parameters[1]),kernel = "radial",
      cross=10))
92 svm.pred2 <- predict(svm.model2, resTest)# predicci n
93 Mconfusion2<-table(pred = svm.pred2, true = fototipo2, dnn = c
      ("Asignado","Real")) # Matriz de confusi n
94 rownames(Mconfusion2)=c("I","II","III","IV","V","VI")
95 colnames(Mconfusion2)=c("I","II","III","IV","V","VI")
96 Resumen2<-summary(svm.model2)# muestra resumen de la
      clasificaci n
97 TotalAccuracy02<-Resumen2$tot.accuracy
98 aciertos2<-sum(diag(Mconfusion2)) / nrow(resTest) * 100 #
      porcentaje de aciertos
99
100 ### SVM Sigmaoidal ###
101 tiempo04<-system.time(svm.model4 <- svm(fototipo ~ ., data =
      resTrain, coef0=as.numeric(obj4$best.parameters[2]),gamma=
```

```

    as.numeric(obj4$best.parameters[1]), cost =as.numeric(obj4$
    best.parameters[3]),kernel = "sigmoid",cross=10))
102 svm.pred4 <- predict(svm.model4, resTest)# predicci n
103 Mconfusion4<-table(pred = svm.pred4, true = fototipo2, dnn = c
    ("Asignado","Real")) # Matriz de confusi n
104 rownames(Mconfusion4)=c("I","II","III","IV","V","VI")
105 colnames(Mconfusion4)=c("I","II","III","IV","V","VI")
106 Resumen4<-summary(svm.model4)# muestra resumen de la
    clasificaci n
107 TotalAccuracy04<-Resumen4$tot.accuracy
108 aciertos4<-sum(diag(Mconfusion4)) / nrow(resTest) * 100 #
    porcentaje de aciertos
109
110 ## Redes Neuronales ##
111
112 # Sintonizaci n de los datos de la RN
113 tiempo6<-system.time(parametrosRN <- train(fototipo ~., data=
    resTrain, method="nnet", trace=F,trControl = trainControl(
    method = "cv"))))
114 size <- parametrosRN$bestTune$size
115 decay <- parametrosRN$bestTune$decay
116 parametrosRN$bestTune
117
118
119 # Entrenamiento de la red neuronal con los valores de train
120 tiempo06<-system.time(modeloRN <- nnet(fototipo ~., size=size,
    decay=decay, trace=F,data=resTrain,maxit = 20000))
121 #modeloRN
122 # Predicci n. Se crea un data frame con las probabilidades y
    los nombres de las
123 # especies
124 prediccionesRN <- data.frame(predict(modeloRN, resTest),
    FOTOTIPOCU=predict(modeloRN,resTest, type="class"))
125 #prediccionesRN

```

```
126 mcRN <- table(prediccionesRN$FOTOTIPOCU, resTest$fototipo, dnn
    = c("Asignado", "Real")) # Matriz de confusi n
127 rownames(mcRN)=c("I", "II", "III", "IV", "V", "VI")
128 colnames(mcRN)=c("I", "II", "III", "IV", "V", "VI")
129 aciertosRN <- sum(diag(mcRN)) / nrow(resTest) * 100 # Aciertos
    en %
130
131 # graficas de las curvas ROC
132 ytest<-resTest[,27]
133
134 truestat1=as.numeric(svm.pred1)
135 png("./imagenes/ROCSVMLLineal26rasgosNormalizados.png")
136 curvaROC1<-ROC(truestat1, ytest)
137 title("ROC-SVMLLineal")
138 dev.off()
139
140 truestat2=as.numeric(svm.pred2)
141 png("./imagenes/ROCSVMRBF26rasgosNormalizados.png")
142 curvaROC2<-ROC(truestat2, ytest)
143 title("ROC-SVMRBF")
144 dev.off()
145
146
147 truestat4=as.numeric(svm.pred4)
148 png("./imagenes/ROCSVMSigmoidal26rasgosNormalizados.png")
149 curvaROC4<-ROC(truestat4, ytest)
150 title("ROC-SVMSigmoidal")
151 dev.off()
152
153
154 truestat6=as.numeric(prediccionesRN$FOTOTIPOCU)
155 png("./imagenes/ROCRN26rasgosNormalizados.png")
156 curvaROC6<-ROC(truestat6, ytest)
157 title("ROC-RN")
158 dev.off()
```

