



UAEM | Universidad Autónoma
del Estado de México

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE
MÉXICO

INGENIERÍA EN COMPUTACIÓN

Detección Automática de Glaucoma y Retinopatía Diabética
Usando Características Cromáticas y Texturales

Tesis que presenta
Maria Fernanda URDAÑEZ CARBAJAL

Para obtener el Grado de
Ingeniero en Computación

Asesor de Tesis:
Dr. Jair CERVANTES

Texcoco, Estado de México.

Diciembre del 2020



AUTONOMOUS UNIVERSITY OF MEXICO STATE COMPUTER
SCIENCE DEPARTMENT

**Automatic Detection of Glaucoma and Diabetic Retinopathy
by Using Chromatic and Textural Features**

Submitted by

Maria Fernanda URDAÑEZ CARBAJAL

As a fulfillment of the requirement for the degree of

Computer Science Engineering

Thesis Advisor:

Ph.D. Jair CERVANTES

Texcoco, Estado de México.

December 2020

Abstract

The application of vision systems in several fields has been increasing in recent years. The main challenges have been to improve performance and reduce computational costs or algorithmic complexity.

The objective of this research is to analyze the several techniques of features extraction of a vision system that allows classifying glaucoma and diabetic retinopathy from textural and chromatic features, by using Local Binary Patterns, Haralick and Histograms of Oriented gradients. In the experiments carried out, different classifiers were used such as: neural networks, decision trees, SVM and Naive bayes, carrying out the evaluation by using cross validation and area under the curve performance metrics, which allowed us to have a comparison of the performance of the classifiers as well as the representation of the rate of false positives versus false negatives. The results obtained are discussed and compared in this thesis, which will also allow making a selection of a technique based on its complexity and performance.

Resumen

La aplicación de los sistemas de visión en diversos campos ha ido creciendo en los últimos años. Los principales retos han sido la mejora del desempeño y disminuir los costos computacionales o complejidad algorítmica. Esta investigación tiene como objetivo analizar las diferentes técnicas de extracción de características de un sistema de visión que permite clasificar el glaucoma y retinopatía diabética a partir de características texturales y cromáticas, de las cuales se utilizaron en este análisis Patrones Binarios Locales, Haralick e Histogramas de Gradientes Ordenados. En los experimentos realizados se utilizaron diferentes clasificadores como: redes neuronales, árboles de decisión, SVM y Naive bayes, realizando la evaluación a partir de técnicas como validación cruzada y área bajo la curva, que nos permitieron tener una comparación del desempeño de los clasificadores así como la representación de la taza de falsos positivos frente a los falsos negativos. Los resultados obtenidos son analizados y comparados en esta tesis, que permitirán también hacer una selección de alguna técnica en base a su complejidad y desempeño.

Índice general

1. Introducción	2
1.1. Problemática	4
1.2. Justificación	4
1.3. Objetivos	4
1.4. Hipótesis	5
1.5. Antecedentes	5
1.5.1. Sistema de Reconocimiento de Objetos	5
1.5.2. Enfermedades de los Ojos	8
2. Preliminares	11
2.1. Técnicas de Clasificación	11
2.1.1. Neural networks	11
2.1.2. Support Vector Machines	15
2.2. Técnicas de Evaluación de desempeño	18
2.2.1. Cross-validation	19
2.2.2. F-Measure	20
2.2.3. Área ROC	21
3. Técnicas de segmentación y extracción de características	23
3.1. Técnicas de segmentación	23
3.1.1. Frontera Global	24
3.1.2. Segmentación adaptativa	24
3.1.3. Segmentación por clustering	25

3.1.4. Segmentación Otsu	26
3.1.5. Técnicas de dilatación y erosión	27
3.1.6. Técnicas para eliminar ruido en imágenes binarias	28
3.2. Técnicas de Extracción de características	29
3.2.1. Características texturales de Haralick	29
3.2.2. Patrones binarios locales	32
3.3. Histograma de Gradientes Ordenados	34
3.3.1. Características cromáticas	38
4. Metodología	39
4.1. Conjunto de datos utilizado	40
4.2. Pre procesamiento	41
4.3. Segmentación	41
4.4. Extracción de características	42
4.4.1. Normalización de datos	43
4.5. Clasificación e identificación	45
4.6. Análisis de desempeño	46
5. Resultados experimentales	47
5.1. Desempeño de clasificadores utilizando Histograma de Gradientes Orientados	47
5.2. Desempeño de clasificadores utilizando características de Haralick	50
5.3. Desempeño de clasificadores utilizando LBP	52
5.4. Desempeño de clasificadores para retinopatía diabética (Conjunto de datos 2)	54
5.4.1. Prueba 1	54
5.4.2. Prueba 2	56
6. Conclusiones	58

Índice de figuras

1.1.	Sistema de detección de objetos	6
2.1.	Conexión entre una neurona presináptica y una postsináptica	13
2.2.	Validación cruzada de 4 iteraciones	20
2.3.	Representación de similitud entre dos clases. El punto de corte t determina el comportamiento del clasificador.	22
3.1.	Muestra de píxel P (central) de una imagen I	32
3.2.	Equivalente binario LBP	33
3.3.	Cadena binaria, resultado de la comparación de los vecinos del píxel P . .	34
3.4.	Dirección y magnitud de gradientes	34
3.5.	Gradiente trasladado al eje de las coordenadas.	35
4.1.	Pasos más usuales en métodos de selección de características por filtrado .	39
4.2.	Conjunto de datos utilizado a) Retinopatía diabética b) Glaucoma c) Saludable	40
4.3.	Segmentación de las imágenes con el método de Otsu	42
5.1.	Gráfica de resultados usando HoG	49
5.2.	Gráfica de resultados usando Haralick	51
5.3.	Gráfica de resultados usando LBP	53
5.4.	Gráfica de resultados para la detección de la retinopatía en cualquiera de sus grados	55

5.5. Gráfica de resultados para la detección de la retinopatía en alguno de sus grados	57
---	----

Índice de cuadros

5.1.	Resultados utilizando HoG	48
5.2.	Resultados utilizando características texturales de Haralick	51
5.3.	Resultados utilizando LBP	52
5.4.	Resultados para la detección de retinopatía en cualquiera de sus grados . .	55
5.5.	Resultados de la detección en grados de retinopatía	56

Capítulo 1

Introducción

El desarrollo de sistemas de visión ha simplificado significativamente tareas en diferentes áreas de las ciencias en las que se desarrolla actividades como aprendizaje máquina, cómputo evolutivo, procesamiento de imágenes, entre otras. En los últimos años se han desarrollado numerosos sistemas de visión que ayudan a resolver problemas en múltiples aplicaciones de la vida real. La visión por computadoras y el procesamiento de imágenes tienen mucho en común, siendo diferente el objetivo de cada una. En el procesamiento de imágenes se busca mejorar la calidad de una imagen, como eliminar defectos, corregir desenfoque, mejorar la calidad de colores, entre otras. En el que se tiene una imagen de entrada y como salida la imagen mejorada. En el sistema de visión se tiene como objetivo obtener la descripción e interpretación por computadoras, teniendo una imagen como entrada, la cual será procesada para extraer atributos y como salida una descripción de la imagen.

En el presente trabajo se hace una comparación y análisis del desempeño de extracción y clasificación de características en imágenes digitales que son obtenidas de pacientes con glaucoma y retinopatía diabética, de las que se obtienen las cualidades más significativas de estas, que son características texturales y cromáticas. El glaucoma y retinopatía diabética son enfermedades oculares que afectan la visión provocando pérdida total. Para la primera afección el humor acuoso contenido en el ojo no drena de manera correcta debido a un fallo del sistema del drenaje ocular provocando una presión intraocular mayor, esto

genera que las fibras nerviosas que se juntan en el disco óptico se dañen y puedan morir, evitando que lleguen señales al nervio óptico (**Maul, 2002**).

Para la retinopatía producida en pacientes diabéticos, los niveles altos en el azúcar causan que los vasos sanguíneos que alimentan a la retina mueran, el ojo intenta crear nuevas de estas estructuras para continuar con la correcta irrigación sanguínea pero estos nuevos vasos son débiles lo que provoca que se fracturen y sangren. En las primeras etapas de este padecimiento los vasos sanguíneos débiles se dilatan, generándose bultos que sobresalen de las paredes. Entre más vasos sanguíneos se dañen más grave será, pues provoca que las fibras nerviosas y la mácula se inflamen, En este grado del padecimiento aún hay tratamiento. El problema se agrava, como ya se ha mencionado, con la muerte de los vasos sanguíneos, provocando el desprendimiento de la retina. La retinopatía puede dar paso al glaucoma, pues el tejido cicatricial afecta al drenaje de líquido acuoso.

Los sistemas de visión automáticos ayudan a una detección eficiente y eficaz sin la intervención de un profesional agilizando los resultados. Por ellos la importancia de la mejor selección de características texturales y cromáticas.

1.1. Problemática

El glaucoma y la retinopatía diabética son enfermedades que afectan a gran parte de la población Mexicana. “en México se estima que existen cerca de 1.5 millones de personas con glaucoma, y hasta 50 mil casos de ceguera a causa de su detección tardía” (**UNAM, 2020**). ”15.5 millones de personas sufren diabetes en América Latina y 10.6 millones en México” (**Forbes, 2015**), lo cual es indicativo que casi el 10 por ciento de la población corre riesgo de padecer retinopatía, estas enfermedades son silenciosas y sin cura, causando ceguera irreversible si no se detecta a tiempo.

1.2. Justificación

Enfermedades oculares como retinopatía diabética y glaucoma requieren un rápido tratamiento para disminuir el progreso de la enfermedad y sus efectos. Un diagnóstico en fase temprana es de gran importancia para este propósito. Sin embargo, un diagnóstico certero solo es posible si se lleva a cabo por un médico especialista cuyo entrenamiento y capacitación requiere varios años.

El propósito de esta tesis es diseñar un sistema de visión que permita realizar detección automática de glaucoma y retinopatía diabética. Se realizará un análisis del desempeño de varios clasificadores con base a diferentes características texturales y cromáticas. Lo que ayudará a mejorar la precisión del diagnóstico así como disminuir falsos positivos y falsos negativos.

1.3. Objetivos

El objetivo general de esta tesis es realizar un sistema de visión automático utilizando características texturales y cromáticas para clasificar a) Glaucoma y b) Retinopatía diabética. Se pretende alcanzar los siguientes objetivos específicos:

1. Realizar un análisis de las diferentes técnicas de extracción de características texturales para la detección de glaucoma y retinopatía

2. Obtener un análisis estadístico de las características para mejorar la precisión de clasificación
3. Probar el funcionamiento del sistema.
4. Discutir resultados y conclusiones obtenidas

1.4. Hipótesis

El uso de características texturales y cromáticas adecuadas pueden mejorar el desempeño de los clasificadores para la detección de retinopatía diabética y glaucoma.

1.5. Antecedentes

A continuación se explicarán algunos trabajos que permitirán el desarrollo de esta tesis, en los que se describe los sistemas de reconocimiento de objetos y su relación en el campo de la medicina, los cuales nos permiten tener un panorama más amplio

1.5.1. Sistema de Reconocimiento de Objetos

Los sistemas de visión han sido una parte importante en la inteligencia artificial, ya que a sistemas automáticos se le dota de cualidades e indicaciones para obtener información acerca de imágenes capturadas del entorno, con el fin de obtener información relevante e ignorar lo que causa ruido a los objetos de interés. Para un humano es relativamente fácil detectar objetos sin importar la posición, color y su intensidad, el cerebro hace una clasificación de la información necesaria y de la que no sirve en ese momento, todo a través de procesos electroquímicos. Pero cuando un sistema artificial tiene que hacer la detección del mismo objeto, se encuentra con una cantidad de información exponencial en la que podemos resaltar una ventaja, esta es capaz detectar objetos que pudieran ser imperceptibles para el ser humano, promoviendo infinidad de aplicaciones (**Shapiro, 2001**).

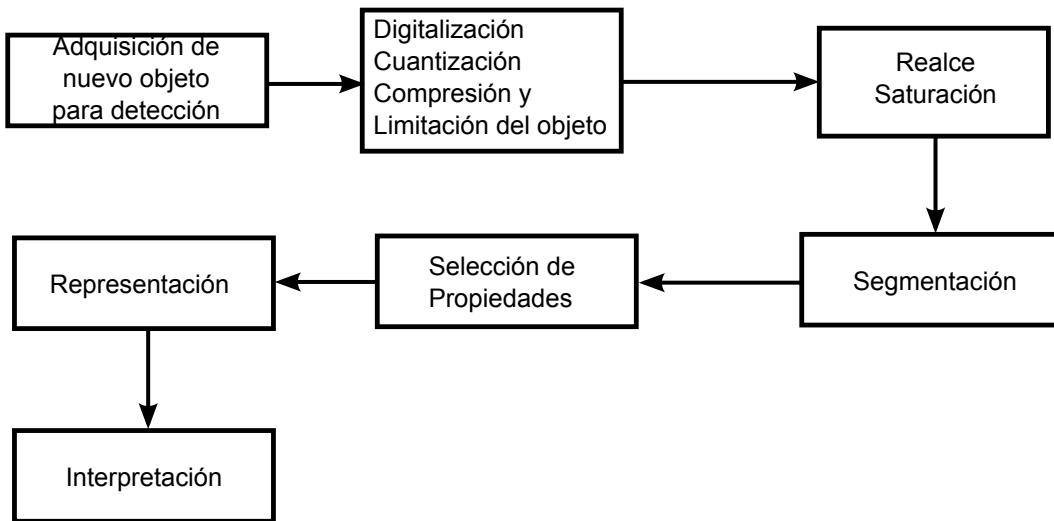


Figura 1.1: Sistema de detección de objetos

Al sistema de visión se le asignan funcionalidades para la adquisición y análisis de las imágenes. La formación de imágenes sucede cuando un sensor, sea ojo o una cámara registra la luz que interacciona con los objetos el cual da como resultado una función bidimensional, donde cada punto de la imagen pertenece a la intensidad en imágenes monocromáticas, al que se le asignan un sistema de coordenadas (x, y) siendo el punto superior izquierdo el origen (Prensky, 2001). En el análisis de la imagen se hace el procesamiento donde se utilizan técnicas para extraer todas las características posibles que identifiquen al objeto de interés, un esquema simple del proceso de reconocimiento de objetos en imágenes puede ser observado en la Figura 1.1.

El análisis de imágenes se puede subdividir en cuatro fases:

- Pre-procesamiento: Es el preliminar del procesamiento, en el cual se busca mejorar la imagen y modificar sus características, como son:
 - Los niveles de gris
 - Contraste
 - Eliminación de ruido
 - Realce de características

- Segmentación: La imagen se divide hasta que se aísle el objeto de interés, para llevar a cabo este procedimiento se evalúan los pixeles con sus vecinos para denotar los cambios en los niveles de gris o de forma contraria, los pixeles similares.
- Descripción: Aquí se extraen las características más significativas del objeto, para que el descriptor tenga suficiente información y se pueda distinguir un objeto de otro.

En este punto es importante destacar que existen diferentes descriptores visuales, en los que definen entre otras características que son color, textura y forma. Los cuales definen a la región correspondiente del objeto, aunque a veces se pueden obtener de toda la imagen.

Las características cromáticas proporcionan mayor información del objeto lo que ayuda a mejorar su clasificación. El humano pude detectar miles de colores con diferentes intensidades y saturaciones, pero solo puede detectar 20 niveles de gris. Por esto se deduce que el color tiene mayor influencia en la clasificación de objetos.

En imágenes los objetos o regiones no son uniformes, teniendo pequeños elementos entrelazados y poco distinguibles conocidos como textura, por lo que en algunos casos detectar los primitivos (texel) tiene mayor complejidad que otros, influyendo en este tipo de características la resolución, pues como se ha mencionado, para detectar los elementos en una resolución alta sería más fácil que en una resolución menor, pues los elementos se pierden llegando a parecer que son una región uniforme.

Aunque en la detección de objetos es importante las características ya mencionadas, el ser humano también puede hacer clasificaciones de formas, detectando las orillas o siluetas. En la visión por computadoras es una parte importante ya que se pueden detectar bordes discretizados por las variaciones de intensidad, ayudando en tareas de bajo nivel como detección de esquinas bordes, y para un nivel intermedio realizar segmentación y separación de diferentes objetos de una imagen.

- Reconocimiento: Consiste en identificar cada objeto de una escena, el cual depende de un aprendizaje previo donde las características de una sola clase tienen un patrón;

posteriormente viene la clasificación, que es el cálculo de proximidades del vector de características x con el patrón m .

1.5.2. Enfermedades de los Ojos

Para poder entender que es el glaucoma y la retinopatía, se definen algunos conceptos que preceden a la explicación del funcionamiento de los ojos.

- Córnea: Primer lente del ojo que brinda protección ocular
- Iris y pupila: Iris es la parte de color del ojo que se expande y contrae para dejar pasar diferentes cantidades de luz a través de la pupila.
- Cristalino: Segunda lente, se encuentra entre el iris y el humor vítreo.
- Retina: Superficie interior del ojo, sensible a la luz.
- Humor vítreo: líquido transparente con una consistencia gelatinosa, comprendido entre el iris y la parte interna de la retina.
- Humor acuoso: líquido transparente que sirve para oxigenar y aportar nutrientes al ojo, que va desde el cristalino hasta la córnea.

Este último es parte de un proceso llamado ciclo de drenaje, cuando se hace de forma correcta, el humor acuoso sale por la pupila y este es absorbido por el torrente sanguíneo a través de una malla de canales de drenado que se encuentran en el borde externo del iris llamado sistema de drenaje óptico, la cantidad del humor acuoso nos permite medir la presión ocular.

Para que el ojo perciba las imágenes del mundo exterior, la luz que pasa por la pupila es dirigida al cristalino para que este a su vez lo enfoque en la retina donde existe fotoreceptores que envían la información al cerebro por medio del nervio óptico.

- Glaucoma

Esta enfermedad consiste en el fallo del sistema de drenaje, se bloquea impidiendo el correcto flujo del humor acuoso aumentando la presión ocular, por ende el nervio óptico se daña así como las fibras nerviosas, al comenzar esto el disco óptico comienza a hacerse hueco rechazando las fibras del disco óptico, el cual adoptará la forma de copa o curva, la cual si se mantiene por mucho tiempo provoca la pérdida de la visión (**Kanski, 2004**).

Generalmente no hay síntomas dolorosos, pero conforme va avanzando se puede presentar: visión borrosa, dolor de ojos y de cabeza, náuseas y vómito, pérdida repentina de la visión.

En (**Encalada, 2016**) realizaron un trabajo de grado para la identificación del glaucoma basándose en las características del nervio óptico, en el que aplicaron técnicas de pre-procesamiento y segmentación para identificar la copa y disco, el inconveniente que ellos mismos definen es que algunas personas tienen el nervio óptico más grande y esto afectaría a los resultados, solucionándolo con una regla que permite medir la relación entre el disco y copa óptica.

- Retinopatía

Una persona con diabetes tiene los niveles altos de azúcar en la sangre, lo que provoca daño en los vasos sanguíneos de la retina que pueden llegar a hincharse cerrarse o morir; el cuerpo en el intento de restaurar lo perdido crea nuevos vasos anormales y frágiles provocando su fractura, lo que genera sangrado y bloqueo de la vista.

En (**Wong, 2016**) hacen mención del sistema para la detección de retinopatía a través de la tecnología de aprendizaje profundo hecho por la IA de Google, la cual ya utilizan para identificar los grandes volúmenes de datos como los búsquedas de interés de usuarios o agrupamiento de imágenes que son cargadas en Google fotos. Esta vez trasladada a un área de medicina, en las que se procesaron y probaron las imágenes de pacientes con retinopatía, superando las pruebas con éxito, en las que ellos señalan algunas limitaciones:

- El conjunto de datos con retinopatía severa es muy pequeño, por ende los resultados no pueden ser representativos para este tipo de clase, en los que hay realmente riesgo de ceguera.

- El algoritmo solo detecta la retinopatía, excluyendo otros padecimientos como el glaucoma o degeneración macular.

Capítulo 2

Preliminares

En esta sección se describen algunas técnicas de clasificación y métodos para evaluar su desempeño.

2.1. Técnicas de Clasificación

La tarea de la clasificación consiste en ordenar N número de individuos en K clases, las cuales se han definido por medio de patrones o similitudes en las características que dependen de valores cuantitativos, es decir entre más cerca está un vector de características de una clase o grupo, este vector se etiquetará con el nombre de este patrón. Es por ello la importancia de definir el método óptimo para obtener una clasificación adecuada.

2.1.1. Neural networks

Las redes neuronales artificiales (RNA) están inspiradas en el cerebro humano, su funcionamiento consiste en recibir una cierta cantidad de información que es procesada por las neuronas, entrando desde las dendritas que son los elementos de entrada de la neurona, pasa por el soma, órgano para llevar a cabo el procesamiento de la información nerviosa, hasta llegar al axón, elemento de salida (**Schmidhuber, 2015**). En la Figura 2.1 se puede visualizar la conexión de una neurona presináptica y una postsináptica, con los elementos de entrada y salida y sus respectivos pesos.

Las redes neuronales artificiales se han estructurado de manera similar para llevar a cabo la clasificación; el conjunto de neuronas o perceptrones agrupados en capa forman la red neuronal. Cada una de las neuronas posee un peso, al tener un dato de entrada lo evalúa y modifica este valor al salir, para que siga su camino a través de la red y obtener una salida final la cual se traduce como la predicción calculada por la red. Para obtener resultados correctos es necesario entrenarla, a través de datos que entran modificando la red de acuerdo a su contribución y margen de error; lo cual es denominado backpropagation.

La topología de una red neuronal es la manera en que se relacionan las neuronas que la constituyen. Estas se conectan por medio de estructuras sinápticas que determinan el comportamiento de la red, es la forma de comunicación que se tiene entre neuronas y la dirección en que viajan van de las neuronas presinápticas a las postsinápticas. Estas, como se ha mencionado forman estructuras denominadas capas, de las que se pueden definir tres tipos:

- Capa de entrada: Reciben los datos procedentes del entorno
- Capa de salida: neuronas que proporcionan respuesta de la red
- Capa oculta: las que no tienen una conexión directa con el entorno.

Una red neuronal está compuesta por los siguientes elementos:

- Un conjunto de entradas que está definido por x_j , con pesos sinápticos asociados a sus entradas.
- Pesos sinápticos denotados por w_{ij} , donde j es la neurona presináptica e i la neurona postsináptica.
- Una función de propagación, es el valor potencial postsináptico h_i , el cual es definido por el conjunto de entradas y el valor de los pesos sinápticos.

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) \quad (2.1)$$

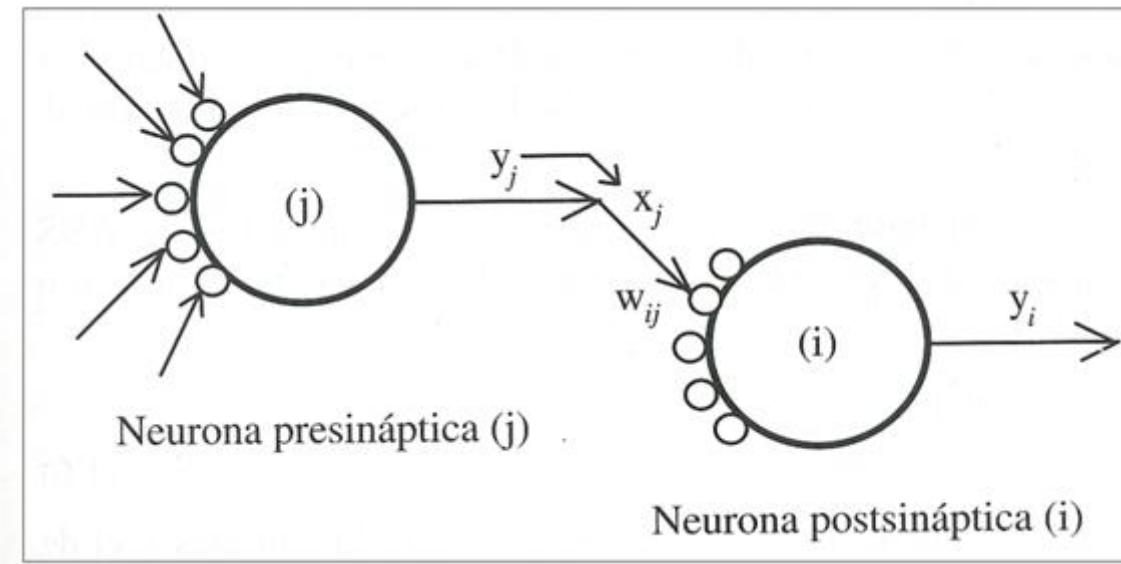


Figura 2.1: Conexión entre una neurona presináptica y una postsináptica

La regla más común es combinar linealmente las entradas y pesos sinápticos de esta manera:

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{i=1}^n x_j w_{ij} \quad (2.2)$$

A este conjunto de pesos es común que se le añada un parámetro adicional θ_i para restarle potencia postsináptica.

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{i=1}^n x_j w_{ij} - \theta_i \quad (2.3)$$

Si iniciamos i y j en 0, y $w_{i0} = \theta_i$ y $x_0 = -1$. La regla de propagación se puede escribir como:

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{i=1}^n x_j w_{ij} = \sum_{i=1}^n x_j w_{ij} - \theta_i \quad (2.4)$$

La función de activación es una representación de la salida en la neurona que a su vez es el comportamiento de esta, la cual se muestra a continuación:

$$y_i = f_i(h_i) = f_i \left(\sum_{i=1}^n x_j w_{ij} \right) \quad (2.5)$$

Se debe determinar la función de activación adecuada, entre las mas habituales está la sigmoidea o tangente hiperbólica.

Entrenamiento.

El objetivo del entrenamiento consiste en asignarle valores al vector de pesos, para que al evaluar los ejemplos el error sea mínimo, sin perder de vista que se debe mantener un equilibrio entre la generalización y la preparación especializada para evitar perder información. Después se procede a probar con valores nuevos, de esta forma se puede observar su comportamiento con nuevas entradas (**Ferreira, 2016**).

Para el entrenamiento de un perceptrón se inicializa sus pesos aleatoriamente para que estos se vayan actualizando en cada iteración:

$$w_i \leftarrow w_i + \Delta w_i \quad (2.6)$$

$$\Delta w_i = n(t - o)x_i \quad (2.7)$$

Dónde:

- n toma el valor del ritmo de aprendizaje, por lo que si se asigna muy pequeño este tardará mucho más en ser entrenada
- t es la salida correcta
- o la salida obtenida por la red

Si un perceptrón arroja una respuesta correcta los pesos se mantendrán pero si es errónea estos se actualizarán.

Con la regla anterior los pesos se ajustarán hasta que todos los elementos queden bien clasificados, los cuales también dependen de la función gradiente descendente de la función de error, que evalúa el valor de variación del vector de pesos haciendo que el error disminuya.