```
159
160
161 ##### Resultados #####
162
163 SVMLineal<-matrix(list(Resumen1$cost, Resumen1$gamma, Resumen1
    $tot.nSV, aciertos1, curvaROC1$AUC, tiempo01[1]))
164 colnames(SVMLineal)<-c("SVM-Lineal")
165 rownames(SVMLineal)<-c("C", "gamma", "nSV", "%Aciertos", "AreaROC",
    , "Tiempo")
166 t(SVMLineal)
167
168 SVMBRF<-matrix(list(Resumen2$cost, Resumen2$gamma, Resumen2$
    tot.nSV, aciertos2, curvaROC2$AUC, tiempo02[1]))
169 colnames(SVMBRF)<-c("SVM-RBF")
170 rownames(SVMBRF)<-c("C", "gamma", "nSV", "%Aciertos", "AreaROC", "
    Tiempo")
171 t(SVMBRF)
172
173 SVMsigmoidal<-matrix(list(Resumen4$cost, Resumen4$gamma,
    Resumen4$coef0, Resumen4$tot.nSV, aciertos4, curvaROC4$AUC,
    tiempo04[1])
174 colnames(SVMsigmoidal)<-c("SVM-Sigmoidal")
175 rownames(SVMsigmoidal)<-c("C", "gamma", "coef", "nSV", "%Aciertos",
    , "AreaROC", "Tiempo")
176 t(SVMsigmoidal)
177
178 RN<-matrix(list(size, decay, aciertosRN, curvaROC6$AUC,
    tiempo06[1]))
179 colnames(RN)<-c("RN")
180 rownames(RN)<-c("Neuronas", "decay", "%Aciertos", "AreaROC", "
    Tiempo")
181 t(RN)
```

APÉNDICE C

Código con Conglomerados-PCA

```
1  rm(list=ls()) ## limpia las variables
2  ## Se cargan las Librerias ##
3  library(xlsx) # leer datos xlsx #
4  library(e1071) # SVM, basado en LIBSVM
5  library(Epi) # curvas ROC
6  library(caret) #entrenamiento de modelos de clasificaci n
7  library(nnet) # Red Neuronal (percentron multicapa)
8  #Leer datos Excel
9  file="./DatosEspectros.xlsx" ## Datos espectrales para 31, 26
   ragos y coeficiente de absorcion
10 res1 <- t(read.xlsx(file, 1,endRow=32))[-1,] # declaraci n de
   variables
11 res2 <- t(read.xlsx(file, 2,endRow=32))[-1,]
12 res3 <- t(read.xlsx(file, 3,endRow=32))[-1,]
13 res4 <- t(read.xlsx(file, 4,endRow=32))[-1,]
14
15 res <- rbind(res1,res2,res3,res4) #matriz de datos espectrales
16 indice<-as.logical(apply(!is.na(res),1,prod))
17 res <- res[indice,-(1:5)]
18 res<-scale(res) # normalizaci n de los datos
19 res<-as.data.frame(res)
```

20