La función de error mide la diferencia entre la salida de la red y la salida correcta

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 \quad (2.8)$$

Para realizar las iteraciones se tienen que determinar una condición de paro, la cual se puede dar bajo diferentes situaciones:

- Se ha llegado a un valor que se considere suficientemente pequeño.
- Sí se llega a un número máximo de iteraciones, en el cual se define como límite para el entrenamiento.
- Si el error es igual a cero.
- El valor del error ya no disminuye, aun incrementando el número de veces el entrenamiento.

2.1.2. Support Vector Machines

Las máquinas de soporte vectorial (SVM por sus siglas en inglés) es un sistema de aprendizaje automático, el cual fue creado en los años 90 por Vapnik para darle solución a problemas de clasificación binaria (**Vapnik, 1995**) , pero con el tiempo su aplicación se extendió hasta clasificación múltiple y regresión siendo de gran utilidad para datos complejos, ruidosos y escasos ya que evalúa a los que proporcionan mayor información sin depender de mínimos locales. El éxito de esta herramienta está fundamentada en el concepto de minimización de riesgo estructural, lo que disminuye el error de clasificación para nuevos datos (**Cuevas, 2010**).

La clasificación de las SVM depende del entrenamiento de datos o patrones en una superficie de decisión óptima representada por un hiperplano el cual se obtiene con la máxima distancia entre él y los datos más cercanos los cuales representan a los vectores de soporte, ahora bien, las SVM evalúan la distribución de los datos para identificar su linealidad. En caso de no ser linealmente separable se usa un kernel, que permite convertir un problema no lineal a uno de clasificación lineal.

Aprendizaje estadístico.

La clasificación depende de los datos de entrenamiento o patrones, como se ha mencionado anteriormente, esta se hace a partir de una superficie de decisión óptima representada por un hiperplano el cual podemos definir como un subespacio plano afín al espacio de p-dimensional, con dimensiones p-1 (**Ostaszewski, 1990**). Esto es, que si tenemos un espacio de dos dimensiones nuestro hiperplano sería una recta que no tendría que pasar por el centro, así los patrones que tienen la forma $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ donde $x \in R^N$ pertenecientes a una clase con la etiqueta $y_i \in \{1, -1\}$, podrían ser ordenados de un lado u otro del hiperplano.

El hiperplano óptimo tiene la forma siguiente:

$$w \cdot z + b = 0 \quad (2.9)$$

Donde w es el vector perpendicular al plano y b es el coeficiente de intersección

Para un caso linealmente separable, se cumple la condición de clasificación siguiente:

- Para las muestras positivas serán mayor a 1

$$w \cdot z + b \geq 1 \quad (2.10)$$

- Las muestras negativas, menor a -1

$$w \cdot z + b \leq -1 \quad (2.11)$$

El hiperplano solución, es la recta media y paralela a los hiperplanos formados por los vectores de soporte, dados por las siguientes funciones:

$$h1 \rightarrow w \cdot z + b = 1 \quad (2.12)$$

$$h1 \rightarrow w \cdot z + b = -1 \quad (2.13)$$

El margen delimitado por los hiperplanos formados por los vectores de soporte se puede calcular a través de la distancia entre ellos, con la siguiente expresión:

$$d(x, w, b) = \frac{1}{|w|} \quad (2.14)$$

Para maximizar la distancia de este margen, va a ser equivalente a obtener el mínimo del problema inverso

$$\min \varphi(w) = \frac{\|w\|^2}{2} \quad (2.15)$$

Dado

$$y_i [(w \cdot x_i) + b] \geq 1 \quad (2.16)$$

Para maximizar la mínima distancia limitada por los valores de entrada ya mencionados, se hace uso de un lagrangiano.

$$L(w, b, a) = \frac{\|w\|^2}{2} - \sum_{i=1}^N \alpha [y_i [(w \cdot x_i) + b] - 1] \quad (2.17)$$

Para la minimización depende de los valores de la distancia, y para la maximización depende de los multiplicadores α

El clasificador final se puede definir directamente de los multiplicadores del hiperplano $w \cdot x$ o a partir de los multiplicadores del lagrangiano, quedando de la siguiente manera.

$$f(x) = \operatorname{sgn}(w \cdot x + b) = \operatorname{sgn} \left(\sum y_i \alpha_i (x_i \cdot x) + b_0 \right) \quad (2.18)$$

Donde el signo resultante indicara a que clase pertenece un determinado dato.

Caso no linealmente separable

Los casos de estudio en el mundo real difícilmente se presentan en dos dimensiones, y pueden tener más características que eviten una separación lineal adecuada como la complejidad en la separación de clases, más de dos categorías para su clasificación entre otras. Por ello se da como solución mapear o representar los datos en un plano dimensional mayor, a través de un kernel, que está dado por la siguiente función.

$$K = R^N \times R^N \rightarrow R \quad (2.19)$$

$$K : (x_i x_j) \rightarrow \varphi(x_i) \cdot \varphi(x_j) \quad (2.20)$$

Lo anterior es el producto punto punto en un espacio de dimensiones indistinto. Para encontrar el hiperplano óptimo de un función no lineal , se hace de la siguiente manera:

$$\text{Max}\theta(\alpha) = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i x_j) \quad (2.21)$$

Y la función de decisión está dada por:

$$f(x) = \text{sgn}(w \cdot x + b) = \text{sgn} \left(\sum y_i \alpha_i K(x_i x) + b_0 \right) \quad (2.22)$$

2.2. Técnicas de Evaluación de desempeño

Para evaluar los clasificadores se implementan diferentes métricas con el fin de compararlos y evaluar su desempeño para así poder anticipar el buen funcionamiento del sistema. Esto depende en gran medida de los datos utilizados para su entrenamiento (training set) y testeo o prueba (test set), seleccionándose los valores de cada grupo de forma aleatoria, siendo el primer grupo el que definirá el aprendizaje y el segundo su rendimiento. Estos dos conjuntos son seleccionados del total de datos que tenemos disponible, cometiendo un error si se utilizan los datos de entrenamiento para las pruebas pues se tendría un sobreajuste y la finalidad de evaluar la predicción se perdería pues los resultados serían casi perfectos.

La evaluación consiste en determinar la exactitud o minimizar el error de predicción, entre la distancia de los datos de prueba y la predicción del modelo bajo estudio. Idealmente debería ser exacta, pero la realmente se desea minimizar la distancia entre ellos.

La matriz de confusión es muy útil para evaluar las medidas de clasificación, donde se tienen las siguientes variables

- verdaderos positivos

- falsos positivos
- verdaderos negativos
- falsos negativos

Estas variables nos permiten obtener los positivos clasificados correctamente y la precisión, que son las predicciones positivas correctas. Es importante recalcar que hay variables que no se valoran por igual, pues en la mayoría de ocasiones es preferible obtener la mayor cantidad de valores aunque no lo sean, pues el costo de un falso negativo es mayor al de un falso positivo. En esta sección se describen técnicas de desempeño utilizadas frecuentemente.

2.2.1. Cross-validation

Evaluar las diferentes configuraciones de subconjuntos del total de datos disponible da pauta a que se genere un sobreajuste, pues se puede llegar a la selección óptima entre el conjunto de prueba y entrenamiento, filtrándose conocimiento de los datos de prueba para el ajuste, lo que evita que se informe correctamente sobre el rendimiento y la generalización.

Para prevenir este problema en el conjunto total se crea un subconjunto más, llamado conjunto de validación. El cual interviene después del conjunto de entrenamiento, para así realizar una evaluación, finalizando con el conjunto de prueba; el inconveniente que se genera con este procedimiento es la reducción drástica en la cantidad de muestras para realizar el aprendizaje.

Para evitar estos problemas debido a la limitación de datos, se utiliza este método, Cross-Validation (CV para abreviar) en el que aparentemente se descarta el subconjunto de validación, pero este se incluye en el subconjunto de entrenamiento (**Lang, 2014**). Esto es, el conjunto total de datos se divide en k subconjuntos, en el que se usa una partición para la prueba y $k-1$ para el entrenamiento, iterando k veces para cada uno de los conjuntos de prueba disponibles. La Figura 2.2 muestra un ejemplo con 4 iteraciones.

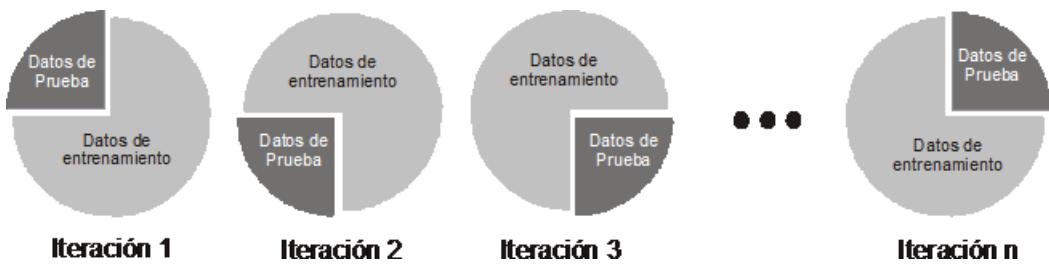


Figura 2.2: Validación cruzada de 4 iteraciones

El resultado del rendimiento se obtiene a través de la media aritmética de los valores obtenidos en cada iteración, siendo computacionalmente costoso, pero evita la perdida de datos ya que no se fija un conjunto de datos arbitrario.

2.2.2. F-Measure

La medida del rendimiento (F_1) para esta prueba es obtenida a través de la media armónica entre precisión y recuperación

$$F1 = \left(\frac{precision^{-1} \cdot recall^{-1}}{2} \right)^{-1} = \frac{2precision * recall}{precision + recall} \quad (2.23)$$

Donde la precisión es la predicción que son los valores positivos clasificados como positivos, y la exhaustividad es la capacidad del modelo para identificar correctamente las predicciones positivas; la identificación del mejor valor para F es cuando da resultado 1, lo que significa que la tasa de aciertos para la precisión y exhaustividad son perfectas; 0 para la peor clasificación (Beitzel, 2006).

Precisión

Como se ha mencionado, con esta métrica se puede identificar los valores positivos clasificados correctamente:

$$precision = \frac{TP}{TP + FP} \quad (2.24)$$

Exhaustividad.

Este valor es definido para la cantidad de verdaderos positivos que es capaz de identificar el modelo.

$$\text{recall} = \frac{TP}{TP + FN} \quad (2.25)$$

2.2.3. Área ROC

Con el fin de evaluar el rendimiento de un clasificador se utiliza la métrica ROC (Receiver Operating Characteristic, por sus siglas en inglés), el cual representa la exactitud diagnóstica a través de la sensibilidad, que es la fracción de valores positivos clasificados correctamente (FVP) con respecto a la especificidad que es la probabilidad de clasificar los positivos de manera errónea, denominado como la fracción de falsos positivos (FFP).

El inconveniente que presenta esta evaluación es que es útil para la clasificación dicotómica, quedando excluida la alta gama de resultados que se encuentran en una escala nóminal, en la que se inclina por valores cualitativos que cuantitativos. La solución cae en crear nuevas secciones de decisión o cortes.

Esta prueba se utiliza para determinar el mejor punto de corte en la curva ROC, el cual contiene el mayor número de verdaderos positivos que falsos positivos, el cual se calcula gráficamente, al proyectar una línea recta del punto (0,1) intersectando de manera perpendicular en un punto a la curva ROC, la cual representa la mínima distancia, dando como resultado mayor área bajo la curva, por ende mayor poder discriminativo (**Zou, 2007**). El comportamiento de las pruebas depende de este corte, si este se desplaza a la derecha (Clase Negativa) disminuye la tasa de falsos positivos y aumenta la tasa de falsos negativos. Inversamente si se desplaza a la izquierda (Clase Positiva) aumenta la tasa de falsos positivos pero disminuye la tasa de falsos negativos.

Si la prueba fuera perfecta, es decir que no exista solapamiento entre clases, la curva solo tendría un punto (0,1). De lo contrario, si la prueba fuera mala la curva sería una diagonal de (0,0) a (1,1).

La Figura 2.3 muestra un ejemplo de similitud que puede existir entre dos clases.

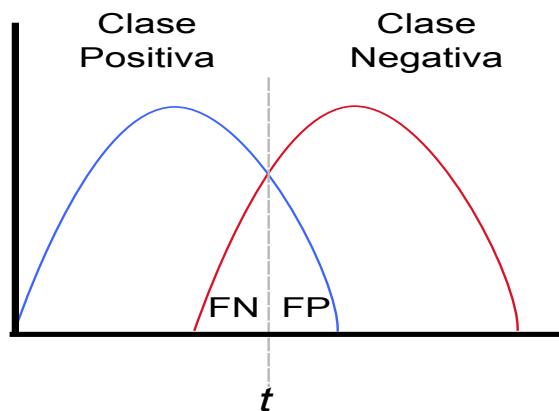


Figura 2.3: Representación de similitud entre dos clases. El punto de corte t determina el comportamiento del clasificador.

Capítulo 3

Técnicas de segmentación y extracción de características

3.1. Técnicas de segmentación

La segmentación de imágenes es una parte importante del procesamiento, ya que separa los objetos de interés del resto basándose en las intensidades o niveles de gris distribuidos en la imagen(**Morris, 2004**). La relevancia de este proceso se centra en poder definir las áreas más significativas de la imagen, las que proporcionan mayor información, pues de esto dependerá el éxito de la extracción de características, ya que cada pixel aporta información significativa así como cambios en los pixeles y niveles de luminosidad, los cuales afectan al procesamiento.

Una imagen se puede definir como una función bidimensional $f(x, y)$, donde x e y es la ubicación espacial del píxel y f representa el nivel de gris en ese punto. Con este proceso se pueden identificar bordes, regiones, líneas o curvas, ya que hay zonas con similitudes de color, textura, etc. Las cuales frecuentemente representan un objeto distinto.

En imágenes monocromáticas, la segmentación se determina a partir de sus propiedades en los niveles de gris que son discontinuidad, donde se divide la imagen de acuerdo a los cambios bruscos de nivel de gris, siendo útil esta para detectar puntos, líneas o bordes (**Linda, 2001**). Y la similitud, que se basa en la regularidad de los píxeles en una región

utilizando umbralización o prueba de regiones

A continuación se describen algunas técnicas de segmentación propuestas para el sistema.

3.1.1. Frontera Global

Es una técnica de umbralización o thresholding que permite extraer una imagen compuesta de un objeto claro con fondo oscuro o viceversa, identificando la imagen a través de un valor umbral de intensidad T , el cual puede ser obtenido bajo diferentes criterios ya sea por selección del usuario o bien , de forma automática, en esta última llevando a cabo dos técnicas :

1. Sumar el valor mínimo y máximo en la imagen y dividirlo entre dos

$$th = (\max(I(x, y)) + \min(I(X, Y)))/2 \quad (3.1)$$

2. Otro método es a través de la normalización del histograma (**Gonzalez, 2002**); en el que se obtienen las probabilidades de existencia en cada pixel, la información proporcionada por este método es la distribución de grises lo que va a permitir transformar una imagen en escala de grises en binaria de acuerdo a la siguiente ecuación, donde th es el valor del umbral:

$$G(x, y) = \begin{cases} 0 & 0 \leq f(x, y) \leq th \\ 1 & th < f(x, y) \leq 255 \end{cases} \quad (3.2)$$

3.1.2. Segmentación adaptativa

Debido a los cambios en los niveles de gris no siempre es recomendable optar por un umbral global, pues los cambios en la iluminación desfavorecen para separar adecuadamente el objeto de interés en la imagen (**Sonka, 1998**). La alternativa a esto es subdividir una imagen y encontrar el valor umbral óptimo para cada una de los componentes, devolviendo una imagen binarizada. El procedimiento es el siguiente; Se comienza seleccionando

una frontera inicial cualquiera $th(i)$, donde $i = 1$. Permitiendo en un principio dividir la imagen en dos grupos que son los pixeles mayores o igual a th y los menores a th , a continuación se muestra el agrupamiento de los pixeles:

$$G_1 = 0, \dots, th \quad (3.3)$$

$$G_2 = th + 1, th + 2, \dots, L \quad (3.4)$$

Donde L es la intensidad máxima en la imagen.

Se obtiene la media de cada grupo:

$$\mu_1 = \sum_{i=0}^{th} \frac{G_1(i)}{th} \quad (3.5)$$

$$\mu_2 = \sum_{i=th+1}^L \frac{G_2(i)}{L} \quad (3.6)$$

Esto nos ayuda a actualizar la frontera:

$$th(i) = \frac{\mu_1 + \mu_2}{2} \quad (3.7)$$

La condición de paro consta en que la diferencia de los umbrales en las últimas dos iteraciones no haya variado, de lo contrario se siguen ejecutando todos los pasos.

3.1.3. Segmentación por clustering

La segmentación por clustering es una técnica de agrupación automática basada en las distancias, segmentando el fondo de la imagen que vamos a utilizar (**Hartigan, 1975**). El algoritmo se presenta a continuación:

1. Se inicia seleccionando el número de grupos k , en los que se escogerá $k - puntos$ que son los centroides.
2. Cada punto del grupo es asignado al centroide más cercano, calculado a través de la distancia euclíadiana.

3. Los centroides se actualizan calculando la media de los puntos obtenidos para cada centroide.

Los pasos 2 y 3 se repiten hasta que los centroides no se muevan.

3.1.4. Segmentación Otsu

Esta es una técnica de umbralización basada en la varianza o dispersión de los valores, buscando que esta sea mínima dentro de las clases pero maximizarla en las que son diferentes.

Se tiene que calcular el histograma normalizado de la imagen, donde:

- Una imagen de $M \times N$ se define como una función $f(x, y)$.
- Niveles de gris en la imagen $0, 1, 2, \dots, L$
- n_i =número de píxeles con intensidad i
- $p_i = \frac{n_i}{MN}$, es la probabilidad de ocurrencia del nivel en la intensidad i .

Calcular las sumas acumulativas

$$w_1(k) = \sum_{i=0}^k p_i \quad (3.8)$$

Si se fija un valor umbral T en la intensidad $K(T = K)$, entonces se dividiría el conjunto de píxeles en dos clases

- c_1 = píxeles con intensidad de $[0, k]$
- c_2 = píxeles con intensidad de $[k+1, L]$

$$w_1(t) = \sum_{z=1}^T p_i \quad (3.9)$$

$$w_2(t) = \sum_{z=T+1}^L p_i \quad (3.10)$$

Calcular las medias acumulativas y la varianza entre clases.

$$\mu_1(t) = \sum_{i=1}^T \frac{ip(i)}{w_1(t)} \quad (3.11)$$

$$\mu_2(t) = \sum_{i=1}^T \frac{ip(i)}{w_2(t)} \quad (3.12)$$

$$\sigma_1^2(t) = \sum_{i=T+1}^T (i - \mu_2(t))^2 \frac{p(i)}{w_1(t)} \quad (3.13)$$

$$\sigma_2^2(t) = \sum_{i=T+1}^T (i - \mu_2(t))^2 \frac{p(i)}{w_2(t)} \quad (3.14)$$

Finalmente la varianza entre clases, la cual se define como:

$$\sigma_w^2 = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \quad (3.15)$$

Se selecciona el valor umbral T correspondiente a la varianza entre clases dada por la ecuación.

3.1.5. Técnicas de dilatación y erosión

La imagen A aún puede contener píxeles que pertenecen al fondo, para poder eliminarlos sin alterar los caracteres, la imagen es erosionada:

$$A_e = A \ominus B = (x', y') \mid (x' + i, y' + j) \in P_A, (i, j) \in P_B \quad (3.16)$$

Donde A es la imagen original, A_e es la imagen erosionada y \ominus denota el operador de erosión empleando la mascara B.

Ahora bien la operación morfológica de dilatación sirve para restaurar posibles caracteres rotos o incompletos, esto es posible realizarlo mediante el siguiente operador:

$$A_d = A \oplus B = (x', y') = (x + i, y + j) \mid (x', y') \in P_A, (i, j) \in P_B \quad (3.17)$$

3.1.6. Técnicas para eliminar ruido en imágenes binarias

Entre las operaciones morfológicas existen técnicas que sirven para eliminar el ruido en imágenes. Las siguientes ecuaciones permiten realizar este trabajo.

1. Quitar ruido (puntos negros) zonas blancas.

$$P_0 = P + (b \cdot g \cdot (d + e) + (d \cdot e \cdot (b + g))) \quad (3.18)$$

2. Quitar ruido (puntos blancos) zonas negras.

$$P_0 = P + (((a + b + d) \cdot (e + g + h)) + ((b + c + e) \cdot (d + f + g))) \quad (3.19)$$

3. Reconstruir esquinas de figuras blancas.

- Superior derecha

$$P_0 = \bar{P} \cdot d \cdot f \cdot g \cdot \overline{(a + b + c + e + h)} + P \quad (3.20)$$

- Inferior derecha.

$$P_0 = \bar{P} \cdot a \cdot b \cdot d \cdot \overline{(e + c + f + g + h)} + P \quad (3.21)$$

- Superior izquierda.

$$P_0 = \bar{P} \cdot e \cdot g \cdot h \cdot \overline{a + b + c + d + f} + P \quad (3.22)$$

- Inferior izquierda.

$$P_0 = \bar{P} \cdot b \cdot c \cdot e \cdot \overline{a + d + f + g + h} + P \quad (3.23)$$

4. Reconstruir esquinas de figura negra

- Superior derecha

$$P_0 = \bar{P} \cdot (d + f + g) + \overline{a \cdot b \cdot c \cdot e \cdot h \cdot P} \quad (3.24)$$

- Inferior derecha.

$$P_0 = \bar{P} \cdot (a + b + d) + \overline{c \cdot e \cdot f \cdot g \cdot h \cdot P} \quad (3.25)$$

- Superior izquierda.

$$P_0 = \bar{P} \cdot (e + g + h) + \overline{a \cdot b \cdot c \cdot d \cdot f \cdot P} \quad (3.26)$$

- Inferior izquierda.

$$P_0 = \bar{P} \cdot (b + c + e) + \overline{a \cdot d \cdot f \cdot g \cdot h \cdot P} \quad (3.27)$$

3.2. Técnicas de Extracción de características

La utilidad de imágenes en el campo de visión artificial ha permitido abarcar un número extenso de aplicaciones, donde se utilizan técnicas que preparan los datos de una imagen para que resulte informativa, eliminando los grandes volúmenes para ser procesados, pues pueden ser redundantes. A esta selección de características se le denomina extracción de características, de las cuales se espera que haya información relevante, lo que permitirá facilitar el aprendizaje, incluso mejorando la interpretación humana. A continuación se describen algunas técnicas de extracción de características comúnmente utilizadas.

3.2.1. Características texturales de Haralick

Las características de textura proporcionan información sobre la disposición espacial de colores o intensidades en la imagen. Los algoritmos de extracción de características de textura buscan patrones repetitivos básicos con estructuras periódicas o aleatorias en las

imágenes. Estas estructuras se obtienen por propiedades de la imagen, tales como la rugosidad, granulación, finura, suavidad, etc (**Presutti, 2004**). La textura repite un patrón a lo largo de una superficie, debido a que las texturas son invariables a los desplazamientos, esto explica por qué la percepción visual de una textura es independiente de la posición. En este documento se utilizaron las características de textura de Haralick, histograma de gradientes ordenados y patrones binarios locales. Estas características consideran la distribución de los valores de intensidad en la región, al obtener la media y el rango de las siguientes variables: media, mediana, varianza, suavidad, sesgo, curtosis, correlación, energía o entropía, contraste y homogeneidad (**Haralick, 1973**).

Los descriptores de Haralick utilizados son mostrados a continuación:

$$f_1 = \sum_i \sum_j [p(i, j)^2] \quad (3.28)$$

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \begin{array}{l} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \\ |i - j| = n \end{array} \right\} \quad (3.29)$$

$$f_3 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [ijp(i, j) - \mu_x \mu_y]}{\sigma_x \sigma_y} \quad (3.30)$$

$$f_4 = \sum_i \sum_j (i - \mu_x)^2 p(i, j) \quad (3.31)$$

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \quad (3.32)$$

$$f_6 = \sum_{i=2}^{2N_g} iP_{x+y}(i) \quad (3.33)$$

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 P_{x+y}(i) \quad (3.34)$$

$$f_8 = - \sum_{i=2}^{2N_g} P_{x+y}(i) \log\{P_{x+y}(i)\} \quad (3.35)$$

$$f_9 = - \sum_i \sum_j p(i,j) \log\{p(i,j)\} \quad (3.36)$$

$$f_{10} = \sum_{i=0}^{N_g-1} (i - f_8)^2 P_{x-y}(i) \quad (3.37)$$

$$f_{11} = - \sum_{i=0}^{N_g-1} P_{x-y}(i) \log\{P_{x-y}(i)\} \quad (3.38)$$

$$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (3.39)$$

$$f_{13} = (1 - e^{[-2(HXY2 - HXY)]})^{\frac{1}{2}} \quad (3.40)$$

$$f_{14} = (\text{Segundo eigenvalor mas grande } Q)^{\frac{1}{2}} \quad (3.41)$$

Donde:

- $p(i,j)$ = Define el valor de la matriz de coocurrencia en la posición (i,j) .
- N_g = Representa los niveles de gris usados en la región segmentada de la imagen.
- $p_x(i)$ = Es la i-ésima entrada en la matriz de probabilidad de coocurrencia que es obtenida sumando columnas en $p(i,j)$.
- R = Número total de elementos en la matriz de coocurrencia.
- μ_x, μ_y = Promedio de P_x y P_y
- σ_x, σ_y = Desviación estándar de P_x y P_y
- H_x, H_y = Entropía de P_x y P_y

3.2.2. Patrones binarios locales

Es una técnica descriptiva simple pero altamente efectiva para la clasificación de objetos dentro de la visión por computadoras que filtra los píxeles adyacentes mediante consideraciones determinadas y obtiene un valor binario representativo. Codifica la relación del píxel central con la intensidad de color de los píxeles aledaños. Debido a su elevada capacidad discriminatoria, constituye una aproximación usual para la solución de multitud de problemas (**Matas, 2010**). Probablemente una de sus características más importantes es la robustez de su invariante frente a cambios lumínicos.

Local Binary Patterns (LBP) es un método muy eficiente para el reconocimiento de texturas, el cual se caracteriza por su simplicidad computacional y alto poder discriminativo.

El operador de análisis de textura de LBP tiene como idea básica que las texturas de superficie bidimensional se puedan describir mediante dos medidas complementarias:

- Patrones espaciales locales.
- Contraste de escala de grises.

El operador consiste en comparar cada píxel P de una imagen con sus ocho vecinos que lo rodean, determinando si cada píxel en esa vecindad tiene una intensidad mayor o menor al píxel P. En la Figura 3.1 se puede apreciar un ejemplo de los píxeles vecinos de un píxel P.

7	5	4
3	7	9
2	8	5

Figura 3.1: Muestra de píxel P (central) de una imagen I

En caso de ser mayor se le asigna el valor de 1, de lo contrario un 0. Tomando como píxel inicial cualquier de la vecindad, sin perder el orden para ninguno píxel de la imagen I.

En el ejemplo de la Figura 3.2 se toma como primer píxel el superior izquierdo, circulando en el sentido de las manecillas del reloj.