```
21 #definición de la etiqueta de los datos (clases)
22 fototipo<-t(cbind(
23   as.vector(read.xlsx(file,1,startRow=35,endRow=36))[-1]
24 ,as.vector(read.xlsx(file,2,startRow=35,endRow=36))[-1]
25 ,as.vector(read.xlsx(file,3,startRow=35,endRow=36))[-1]
26 ,as.vector(read.xlsx(file,4,startRow=35,endRow=36))[-1]
27   ) )
28 rownames(fototipo )<-NULL
29 fototipo[fototipo=="I", ]<-1
30 fototipo[fototipo=="II", ]<-2
31 fototipo[fototipo=="III", ]<-3
32 fototipo[fototipo=="IV", ]<-4
33 fototipo[fototipo=="V", ]<-5
34 fototipo[fototipo=="VI", ]<-6
35
36 colores<-c("black","red","yellow","green","blue","pink") [
   unclass(as.factor(fototipo))] # se fijan los colores
   asociados para cada clase
37 png("./imagenes/Nombre de la imagen.png") # guarda imagen
38 matplot(seq(450,700,by=10),t(res),type="l",col=colores,xlab =
   expression(lambda),ylab="% R",lty=1) #grafica de los datos
   espectrales
39 legend("topleft", legend = c("I","II","III","IV","V","VI"),
   col =c("black","red","yellow","green","blue","pink"),lwd=2)
40 dev.off()
41
42 #####Conglomerados###
43 i<-1
44 res1<-res
45 kmedias2<-kmeans(res1, centers=6)
46 eje<-seq(400,700,by=10)
47 orden<-as.numeric(names(sort(kmedias2$centers[,1],decreasing =
   TRUE)))
48 color<-c("black","red","yellow","green","blue","pink")
49
```

```
50 ##### Reordenamiento de los conglomerados respecto a los
    fototipos cutaneos #####
51 color1<-color[orden]
52 colores<-color1[unclass(as.factor(kmedias2$cluster))]
53 colores[kmedias2$cluster==orden[1]]<-color[1]
54 colores[kmedias2$cluster==orden[2]]<-color[2]
55 colores[kmedias2$cluster==orden[3]]<-color[3]
56 colores[kmedias2$cluster==orden[4]]<-color[4]
57 colores[kmedias2$cluster==orden[5]]<-color[5]
58 colores[kmedias2$cluster==orden[6]]<-color[6]
59
60
61 ##### Componentes Principales #####
62
63 pca<-princomp(res1)
64 componentes<-loadings(pca)
65
66 ##### Comparacion entre criterio medico y conglomerados
    #####
67 fototipo[colores=="black"]<-"1"
68 fototipo[colores=="red"]<-"2"
69 fototipo[colores=="yellow"]<-"3"
70 fototipo[colores=="green"]<-"4"
71 fototipo[colores=="blue"]<-"5"
72 fototipo[colores=="pink"]<-"6"
73
74 datostotal<-c(sum(as.numeric(fototipo==1)),sum(as.numeric(
    fototipo==2)),sum(as.numeric(fototipo==3)),sum(as.numeric(
    fototipo==4)),sum(as.numeric(fototipo==5)),sum(as.numeric(
    fototipo==6)))
75 clase<-c("I","II","III","IV","V","VI")
76 cbind(datostotal,clase)
77 medicoKmedia<-cbind(datostotaloriginal,datostotal,clase)
78 colnames(medicoKmedia)=c("Criterio-Medico","Criterio-Kmedia",
    fototipo")
```

```
79
80 claseoriginal<-data.frame(res,fototipooriginal)[,32]
81 claseconglomerado<-data.frame(res,fototipo)[,32]
82 Mconfusionconglomerado<-table(pred = claseconglomerado , true
    = claseoriginal, dnn = c("Asignado","Real"))
83 rownames(Mconfusionconglomerado)=c("I","II","III","IV","V","VI
    ")
84 colnames(Mconfusionconglomerado)=c("I","II","III","IV","V","VI
    ")
85 aciertosconglomerado<-sum(diag(Mconfusionconglomerado)) /
    length(claseoriginal) * 100
86
87
88 ##### SVM
89 index <- 1:nrow(pca$score)
90 #testindex <- sample(index, trunc(length(index)/3))
91 testset <- pca$score[testindex,]
92 trainset <- pca$score[-testindex,]
93 fototipo1<-fototipo[-testindex]
94 resTrain<-data.frame(pca$score[-testindex,1],pca$score[-
    testindex,2],fototipo1) #datos de entranamiento
95 colnames(resTrain)<-c("comp1","comp2","fototipo")
96 fototipo2<-fototipo[testindex]
97 resTest<-data.frame(pca$score[testindex,1],pca$score[testindex
    ,2],fototipo2) #datos de prueba
98 colnames(resTest)<-c("comp1","comp2","fototipo")
99 datos<-data.frame(pca$score[,1],pca$score[,2],fototipo)
100 colnames(datos)<-c("comp1","comp2","fototipo")
101
102 ## Parametros de las SVM
103     gamma1<-c(0.001,0.01,0.1,1,1.5,2,2.5)
104     c1<-c(0.001,0.01,0.1,1,10,100,1000)
105     r1<- c(0.001,0.01,0.1,1,1.5,2,2.5)
106     d<-seq(1,5,by=1)
107 ### Sintonazaci n de parametros de las SVM
```