1	0	0
0		1
0	0	0

Figura 3.2: Equivalente binario LBP

Esto da como resultado una cadena binaria, conteniendo en total 8 bits que indican la distribución de las intensidades a su alrededor sustituyendo el píxel P por el resultado de la cadena binaria convertida a decimal, en la Figura 3.3 se observa un ejemplo para el cálculo de la cadena binaria el cual es resultado de la comparación dada en el ejemplo de la Figura 3.2, finalmente el pixel P debe ser sustituido por el resultado de la cadena binaria en su equivalente decimal, lo que da suficiente información de los píxeles con relación a sus vecinos.

Como ya se ha mencionado, este algoritmo tiene que ser repetido para cada píxel de la imagen, presentando una inconsistencia en los bordes, pues no se podrían completar los vecinos, teniendo dos soluciones para esto:

- Eliminar los bordes y no calcularlos
- Añadir una capa de píxeles en los bordes, tipo espejo.

Obtenido el LBP para cada píxel, se procede al cálculo del histograma el cual es obtenido a través de las frecuencias de cada intensidad de la imagen I. Para mejorar el resultado es recomendable dividir la imagen en bloques para que este sea mas discriminativo, del cual tendremos 256 entradas, 0-255.

0	0	0	0	1	0	0	1	=9
0	0	0	0	8	0	0	1	

Figura 3.3: Cadena binaria, resultado de la comparación de los vecinos del píxel P

3.3. Histograma de Gradientes Ordenados

HOG(Histograma de Gradientes Ordenados) es un descriptor de características que tiene como objetivo generalizar un objeto de tal forma que se produzca lo más parecido posible al descriptor, cuando es visto bajo diferentes condiciones, así la tarea de clasificación será mucho más fácil.

Esta técnica utiliza la información de los gradientes como dato básico (**Carrillo, 2017**). El cual es definido como el cambio de intensidad en una imagen en una cierta dirección, calculado para cada píxel dado por los siguientes valores:

- Dirección del gradiente donde el cambio de intensidad es máximo.
- Magnitud del cambio que es afectado por la cantidad de intensidad.

En la Figura 3.4 muestra un ejemplo donde en el inciso a) se puede observar que el gradiente está en la misma dirección que la imagen c) pero con menor magnitud ya que es menor el cambio de intensidad; en la imagen b) cambia la dirección y mantiene las misma magnitud que c).

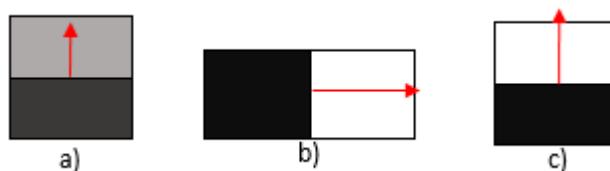


Figura 3.4: Dirección y magnitud de gradientes

Para el cálculo del gradiente en un píxel P, se realiza con el cambio de intensidad en los píxeles vecinos de manera horizontal y vertical, como se muestra en las ecuaciones siguientes:

$$d_x = I(x + 1, y) - I(x - 1, y) \quad (3.42)$$

$$d_y = I(x, y + 1) - I(x, y - 1) \quad (3.43)$$

A partir de estas diferencias se puede calcular la orientación y magnitud del gradiente, ya que esta se puede trasladar al eje de las coordenadas, como se puede apreciar en la Figura 3.5.

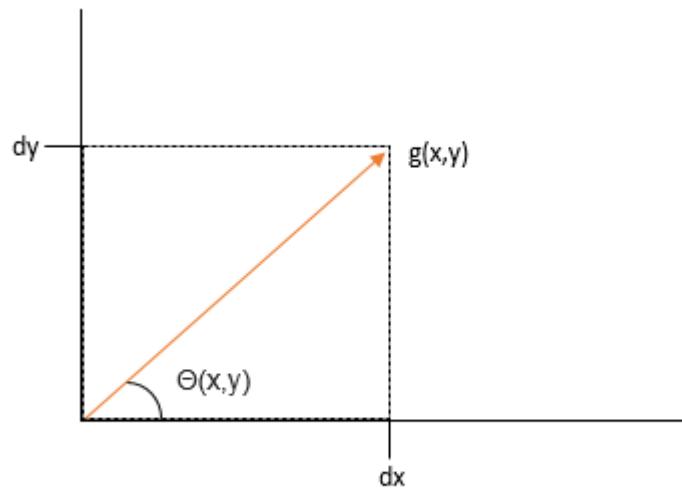


Figura 3.5: Gradiente trasladado al eje de las coordenadas.

Para la obtención de la orientación y magnitud se utilizan las siguientes fórmulas.

$$\theta(x, y) = \tan\left(\frac{d_y}{d_x}\right) \quad (3.44)$$

$$g(x, y) = \sqrt{d_x^2 + d_y^2} \quad (3.45)$$

El cálculo de los gradientes para imágenes de color se tiene que hacer para los tres canales (R,G,B) y al final seleccionar el canal con mayor magnitud.

Esta información es muy importante para conocer características del objeto, siendo que los valores altos en la magnitud representan una silueta o contorno y la orientación representa la forma o localización.

La información de los gradientes tiene que ser generalizada, lo que da como resultado un vector de características que es el descriptor global, el proceso es el siguiente:

- Dividir la imagen en un número fijo de celdas 8×8 ó 6×6 .
- Calcular los histogramas para cada una de las celdas.
- Combinar los histogramas de todas las celdas para obtener el vector de características.

Para la obtención de los histogramas en cada celda se tendrá que dividir el rango de orientaciones, en los que se asignará cada gradiente al rango correspondiente dependiendo de su orientación, esto nos permite obtener la frecuencia en cada uno de los intervalos para cada celda. El inconveniente para gradientes con orientaciones similares es que son asignados a intervalos muy diferentes, una de las formas de mitigar este problema es asignarlo a los dos rangos más cercanos.

Con el objetivo de conseguir la máxima invarianza de cambio en la imagen es conveniente normalizar los histogramas en los bloques de $b \times b$ celdas, a través del siguiente procedimiento:

- Concatenar los histogramas de ese bloque.
- Normalizar el vector utilizando la norma L2

$$v = \frac{v}{\sqrt{\|v\|}} \quad (3.46)$$

Esta normalización contribuye a disminuir las diferencias entre imágenes similares.

El descriptor final se define agrupando celdas para formar un bloque de 2x2 celdas.

Teniendo como parámetro los siguientes valores:

- Tamaño de celda.
- Número de intervalos del histograma.
- Número de celdas por bloque.

Lo anterior nos permite determinar la dimensión del vector final con la siguiente expresión.

$$N = NoBloques \times NoCeldasBloque \times NoIntervalosHistogramas \quad (3.47)$$

$$NoBloques = NoCeldas - NoCeldasBloque + 1 \quad (3.48)$$

3.3.1. Características cromáticas

Las características cromáticas proporcionan información de la intensidad de color de una región segmentada. Estas características se pueden calcular para cada canal de intensidad, por ejemplo, rojo, verde, azul, escala de grises, tono (Tono), saturación (Saturación) e intensidad (Valor), etc. Las características utilizadas fueron:

- Características de intensidad estándar.
- Desviación estándar de intensidad.
- Primera y segunda derivada en la región segmentada.
- Momentos de Hu con información.
- Características de Gabor basadas en funciones 2D de Gabor.

En los experimentos se obtuvieron 122 características para cada canal, pues estos se realizaron en RGB, utilizando así 366 características cromáticas.

El vector de características cromáticas X_c se puede definir por:

$$X_c = [x_1, x_2, \dots, x_{117}] \quad (3.49)$$

$$X_c = [x_{Re}, x_{RHu}, x_{Ge}, x_{GHu}, x_{Be}, x_{BHu}] \quad (3.50)$$

Dónde:

- x_{Re} , x_{Ge} , x_{Be} : representa las características de color elementales en los canales R, G y B respectivamente.
- x_{RHu} , x_{GHu} , x_{BHu} : representa los momentos de color invariables de Hu en los canales R, G y B respectivamente.

Se puede encontrar una descripción detallada sobre los momentos Hu con color en (Mingqiang, 2008).

Capítulo 4

Metodología

En este capítulo se escribe paso a paso la metodología utilizada para resolver el problema. La clasificación en un sistema de visión es considerado una parte importante en la identificación de objetos. La fiabilidad y eficiencia del sistema de reconocimiento depende directamente de cada uno de los pasos llevados a cabo, desde las técnicas de pre-procesamiento utilizadas, métodos de segmentación, características extraídas y clasificadores para el reconocimiento. En los experimentos llevados a cabo se trabajaron con diferentes técnicas de selección de características, así como diferentes clasificadores. La Figura 4.1 muestra la metodología en el desarrollo de los experimentos llevados a cabo en esta tesis. Cada uno de los procesos se detalla en este capítulo.

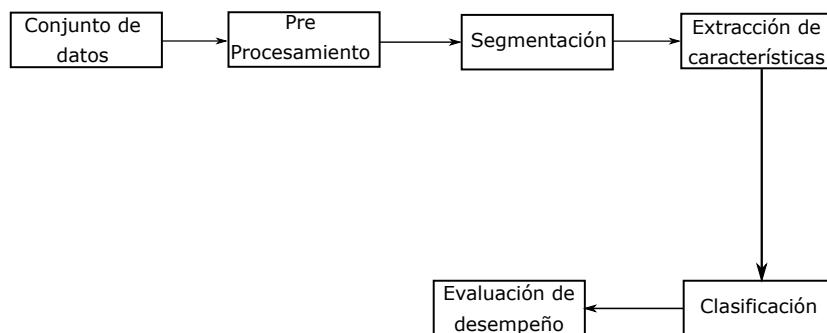


Figura 4.1: Pasos más usuales en métodos de selección de características por filtrado

4.1. Conjunto de datos utilizado

En los experimentos realizados, se utilizaron dos conjuntos de imágenes, el primer conjunto compuesto de tres clases de datos el cual fue obtenido de <https://www5.cs.fau.de/en/our-team/bo-ckruegler/automated-glaucoma-risk-index/>. De acuerdo a sus recopiladores el conjunto fue generado bajo condiciones específicas en un ambiente completamente controlado. La Figura 4.2 muestra el conjunto de datos utilizado. La fila a) muestra el conjunto de imágenes con problemas de retinopatía diabética, la fila b) muestra el conjunto de imágenes con padecimiento de glaucoma y la fila c) muestra el conjunto de imágenes saludables.

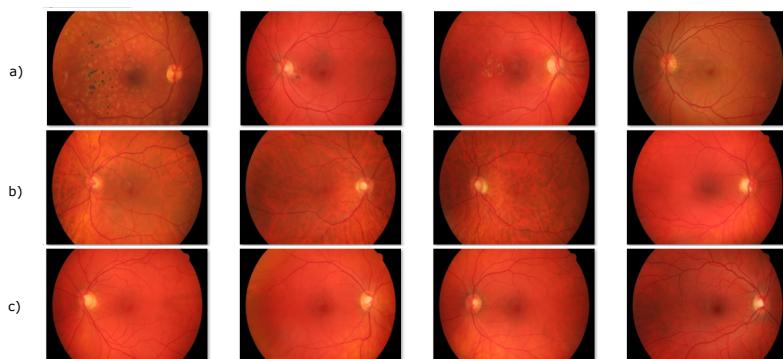


Figura 4.2: Conjunto de datos utilizado a) Retinopatía diabética b) Glaucoma c) Saludable

El segundo conjunto de datos fue obtenido de (**Diabetic Retinopathy Image Dataset, 2018**) . El cual está dividido en tres partes:

- A. Segmentación: Consiste de imágenes de fondo de ojo a color (81 imágenes divididas en conjuntos de entrenamiento y prueba). Imágenes de lesiones (microaneurismas, hemorragias, exudados duros y exudados blandos divididos en conjuntos de entrenamiento y prueba) y disco óptico (dividido en conjunto de entrenamiento y prueba)
- B. Clasificación de enfermedades: consiste en Imágenes de fondo de ojo a color originales, 516 imágenes divididas en conjuntos de entrenamiento (413 imágenes) y conjunto de prueba (103 imágenes). Etiquetas del grado de enfermedad de la retinopatía diabética y el edema macular diabético (dividido en conjuntos de entrenamiento y prueba)

C. Localización: consta de imágenes de fondo de ojo a color originales 516 imágenes divididas en conjuntos de entrenamiento (413 imágenes) y conjunto de prueba (103 imágenes). Etiquetas para la ubicación del centro del disco óptico (dividido en conjuntos de entrenamiento y prueba). Etiquetas para la ubicación del centro de Fovea (dividido en conjuntos de entrenamiento y prueba)

Se utilizó el inciso B pra realizar dos experimentos:

1. La primer prueba se realizó para la detección de retinopatía en cualquiera de sus grados. Se dividió el conjunto de datos con imágenes de fondo de ojo con algún grado de retinopatía (348) e imágenes de fondo ocular sano (148).
2. En la segunda se incluyeron imágenes de fondo ocular con grados de retinopatía (grado 1(25), grado 2(168), grado 3(93) y grado 4(62)) e imágenes de fondo ocular de personas sanas (168).

4.2. Pre procesamiento

Para el mejoramiento de la imagen, en el que se busca resaltar aspectos esenciales y eliminar los no deseados, se aplican diferentes técnicas que ayudan a estas tareas. En este caso se utilizaron las imágenes originales, en las que se aplicó un filtro Gaussiano de 3x3 para quitar algunas imperfecciones o diminutos puntos negros en las imágenes, lo que reduce las altas frecuencias. En los resultados se observó que no variaron de manera significativa, ya que las imágenes fueron tomadas en un ambiente totalmente controlado.

4.3. Segmentación

Una vez pre-procesadas las imágenes estas fueron segmentadas. En los experimentos realizados solo se utilizó la segmentación para eliminar los bordes y trabajar solo con la parte central. Debido a que el procedimiento es muy sencillo se optó por aplicar el método de Otsu.

La Figura 4.3 muestra el proceso de segmentación. La parte en negro es la parte que no se utilizó en el proceso de extracción de características, mientras que la parte en blanco representa a toda la parte utilizada en la extracción de características.

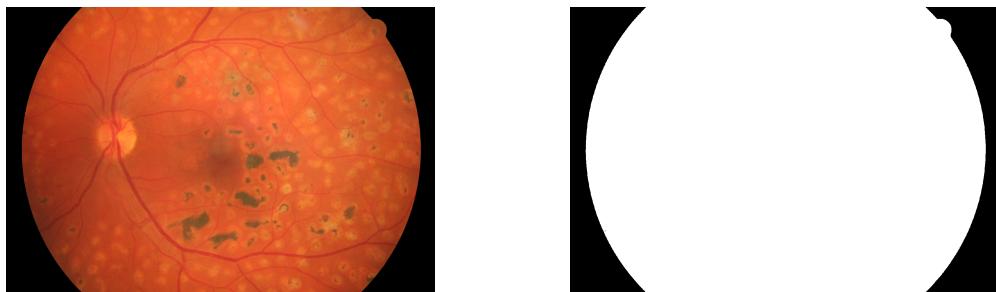


Figura 4.3: Segmentación de las imágenes con el método de Otsu

4.4. Extracción de características

Un vector de características en un conjunto de valores cuantificables que representa una imagen, este vector es calculado utilizando varias técnicas de extracción de características, las cuales se definen de convertir la imagen en una forma única, distintiva y compacta. Este proceso requiere que la imagen esté segmentada. La extracción de características es un proceso realizado después de la segmentación para obtener las propiedades más importantes de la imagen con el propósito de reconocer el objeto de la imagen. En general, las características se pueden clasificar en el tipo de componente (global o local) o en el tipo de características a extraer (texturales, geométricas o cromáticas). Las características locales son extraídas a partir de una cierta localidad de la imagen. Las características globales son obtenidas para toda la imagen.

En los experimentos realizados se trajeron dos tipos de características.

1. Texturales.
2. Cromáticas

Para las características cromáticas se trajeron:

- Momentos de Hu con intensidad para cada canal en RGB y HSV.

- Características HoG piramidales con intensidad.
- Características de Gabor.
- Media de intensidad para cada canal.
- Desviación estándar para cada canal.
- Contraste para cada canal.
- Gradiente en el borde para cada canal.
- Patrones binarios locales para cada canal.
- Intensidad de curtosis para cada canal.

Mientras que para las características texturales se trajeron:

- • Descriptores texturales de Haralick.

En total se trajeron 1992 características texturales y cromáticas.

4.4.1. Normalización de datos

En reconocimiento de patrones, la normalización puede ser realizada por diferentes razones. En los casos más simples, la normalización de las clasificaciones significa ajustar los valores medidos en diferentes escalas a una escala teóricamente común, a menudo antes de promediar. En casos más complicados, la normalización puede referirse a ajustes más sofisticados donde la intención es alinear todas las distribuciones de probabilidad de los valores ajustados.

La normalización se refiere a la creación de versiones de estadísticas modificadas, donde la intención es que estos valores normalizados permitan la comparación con otros valores correspondientes para diferentes conjuntos de datos de una manera que elimine los efectos de ciertas influencias generales.

En los experimentos llevados a cabo, cada conjunto de datos fue normalizado dentro de un determinado rango de valores para que los atributos sean comparables en algún

sentido. Debido a que los parámetros calculados se encuentran en rangos distintos, es necesario realizar una transformación de datos. Para ello se implementó la normalización estándar, si la media de la población y la desviación estándar son conocidas la puntuación estándar de un vector x es calculada como:

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$

Donde:

- μ : Es la media de la población .
- σ : Es la desviación estándar de la población.

El valor absoluto de z representa las distancia entre la puntuación bruta y la media de la población en unidades de desviación estándar, z es negativo cuando la puntuación bruta está por debajo de la media, positivo cuando está arriba de la media.

Calcular z usando esta fórmula requiere de la media y desviación estándar de la población, no de la muestra. Conocer la media y desviación estándar real de una población a menudo no es realista, excepto en casos como las pruebas estandarizadas, donde se mide a toda la población.

Cuando no se conoce la media o la desviación estándar de la población, la puntuación estándar se puede calcular utilizando la media de la muestra y la desviación estándar de la misma como estimaciones de los valores de la población. En estos casos, la puntuación z es:

$$z = \frac{x - \bar{x}}{S} \quad (4.2)$$

Donde:

- \bar{x} : es la media de la muestra.
- S : es la desviación estándar de la muestra.

Hay dos ventajas principales de tener un conjunto de datos normalizado:

1. Mayor consistencia. La información se almacena en un solo lugar, lo que reduce la posibilidad de inconsistencias.
2. Mapeo de objeto a datos más fácil. Los conjuntos de datos normalizados en general están más cercanos conceptualmente a los conjuntos orientados a objetos porque estos promueven una alta cohesión y un acoplamiento flexible entre clases dando como resultado soluciones similares.

La principal desventaja de la normalización es un rendimiento más lento.

4.5. Clasificación e identificación

En un sistema de reconocimiento de objetos, el clasificador juega un importante papel en el reconocimiento de objeto. La eficiencia y la precisión del sistema dependerán del clasificador utilizado. En los experimentos realizados se utilizaron varios clasificadores y se obtuvo un análisis comparativo de su desempeño. Los modelos de clasificación utilizados en esta tesis fueron: redes neuronales, SVM, árboles de decisión y Naive Bayes. Con el objetivo de validar los resultados se utilizó validación cruzada $k - fold$ con $k = 10$.

La validación cruzada o cross-validation (**Hagan, 1996**) es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utilizan en entornos donde el objetivo principal es la predicción y se quiere estimar que tan preciso es un modelo que se llevará a cabo en la práctica. Es una técnica muy utilizada en proyectos de inteligencia artificial para validar modelos generados. En la validación cruzada de k iteraciones o k -fold cross-validation los datos de muestra se dividen en k subconjuntos, uno de los subconjuntos se utilizan como datos de prueba y el resto ($k - 1$) como datos de entrenamiento, el proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de prueba. Finalmente se obtiene la media aritmética de los resultados de cada iteración para obtener un único resultado.

4.6. Análisis de desempeño

Es conveniente establecer una forma de comparar los resultados que se obtengan de la ejecución de los métodos. Supongamos que se realiza la consulta de una imagen sobre una base de datos, produciéndose varias imágenes como resultado. En una aplicación de esta clase, hay dos tipos de errores que nos interesa diferenciar:

1. Falsos positivos. Se trata de resultados que han sido devueltos incorrectamente, esto es, que se han reconocido como pertenecientes a la categoría consultada pero que no lo son.
2. Falsos negativos. Son imágenes que no son reconocidas como pertenecientes a la categoría consultada, o lo son con un bajo nivel de aceptación, cuando en realidad deberán ser aceptadas.

Para representar la variación de estos radios con diferentes ajustes de las técnicas de comparación, usaremos una métrica de comparación que se denomina área bajo la curva AUC-ROC (Receiver Operating Characteristic) que representa la taza de falsos positivos frente a los falsos negativos. En estas curvas, se ha de maximizar el área que cada curva deja por encima, minimizando así el error global cometido. También es importante conocer en qué punto de la curva intersecciona con la recta $f(x) = x$. Esta intersección significa que, para ese ajuste de parámetros concreto, el número de falsos positivos y negativos se iguala, lo que puede ser interesante. Se utilizará como medida de bondad alternativa el área sobre la curva ROC.

Capítulo 5

Resultados experimentales

En esta sección se muestran las pruebas realizadas en el conjunto de datos definido. Se realiza una comparativa entre los resultados obtenidos únicamente con las características extraídas mediante el método de HoG y las demás definidas en el capítulo 4. En esta tesis no se prevé el uso de algoritmos de selección de características para reducir el tiempo de entrenamiento y mejorar la precisión de clasificación. En su totalidad, las características extraídas son utilizadas para entrenar el clasificador. En los resultados experimentales se utilizaron 5 clasificadores para comparar los resultados obtenidos.

5.1. Desempeño de clasificadores utilizando Histograma de Gradientes Orientados

En esta sección se muestran los resultados de precisión obtenidos en el conjunto de datos utilizado. La Tabla 5.1 muestra los resultados de precisión utilizando solo histograma de Gradientes orientados. Las columnas de la Tabla muestran las precisiones obtenidas, donde tenemos las siguientes columnas en la Tabla: VP que representa el porcentaje de verdaderos positivos obtenidos para cada una de las clases, FP el porcentaje de falsos positivos, la precisión, AUC-ROC que representa el área bajo la curva, F-Measure y la clase sobre la que se obtuvieron los resultados mostrados.

Mientras que la primer columna contiene la técnica de clasificación utilizada. En los

resultados experimentales llevados a cabo, se hicieron experimentos con árboles de decisión, Redes Bayesianas, Redes Neuronales, Regresión logística y Support Vector Machines utilizando el algoritmo de Optimización mínima Secuencial (SMO).

La Tabla 5.1 y la Figura 5.1 muestran los resultados obtenidos utilizando Histogramas de Gradientes Orientados. Como lo muestran las Tablas referidas el mejor desempeño es obtenido utilizando redes neuronales, particularmente las clases 1 y 3 son clasificadas con la totalidad de verdaderos positivos, mientras que las clases 2 y 3 con 0 falsos positivos. Por otro lado, el clasificador que peor desempeño tuvo en promedio fue regresión logística, mientras que la peor clase clasificada fue la clase 2 utilizando clasificación Bayesiana.

Cuadro 5.1: Resultados utilizando HoG

	VP	FP	Precisión	AUC-ROC	F-Measure	Clase
Regresión logistica	0.600	0.167	0.643	0.667	0.621	1
	0.400	0.233	0.462	0.540	0.429	2
	0.733	0.233	0.611	0.818	0.667	3
Bayes	0.533	0.167	0.615	0.813	0.571	1
	0.533	0.400	0.400	0.693	0.457	2
	0.600	0.100	0.750	0.782	0.667	3
SMO	0.667	0.133	0.714	0.823	0.690	1
	0.333	0.167	0.500	0.644	0.400	2
	0.733	0.333	0.524	0.767	0.611	3
NN	1	0.033	0.938	1	0.968	1
	0.933	0	1	0.971	0.966	2
	1	0	1	1	1	3
DT	0.800	0.167	0.706	0.807	0.750	1
	0.467	0.167	0.583	0.626	0.519	2
	0.733	0.167	0.688	0.824	0.710	3

En cuanto a los resultados obtenidos en la métrica Precisión, es importante puntualizar que los resultados obtenidos con esta métrica son importantes cuando los conjuntos de datos son balanceados sin embargo, no nos dan ninguna medida de la sensitividad y especificidad del clasificador. La Tabla 5.1 muestra los resultados obtenidos con F-measure, que es otra métrica de desempeño utilizada en aprendizaje máquina.

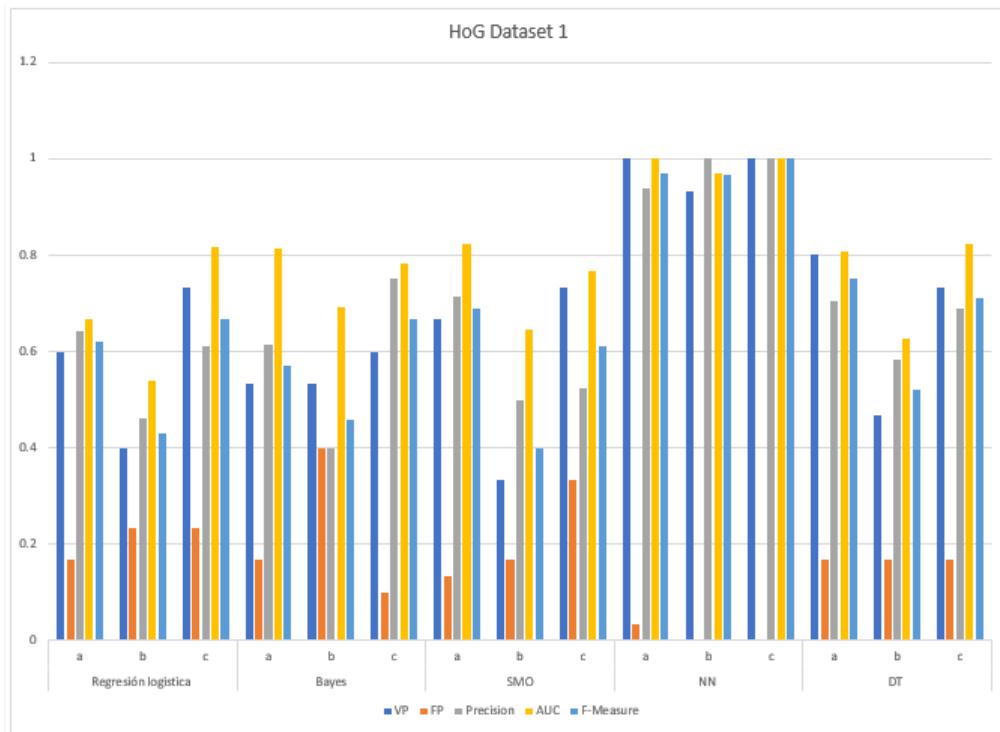


Figura 5.1: Gráfica de resultados usando HoG

Cuando se construye un modelo para un problema de clasificación, casi siempre se obtiene la precisión de ese modelo como el número de predicciones correctas de todas las predicciones realizadas. Esta precisión se podría interpretar como la exactitud de la clasificación. Sin embargo, la precisión de la clasificación por sí sola no suele ser suficiente información para tomar esta decisión. Para ello es necesario, además evaluar la solidez de modelo mediante el uso de validación cruzada a través de otros métodos de desempeño para datos no vistos, como F-measure también conocida como F-score, la cual se utiliza para medir la precisión de una prueba y equilibra el uso de la precisión y el recall para hacerlo. La métrica F-measure puede proporcionar una media más realista del rendimiento de una prueba utilizando la precisión como recall.