```
108 tiempo1<-system.time(obj1 <- tune.svm(fototipo ~ ., data =
    resTrain, cost = c1, kernel = "linear"))
109 tiempo2<-system.time(obj2 <- tune.svm(fototipo ~ ., data =
    resTrain, gamma = gamma1, cost = c1, kernel = "radial"))
110 tiempo4<-system.time(obj4 <- tune.svm(fototipo ~ ., data =
    resTrain, coef0=r1, gamma=gamma1, cost =c1,, kernel = "
    sigmoid"))
111 #####
112
113 ##### Entrenamiento de las SVM con los mejores parametros de
    la sintonizaci n #####
114
115 ### SVM Lineal ###
116 tiempo01<-system.time(svm.modelo1 <- svm(fototipo ~ ., data =
    resTrain, cost = as.numeric(obj1$best.parameters[1]), kernel
    = "linear", cross=10))
117 svm.pred1 <- predict(svm.modelo1, resTest) # predicci n
118 Mconfusion1<-table(pred = svm.pred1, true = fototipo2, dnn = c
    ("Asignado", "Real")) # Matriz de confusi n
119 rownames(Mconfusion1)=c("I", "II", "III", "IV", "V", "VI")
120 colnames(Mconfusion1)=c("I", "II", "III", "IV", "V", "VI")
121 Resumen1<-summary(svm.modelo1) # muestra resumen de la
    clasificaci n
122 TotalAccuracy01<-Resumen1$tot.accuracy
123 aciertos1<-sum(diag(Mconfusion1)) / nrow(resTest) * 100 #
    porcentaje de aciertos
124
125 ### SVM RBF ###
126 tiempo02<-system.time(svm.modelo2 <- svm(fototipo ~ ., data =
    resTrain, cost = as.numeric(obj2$best.parameters[2]), gamma
    = as.numeric(obj2$best.parameters[1]), kernel = "radial",
    cross=10))
127 svm.pred2 <- predict(svm.modelo2, resTest)# predicci n
128 Mconfusion2<-table(pred = svm.pred2, true = fototipo2, dnn = c
    ("Asignado", "Real")) # Matriz de confusi n
```

```
129 rownames(Mconfusion2)=c("I","II","III","IV","V","VI")
130 colnames(Mconfusion2)=c("I","II","III","IV","V","VI")
131 Resumen2<-summary(svm.model2)# muestra resumen de la
      clasificaci n
132 TotalAccuracy02<-Resumen2$tot.accuracy
133 aciertos2<-sum(diag(Mconfusion2)) / nrow(resTest) * 100 #
      porcentaje de aciertos
134
135 ### SVM Sigmaoidal ###
136 tiempo04<-system.time(svm.model4 <- svm(fototipo ~ ., data =
      resTrain, coef0=as.numeric(obj4$best.parameters[2]),gamma=
      as.numeric(obj4$best.parameters[1]), cost =as.numeric(obj4$
      best.parameters[3]),kernel = "sigmoid",cross=10))
137 svm.pred4 <- predict(svm.model4, resTest)# predicci n
138 Mconfusion4<-table(pred = svm.pred4, true = fototipo2, dnn = c
      ("Asignado","Real")) # Matriz de confusi n
139 rownames(Mconfusion4)=c("I","II","III","IV","V","VI")
140 colnames(Mconfusion4)=c("I","II","III","IV","V","VI")
141 Resumen4<-summary(svm.model4)# muestra resumen de la
      clasificaci n
142 TotalAccuracy04<-Resumen4$tot.accuracy
143 aciertos4<-sum(diag(Mconfusion4)) / nrow(resTest) * 100 #
      porcentaje de aciertos
144
145 ## Redes Neuronales ##
146
147 # Sintonizaci n de los datos de la RN
148 tiempo6<-system.time(parametrosRN <- train(fototipo ~., data=
      resTrain, method="nnet", trace=F,trControl = trainControl(
      method = "cv"))))
149 size <- parametrosRN$bestTune$size
150 decay <- parametrosRN$bestTune$decay
151 parametrosRN$bestTune
```

152

153

```
154 # Entrenamiento de la red neuronal con los valores de train
155 tiempo06<-system.time(modeloRN <- nnet(fototipo ~., size=size,
      decay=decay, trace=F,data=resTrain,maxit = 20000))
156 #modeloRN
157 # Predicci n. Se crea un data frame con las probabilidades y
      los nombres de las
158 # especies
159 prediccionesRN <- data.frame(predict(modeloRN, resTest),
      FOTOTIPOCU=predict(modeloRN,resTest, type="class"))
160 #prediccionesRN
161 mcRN <- table(predict(modeloRN$resTest$FOTOTIPOCU, resTest$fototipo, dnn
      = c("Asignado","Real")) # Matriz de confusi n
162 rownames(mcRN)=c("I","II","III","IV","V","VI")
163 colnames(mcRN)=c("I","II","III","IV","V","VI")
164 aciertosRN <- sum(diag(mcRN)) / nrow(resTest) * 100 # Aciertos
      en %
165
166 # graficas de las curvas ROC
167 ytest<-resTest[,27]
168
169 truestat1=as.numeric(svm.pred1)
170 png("./imagenes/ROCSVMLLineal26rasgosNormalizados.png")
171 curvaROC1<-ROC(truestat1, ytest)
172 title("ROC-SVMLLineal")
173 dev.off()
174
175 truestat2=as.numeric(svm.pred2)
176 png("./imagenes/ROCSVMBRF26rasgosNormalizados.png")
177 curvaROC2<-ROC(truestat2, ytest)
178 title("ROC-SVMRBF")
179 dev.off()
180
181
182 truestat4=as.numeric(svm.pred4)
183 png("./imagenes/ROCSVMSigmoidal26rasgosNormalizados.png")
```

```
184 curvaROC4<-ROC(truestat4, ytest)
185 title("ROC-SVMSigmoidal")
186 dev.off()
187
188
189 truestat6=as.numeric(prediccionesRN$FOTOTIPOCU)
190 png("./imagenes/ROCRN26rasgosNormalizados.png")
191 curvaROC6<-ROC(truestat6, ytest)
192 title("ROC-RN")
193 dev.off()
194
195
196 ##### Resultados #####
197
198 SVMlineal<-matrix(list(Resumen1$cost, Resumen1$gamma, Resumen1
   $tot.nSV, aciertos1, curvaROC1$AUC, tiempo01[1]))
199 colnames(SVMlineal)<-c("SVM-Lineal")
200 rownames(SVMlineal)<-c("C", "gamma", "nSV", "%Aciertos", "AreaROC",
   , "Tiempo")
201 t(SVMlineal)
202
203 SVMBRF<-matrix(list(Resumen2$cost, Resumen2$gamma, Resumen2$
   tot.nSV, aciertos2, curvaROC2$AUC, tiempo02[1]))
204 colnames(SVMBRF)<-c("SVM-RBF")
205 rownames(SVMBRF)<-c("C", "gamma", "nSV", "%Aciertos", "AreaROC", "
   Tiempo")
206 t(SVMBRF)
207
208 SVMsigmoidal<-matrix(list(Resumen4$cost, Resumen4$gamma,
   Resumen4$coef0, Resumen4$tot.nSV, aciertos4, curvaROC4$AUC,
   tiempo04[1])
209 colnames(SVMsigmoidal)<-c("SVM-Sigmoidal")
210 rownames(SVMsigmoidal)<-c("C", "gamma", "coef", "nSV", "%Aciertos",
   , "AreaROC", "Tiempo")
211 t(SVMsigmoidal)
```