Por otro lado, el área bajo la curva muestra el rendimiento de un modelo de clasificación en todos sus umbrales. La curva ROC se desarrolló para medir la eficacia en la detección de objetos enemigos en campos de batalla mediante pantallas de radar. El análisis ROC se aplicó posteriormente en diferentes campos de investigación como medicina, radiología, psicología y otras áreas durante varias décadas y recientemente ha encontrado aplicación

en áreas de aprendizaje automático y minería de datos. El área bajo la curva está definida bajo dos parámetros: taza de verdaderos positivos (TPR) y taza de falsos positivos (FPR). Una curva ROC representa TPR frente a FPR en diferentes umbrales de clasificación. Reducir el umbral de clasificación clasifica más elementos como positivos, por lo que aumentarán tanto los falsos positivos como los verdaderos positivos.

5.2. Desempeño de clasificadores utilizando características de Haralick

En esta sección se discuten los resultados obtenidos utilizando solo características de Haralick. En la primer columna de la Tabla 5.2 define el clasificador utilizado. En los experimentos llevados a cabo se utilizaron los clasificadores Árboles de decisión, Redes Bayesianas, Redes Neuronales, Regresión logística y Support Vector Machines utilizando el algoritmo de Optimización mínima Secuencial (SMO).

La Tabla 5.2 y la Figura 5.2 muestran los resultados obtenidos extrayendo características texturales de Haralick de la región de los globulos oculares. Es posible ver tanto en la Tabla como en la gráfica que los resultados son más estables. Tanto la tasa de verdaderos positivos como la tasa de falsos positivos tienen un comportamiento similar en casi todos los clasificadores, sin embargo en este caso el clasificador con mejor desempeño son los árboles de decisión, mientras que el clasificador con peor desempeño es la regresión logística, esto al observar los resultados con la métrica de desempeño f-measure, la tasa de verdaderos positivos y la tasa de falsos positivos. Sin embargo, de forma general las características texturales en general permiten obtener mejores precisiones en cada una de las métricas en comparación con HoG.

Cuadro 5.2: Resultados utilizando características texturales de Haralick

	VP	FP	Precisión	AUC-ROC	F-Measure	Clase
Regresión logistica	0.867	0.100	0.813	0.971	0.839	1
	1	0.033	0.938	0.996	0.968	2
	0.733	0.067	0.846	0.931	0.786	3
Bayes	0.933	0.067	0.875	0.964	0.903	1
	1	0.067	0.882	0.967	0.938	2
	0.733	0.033	0.917	0.927	0.815	3
SMO	0.867	0.067	0.867	0.949	0.867	1
	0.933	0.000	1	0.992	0.966	2
	0.867	0.1	0.813	0.883	0.839	3
NN	0.867	0.033	0.929	0.984	0.897	1
	1	0.033	0.938	0.993	0.968	2
	0.867	0.067	0.867	0.960	0.867	3
Random Forrest	1	0.033	0.938	1	0.968	1
	1	0.000	1.000	1	1	2
	0.933	0.000	1	1	0.966	3

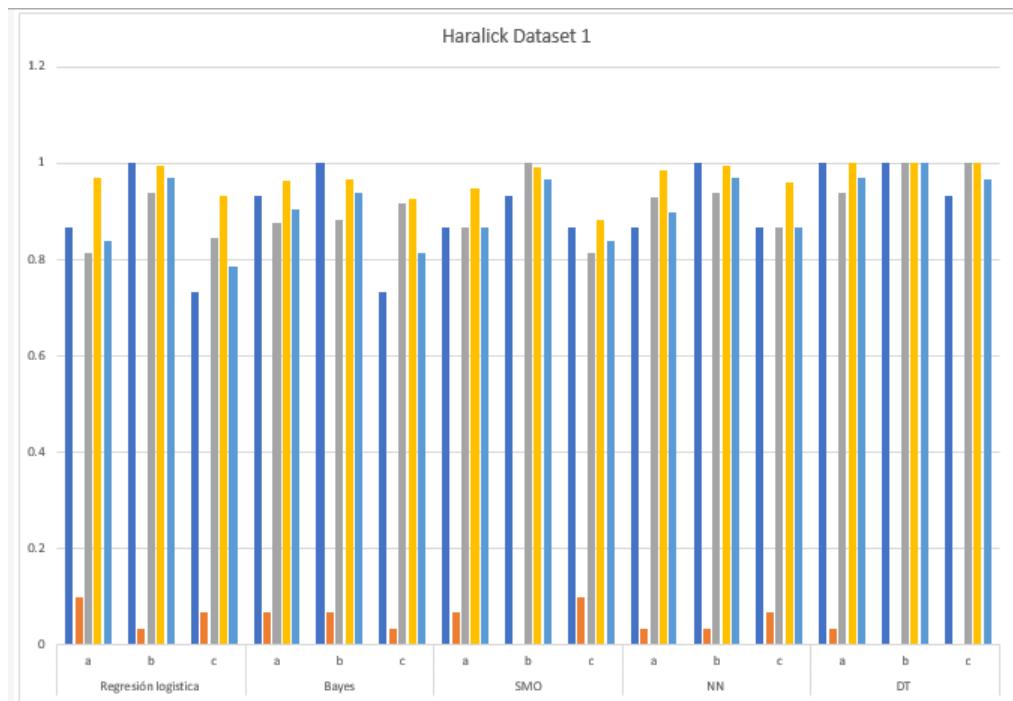


Figura 5.2: Gráfica de resultados usando Haralick

En la Tabla 5.2 se muestran los resultados obtenidos en las distintas métricas mencionadas anteriormente. La tabla muestra un buen desempeño de los clasificadores al utilizar solamente características de Haralick. En conjuntos de datos grandes, la selección de características y la reducción de dimensionalidad puede ayudar a minimizar los tiempos de entrenamiento y en algunos casos mejorar la precisión de clasificación. Sin embargo, en los conjuntos de datos utilizados no es posible apreciar una ventaja significativa.

5.3. Desempeño de clasificadores utilizando LBP

En esta sección se discuten los resultados obtenidos utilizando solo Patrones Binarios Locales (LBP). En la primer columna de la Tabla 5.3 define el clasificador utilizado. En los experimentos llevados a cabo se utilizaron los clasificadores Árboles de decisión, Redes Bayesianas, Redes Neuronales, Regresión logística y Support Vector Machines utilizando el algoritmo de Optimización mínima Secuencial (SMO).

Cuadro 5.3: Resultados utilizando LBP

	VP	FP	Precisión	AUC-ROC	F-Measure	Clase
Regresión logistica	0.867	0.133	0.765	0.909	0.813	1
	0.800	0.033	0.923	0.938	0.857	2
	0.867	0.067	0.867	0.976	0.867	3
Bayes	0.867	0.233	0.650	0.816	0.743	1
	0.533	0.033	0.889	0.649	0.667	2
	0.667	0.200	0.625	0.812	0.645	3
SMO	0.800	0.267	0.600	0.803	0.686	1
	0.533	0.133	0.667	0.741	0.593	2
	0.667	0.100	0.769	0.804	0.714	3
NN	0.733	0.167	0.688	0.813	0.710	1
	0.600	0.133	0.692	0.851	0.643	2
	0.733	0.167	0.688	0.916	0.710	3
Random forest	0.800	0.233	0.632	0.818	0.706	1
	0.533	0.100	0.727	0.754	0.615	2
	0.533	0.233	0.533	0.850	0.533	3

La Tabla 5.3 y la Figura 5.3 muestran los resultados obtenidos utilizando Patrones

Binarios Locales de la región de los globulos oculares. Es posible ver tanto en la Tabla como en la gráfica que los resultados con LBP son los peores obtenidos con las tres técnicas, aunque en promedio son muy similares, todos los clasificadores tienen un desempeño muy pobre. Aun con ello, el mejor clasificador según la gráfica mostrada en 5.3 es el método de regresión logística, mientras que el peor desempeño es para el clasificador Bayesiano y árboles de decisión. Sin embargo, como se menciono anteriormente en todos los clasificadores el desempeño es muy pobre en comparación con HoG y Características texturales de Haralick.

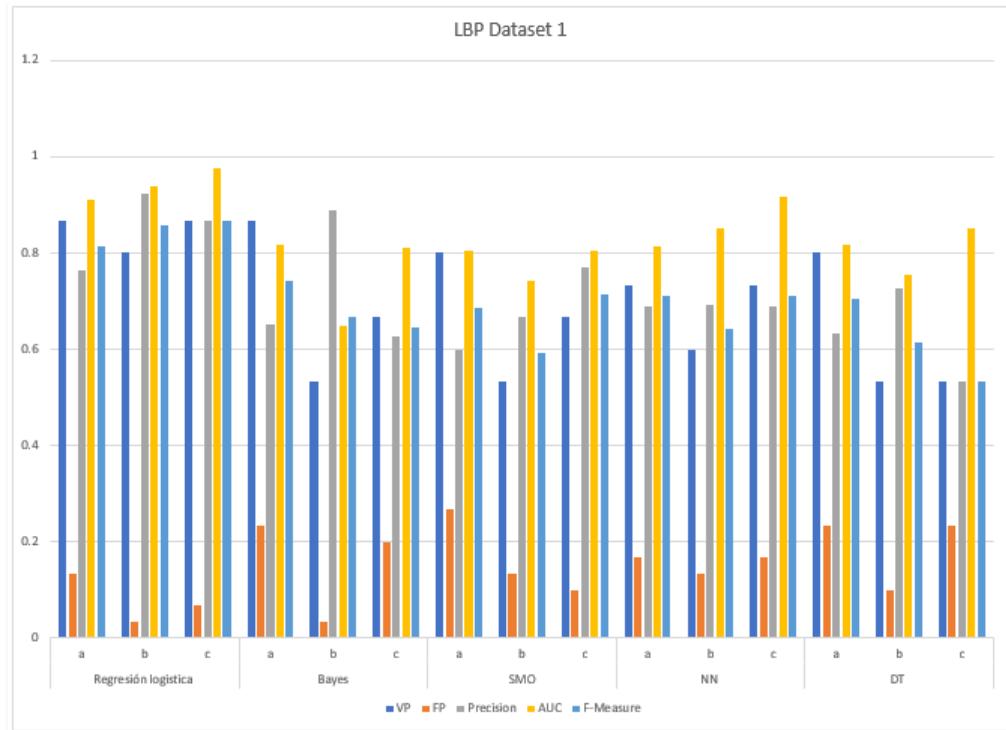


Figura 5.3: Gráfica de resultados usando LBP

En la Tabla 5.3 se muestran los resultados obtenidos con las distintas métricas mencionadas anteriormente. La tabla muestra que LBP no es un buen método para resolver este problema. El desempeño con casi todos los clasificadores, aunque no es muy malo, si está muy por debajo de los desempeños obtenido utilizando HoG y Haralick.

5.4. Desempeño de clasificadores para retinopatía diabética (Conjunto de datos 2)

El segundo conjunto de datos fue obtenido de (**Porwal, 2018**) y este puede ser obtenido en <http://dx.doi.org/10.21227/H25W98>. El conjunto de datos contiene 516 imágenes con diferentes grados de retinopatía. Cada una de las clases contiene un diferente número de datos, es decir el conjunto de datos no está balanceado. Se seleccionó este conjunto de datos porque es un conjunto de datos más grande y con ello los resultados pueden ayudar a obtener una mejor idea del desempeño tanto de los clasificadores como de los descriptores utilizados. Esto debido a que en el conjunto de datos 1 los datos están completamente balanceados, además el conjunto de datos es muy pequeño. En los experimentos realizados, se realizaron dos tipos de pruebas con este conjunto de datos. En la primera prueba, se dividió el conjunto original en solo dos clases (sano y retinopatía), mientras que en la segunda prueba se utilizaron todas las clases. La clase sana y cuatro grados de retinopatía.

5.4.1. Prueba 1

En esta prueba como mencionamos antes solo tenemos dos clases. La clase de imágenes de personas sanas y la clase de imágenes que padecen retinopatía en alguno de sus grados. El conjunto de datos es claramente desbalanceado, con más datos en la clase positiva que en la clase negativa.

La Tabla 5.4 muestra los resultados obtenidos con el conjunto de datos 2 con solo dos clases, la clase con imágenes de personas sanas y la clase con imágenes de personas que sufren retinopatía en cualquiera de sus grados. La Tabla muestra los resultados con los tres tipos de características texturales aplicados en el conjunto de datos 1 (LBP, HoG y Haralick). En la Tabla, es posible ver que los mejores resultados son obtenidos con SVM utilizando HoG, con una mejor área bajo la curva, F-measure y precisión. Solo en una métrica fueron mejores los árboles de decisión. Los resultados obtenidos con LBP y Haralick están siempre por debajo de HoG pero con las características de Haralick son las que peores resultados arrojan de los tres tipos de características utilizados.

Cuadro 5.4: Resultados para la detección de retinopatía en cualquiera de sus grados

	Clasificador	Métrica de desempeño				
		Acc	Precisión	Recall	F-measure	ROC
Haralick	SVM	67.05	0.616	0.671	0.590	0.626
	DT	67.44	0.653	0.674	0.658	0.634
	Bayes	65.50	0.653	0.655	0.654	0.639
	NN	68.79	0.664	0.688	0.665	0.618
LBP	SVM	72.86	0.716	0.729	0.699	0.727
	DT	73.64	0.730	0.736	0.703	0.769
	Bayes	68.60	0.667	0.686	0.671	0.636
	NN	73.25	0.721	0.733	0.722	0.743
HoG	SVM	76.93	0.761	0.769	0.760	0.818
	DT	77.51	0.768	0.775	0.763	0.815
	Bayes	72.48	0.739	0.725	0.730	0.747
	NN	74.61	0.740	0.746	0.742	0.698

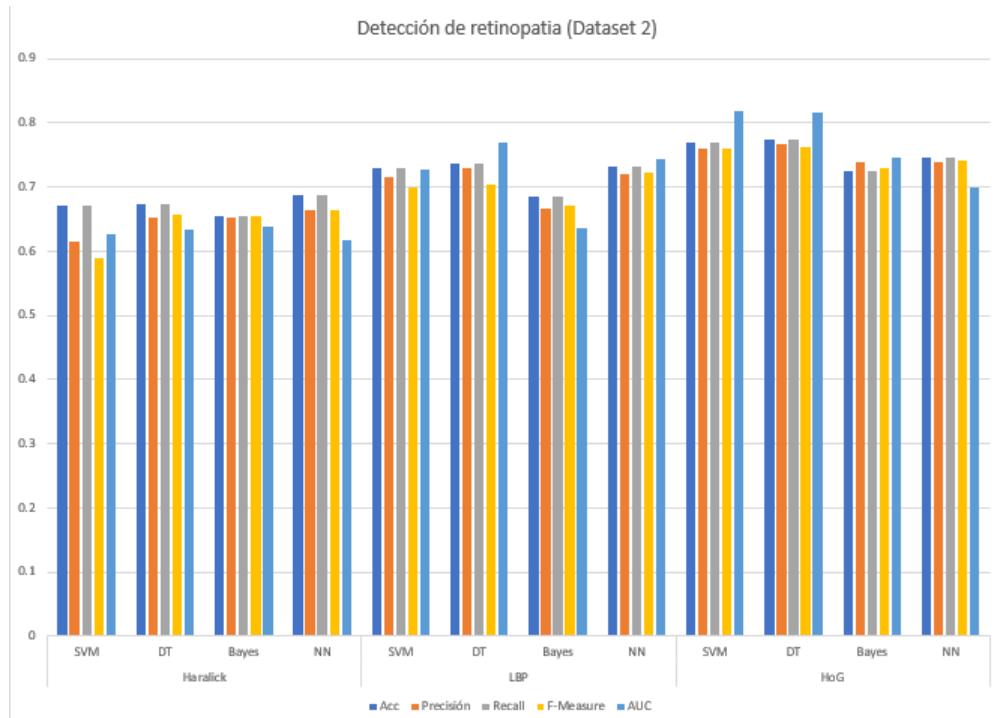


Figura 5.4: Gráfica de resultados para la detección de la retinopatía en cualquiera de sus grados

Por otro lado, es mas claro ver en la Figura 5.4 que los clasificadores entrenados con las características HoG tienen un mejor desempeño en comparación con las características de Haralick y LBP. La gráfica muestra que el mejor desempeño es obtenido por SVM utilizando HoG, mientras que el peor desempeño es obtenido con el clasificador Bayesiano utilizando LBP.

5.4.2. Prueba 2

En esta prueba tenemos cinco clases. la clase con imágenes de personas sanas y 4 clases con imágenes de personas que sufren retinopatía en alguno de sus grados.

Cuadro 5.5: Resultados de la detección en grados de retinopatía

		Clasificador Métrica de desempeño				
		Acc	Precisión	Recall	F-measure	ROC
Haralick	SVM	37.98	0.248	0.380	0.300	0.556
	DT	35.07	0.323	0.351	0.330	0.600
	Bayes	32.17	0.209	0.322	0.239	0.490
	NN	37.59	0.262	0.376	0.303	0.599
LBP	SVM	42.05	0.278	0.421	0.333	0.587
	DT	40.69	0.361	0.407	0.371	0.650
	Bayes	30.23	0.329	0.302	0.310	0.579
	NN	42.05	0.341	0.421	0.369	0.678
HoG	SVM	45.93	0.407	0.459	0.423	0.703
	DT	46.12	0.438	0.461	0.405	0.694
	Bayes	34.10	0.440	0.341	0.342	0.637
	NN	44.37	0.401	0.444	0.409	0.658

La Tabla 5.5 muestra los resultados con los tres tipos de características texturales aplicados en el conjunto de datos 1 (LBP, HoG y Haralick). En la Tabla es posible ver que todos los resultados obtenidos con este conjunto de datos son muy malos. Las precisiones obtenidas estan por debajo del 50 % e incluso el área bajo la curva esta apenas por encima de 0.5 en la mayoria de los casos. Esto da a entender que los clasificadores casi estan adivinando, lo que en su conjunto demuestra un bajo desempeño. Una posible razón podría

deberse al desbalance en el conjunto de datos. Algunas técnicas que pueden tratar con este tipo de datos podrían ayudar a mejorar el desempeño de los clasificadores. En trabajos futuros se espera mejorar la precisión de clasificación utilizando más características y seleccionando las mejores mediante la ayuda de genéticos o algoritmos híbridos.

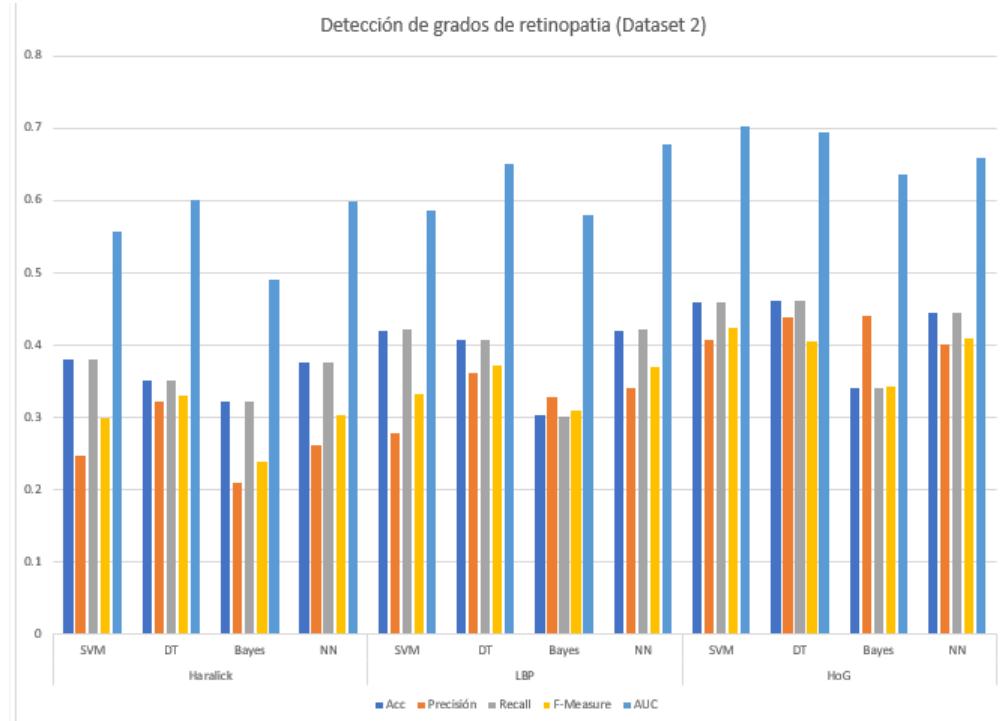


Figura 5.5: Gráfica de resultados para la detección de la retinopatía en alguno de sus grados

Capítulo 6

Conclusiones

En esta tesis se han realizado diferentes pruebas con diferentes algoritmos en el estado del arte para clasificación de retinopatía y glaucoma. En los experimentos llevados a cabo se realizan pruebas con las técnicas de extracción de características y técnicas de clasificación. Sobre las técnicas de extracción de características, se utilizaron solo dos conjuntos de técnicas (utilizando histograma de gradientes ordenados y las definidas en el capítulo 4) y técnicas de clasificación (árboles de decisión –random forest-, redes Bayesianas, redes neuronales y SVM) aunado a eso los resultados se compararon con tres métricas de desempeño diferentes y se presentan los resultados.

De los resultados obtenidos y mostrados en el Capítulo anterior, se pueden argumentar que al utilizar solo un conjunto de características los clasificadores no tienen un buen desempeño, sin embargo al utilizar todas las características enumeradas en el Capítulo 4 los clasificadores mejoran significativamente su desempeño. Un estudio más profundo sobre las características en la clasificación se debería realizar para reducir el tiempo de extracción de características y clasificación. Esto además podría proporcionar una mejor idea del comportamiento de las técnicas de selección de características.

En trabajos futuros se debería de considerar eliminar características mediante técnicas actuales de selección de características, haciendo énfasis en la contribución de la combinación de características, para ello es indispensable hacer uso de técnicas evolutivas. En los experimentos llevados a cabo, se utilizó un conjunto de datos con solo 45 imágenes,

siendo importante aumentar el número e imágenes en el conjunto de datos.

Bibliografía

[UNAM, 2020] UNAM Global, EN MÉXICO, CERCA DE 1.5 MILLONES DE PERSONAS TIENEN GLAUCOMA Boletín: Número 221, CDMX, Ciudad Universitaria.

[Forbes, 2015] Forbes Staff, 10.6 millones de mexicanos padecen diabetes, Forbes México, Recuperado el 8 de Mayo del 2020 , de <https://www.forbes.com.mx/10-6-millones-de-mexicanos-padecen-diabetes/>.

[Encalada, 2016] Mirian Susana Encalada Ojeada y Edwin Gerardo Gamboa Vinueza , Desarrollo de un sistema de soporte a la detección del glaucoma a través de procesamiento digital de imágenes biomédicas del fondo de ojo y uso de software libre, Tesis de Licenciatura, Universidad Politécnica de Salesiana.

[Wong, 2016] Wong TY, Bressler NM. Artificial Intelligence With Deep Learning Technology Looks Into Diabetic Retinopathy Screening. JAMA. 2016;316(22):2366–2367. doi:10.1001/jama.2016.17563

[Maul, 2002] Eugenio Maul de la Puente: Glaucoma, Universidad Católica de Chile, 2002.

[Kanski, 2004] Jack J. Kanski: Oftalmología clínica, 5^a edición, 2004, ISBN 978-84-8174-758-4

[Shapiro, 2001] Linda G. Shapiro and George C. Stockman (2001). Computer Vision. Prentice Hall. ISBN 0-13-030796-3.

[Prensky, 2001] Prensky, Marc (2001). «Digital Natives, Digital Immigrants». On the Horizon, MCB University Press

- [Cuevas, 2010] Cuevas, E., Zaldívar, D., Pérez Cisneros, M. (2010). Procesamiento Digital de Imágenes con Matlab y Simulink. México,D.F.: Alfaomega Grupo Editor.
- [Schmidhuber, 2015] Schmidhuber, 2015) Schmidhuber, J. (2015). «Deep Learning in Neural Networks: An Overview». *Neural Networks* 61: 85-117.
- [Ferreira, 2016] Ferreira, C. (2006). «Designing Neural Networks Using Gene Expression Programming». In A. Abraham, B. de Baets, M. Köppen, and B. Nickolay, eds., *Applied Soft Computing Technologies*:
- [Ostaszewski, 1990] Ostaszewski, Adam (1990). Advanced mathematical methods. Cambridge; Nueva York: Cambridge University Press. p. 129. ISBN 0521289645.
- [Lang, 2014] Jean-Philippe Lang, Predictors tutorial Archivado el 3 de enero de 2014 en la Wayback Machine., Bioinformatic Department Projects,
- [Beitzel, 2006] Beitzel., Steven M. (2006). On Understanding and Classifying Web Queries (Ph.D. thesis). IIT.
- [Zou, 2007] Zou KH, O'Malley AJ, Mauri L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models.
- [Morris, 2004] Tim Morris (2004). Computer Vision and Image Processing. Palgrave Macmillan
- [Linda, 2001] Linda G. Shapiro and George C. Stockman (2001). Computer Vision. Prentice Hall.
- [Gonzalez, 2002] Rafael C. Gonzalez, Richard E. Woods 2002 “Digital Image Proccesing Second Edition Prentice Hall
- [Sonka, 1998] Milan Sonka, Vaclav Hlavac, Roger Boyle 1998 “Image Processing, Analysis, and Machine Vision” (2nd Edition)
- [Hartigan, 1975] Hartigan (1975). Clustering algorithms. John Wiley Sons, Inc.w

[Presutti, 2004] M. Presutti. la matriz de co-ocurrencia en la clasificación multiespectral: tutorial para la enseñanza de medidas texturales en cursos de grado universitario. Universidad Nacional de La Plata, Agosto, 2004.

[Matas, 2010] Trefný, Jirí, and Jirí Matas..^Ex tended set of local binary patterns for rapid object detection."Proceedings of the Computer Vision Winter Workshop. Vol. 2010. 2010.

[Carrillo, 2017] Histograma orientado a gradientes con máquina de soporte vectorial, en la clasificación del alfabeto dactilológico,J. M. Carrillo,2017

[Vapnik, 1995] Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks"(PDF). Machine Learning. 20 (3): 273–297. CiteSeerX

[Haralick, 1973] Haralick, R. M., Shanmugan, K., and Dinstein, I. 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, Vol.3 n.6, pp.610-621.