```
212  
213 RN<-matrix(list(size, decay , aciertosRN ,curvaROC6$AUC,  
    tiempo06 [1]))  
214 colnames(RN)<-c("RN")  
215 rownames(RN)<-c("Neuronas", "decay", "%Aciertos", "AreaROC", "  
    Tiempo")  
216 t(RN)
```

REFERENCIAS

- Boser., B. E. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th annual ACM workshop on Computational Learning Theory*, pages 144–152. ACM Press.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. In *Machine Learning*, pages 273–297.
- Cuadras, C. (1991). *Métodos de Análisis Multivariante*.
- Fitzpatrick, T. and Freedberg, I. (2005). *Dermatología - Dermatología en medicina general*. Madrid.
- Geneser, F. (2000). *Histología TEXbook*. Médica Panamericana S.A.
- González, F. J., Martínez-Escaname, M., Muñoz, R. I., Torres-Alvarez, B., and Moncada, B. (2010). Diffuse reflectance spectrophotometry for skin phototype determination. *Skin Research and Technology*, 16:397–400.
- Hilera and Martínez (1995). *Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones*. RA-MA.
- Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2010). A practical guide to support vector classification.
- Hunter, J. A. A., Savin, J. A., and Dahl., M. V. (1995). *Clinical Dermatology TEXbook*. Blackwell Science.

- Hurtado, J. (2010). *Metodología de la Investigación*. SYPAL.
- Ishimaru, A. (1999). *Wave Propagation and Scattering in Random Media*. An IEEE OUP classic reissue. Wiley.
- Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., and Murthy, K. R. K. (2001). Improvements to platt's smo algorithm for svm classifier design. *Neural Comput.*, 13(3):637–649.
- Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, Berkeley, Calif. University of California Press.
- McCulloch and Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133.
- Meglinski, I. and Matcher, S. (2003). Computer simulation of the skin reflectance spectra. *Computer Methods and Programs in Biomedicine*, 70:179–186.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2015). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.6-7.
- Narea, F., Vivas, S., and Muñoz, A. (2015). Retrieving the absorption coefficient of epidermis in human skin. *Optica pura y aplicada*, 48:207–214.
- Orozco., E. E., Garcia, G. I., y Montiel, S. V., Delgado-Atencio, J. A., Ramos, J. C., and Delgado, F. G. (2010). Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa. *Revista Ingeniería Biomédica*, 4:34–40.
- Pardo, E., Vivas, S., and Muñoz, A. (2015). *Firmas espectrales de fototipos cutáneos de habitantes de Venezuela. Valencia-estado Carabobo-Venezuela. 2015*.
- Parikh, K. S. and Shah, T. P. (2016). 3rd international conference on innovations in automation and mechatronics engineering 2016, iciame 2016 05-06 february, 2016 support vector machine – a large margin classifier to diagnose skin illnesses. *Procedia Technology*, 23:369 – 375.

- Prahl, S. A., Keijzer, M., Jacques, S. L., and Welch, A. J. (1989). A monte carlo model of light propagation in tissue. In *SPIE Proceedings of Dosimetry of Laser Radiation in Medicine and Biology*.
- Prati, R. C., Batista, and Monard, M. C. (2008). Curvas ROC para avaliação de classificadores. *Revista IEEE América Latina*.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Schölkopf, B. and Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive computation and machine learning. MIT Press.
- Searle (1994). *Neural Networks and Statistical Models*, volume 1. USA.
- Vapnik, V. and Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24:774–780.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.
- Viñuela, P. and León, I. (2004). *Redes de neuronas artificiales: un enfoque práctico*. Pearson Educación.
- W. F. Cheong, S. A. P. and Welch, A. J. (1990). Review of the optical properties of a biological tissues. *IEEE J. of Quan. Elec.*, 26:2166–2185.
- Wang, L.-H. and Jacques, S. (1993). Hybrid model of monte carlo simulation and diffusion theory for light reflectance by turbid media. *J. Opt. Soc. Am. A*, 10:1746–1752.
- Westland, S., Ripamonti, C., and Cheung, V. (2012). *Computational Colour Science Using MATLAB*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley.
- Wikipedia (2017). Perceptrón — wikipedia, la enciclopedia libre. [Internet; descargado 8-junio-2017].