



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA  
INGENIERÍA ELÉCTRICA –PROCESAMIENTO DIGITAL DE SEÑALES

DESCRIPCIÓN Y CLASIFICACIÓN DE TEXTURAS BASADA EN POLINOMIOS  
ORTOGONALES DISCRETOS

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
DOCTOR EN INGENIERÍA

PRESENTA:  
M.I. GERMÁN ABRAHAM GONZÁLEZ SÁNCHEZ

TUTOR PRINCIPAL  
DR. BORIS ESCALANTE RAMÍREZ  
FACULTAD DE INGENIERÍA

COMITÉ TUTOR  
DRA. LUCÍA MEDINA GÓMEZ  
FACULTAD DE CIENCIAS  
DR. FERNANDO ARÁMBULA COSÍO  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN  
SISTEMAS (IIMAS)

CIUDAD UNIVERSITARIA, CDMX. OCTUBRE 2018



**JURADO ASIGNADO:**

Presidente: Dra. Medina Gómez Lucía  
Secretario: Dr. Garduño Ángeles Edgar  
Vocal: Dr. Escalante Ramírez Boris  
1er. Suplente: Dr. Rubio Acosta Ernesto  
2do. Suplente: Dr. Moya Albor Ernesto

Lugar donde se realizó la tesis: Facultad de Ingeniería, Ciudad Universitaria, UNAM.

**TUTOR DE TESIS:**

Dr. Escalante Ramírez Boris

-----  
**FIRMA**



- A Dios.
- A mi mamá, mi papá, tía Rosita y mi hermano.
- A mis amigos, quienes me siguen aguantando.

## *Agradecimientos*

En primer lugar, quiero agradecer el gran apoyo, la confianza y la paciencia de mi tutor Boris Escalante. A veces uno no valora las cosas que tiene ni el entorno en donde está, pero el Dr. Boris se ha encargado de que el Laboratorio Avanzado de Procesamiento de Imágenes (LAPI) sea un lugar muy cómodo para trabajar, tanto en equipo como en el trato entre compañeros.

Al comité tutorial y sinodales por haberse tomado el tiempo para revisar este trabajo y contribuir a mejorar la calidad del mismo a través de sus comentarios y correcciones. Gracias por ser siempre amables y atentos.

A mis padres por haberme cuidado y educado, me he dado cuenta que soy un privilegiado gracias a su esfuerzo, no el mío. A mi mamá por el apoyo constante y amor incondicional. A mi papá por mostrarme el valor de la responsabilidad y el trabajo.

A mis compañeros en el laboratorio, los que estuvieron y los que están, por hacer que los días en él sean relajados y amenos. En especial, un agradecimiento a Uriel por ayudarme a encontrar el tema de este trabajo de investigación cuando más oscuro parecía el panorama, por tomarme en cuenta para la elaboración de algunos artículos y por siempre estar dispuesto a compartir su experiencia y conocimiento.

*A na Laura pel suport i ànims constants. Gràcies per la pau, la tranquil·litat i l'amor que existeixen entre nosaltres, ets valuosíssima. Record a que per mi, tu ets un regal immerescut de la vida. Et vull bé.*

A la Universidad Nacional Autónoma de México por formarme desde hace ya 15 años. Estoy muy orgulloso de pertenecer a esta comunidad y por la relevancia que tiene a nivel nacional en los ámbitos científicos, culturales y sociales. Además, estoy muy agradecido por lo que representa para mí a nivel personal: por todos los amigos, compañeros y profesores que he conocido en sus instalaciones.

A mi Nina por su invaluable compañía durante las tardes que dediqué a la escritura de este trabajo de investigación.

A CONACYT por el soporte económico dado durante mis estudios bajo el número CONACYT263921. Además, quiero agradecer el gran apoyo de los proyectos PAPIIT IN116917 y SECITI 110/2015.

# DESCRIPCIÓN Y CLASIFICACIÓN DE TEXTURAS BASADAS EN POLINOMIOS ORTOGONALES DISCRETOS

## Resumen

En los últimos años, los momentos ortogonales discretos (DOMs) han atraído la atención de la comunidad científica porque han sido una herramienta adecuada para la extracción de características. Los DOMs tienen el atributo de describir características locales y globales en imágenes de manera eficiente. La representación generada a través de los DOMs se considera óptima porque tienen redundancia mínima de información. Esta característica tiene origen en la condición de ortogonalidad de la base polinomial con la que se calculan los momentos. Los DOMs, al definirse en el dominio discreto, su cálculo no requiere una cuantización espacial. Sin embargo, la inestabilidad numérica que surge debido al cálculo de momentos de orden alto es el principal inconveniente que limita su aplicación. Dado que los DOMs se calculan como la proyección de la imagen en un núcleo ponderado o un conjunto de polinomios ortogonales, su cálculo generalmente está relacionado con el tamaño de la imagen. Cuanto mayor sea el orden del momento, mayor será el error numérico. En esta tesis, proponemos métodos de descripción, y posterior clasificación, de textura de imágenes que evitan los errores numéricos basados en los momentos discretos de Tchebichef, Krawtchouk y Dual Hahn. Estas familias de momentos son las más comunes en la literatura. También presentamos los momentos discretos de Shmaliy, que son una nueva familia de momentos ortogonales discretos basados en los polinomios Shmaliy. Los diferentes esquemas propuestos para la descripción de texturas son aplicados en algunos conjuntos de imágenes biológicas incluidos en la base de datos pública IICBU-2008 y se validan a través del esquema de validación cruzada. Los resultados se calculan utilizando los siguientes clasificadores: *Random forest*, máquinas de soporte vectorial, *naive* Bayes y  $K$  vecinos más cercanos. Los resultados de clasificación obtenidos con la descripción basada en DOMs son comparados con métodos encontrados en la literatura y muestran mejoría respecto a ellos, mostrando así la efectividad de nuestra propuesta para la descripción de texturas.

# TEXTURE DESCRIPTION AND CLASSIFICATION BASED ON DISCRETE ORTHOGONAL POLYNOMIALS

## Abstract

In recent years, discrete orthogonal moments (DOMs) have attracted the attention of the scientific community because they are a suitable tool for feature extraction. DOMs possess the attribute of describing efficiently local and global features in images efficiently. The DOMs are optimal in the sense that they represent images with minimal information redundancy. This characteristic is originated from the orthogonality condition of the polynomial basis where the moments are computed. Furthermore, contrary to the continuous moments, the DOMs are defined in the discrete domain. Hence, their computation does not require spatial quantization. However, the numerical instability that arises because of the computation of high-order moments is the main drawback that limits their wider application. Since the DOMs are calculated as the projection of the image onto a weighted kernel or set of orthogonal polynomials, their computational complexity is usually linked to the size of the image. The higher the order of the moment, the higher the numerical error. In this thesis, we propose image classification methods that avoid numerical errors based on Tchebichef, Krawtchouk, and Dual Hahn discrete moments. These moment families are the most common in the literature. We also introduced discrete Shmaliy moments, which are a new family of moments derived from Shmaliy polynomials. The proposed schemes for texture description are applied in some biological image datasets included in the IICBU-2008 benchmark and they are carried out through cross-validation. The results are computed using random forest, support vector machines, naïve Bayes, and k-nearest neighbors classifiers. The classification results reached out with the DOMs-based description are compared with methods found in the literature and show improvement with respect to them, thus showing the effectiveness of our proposal for image texture description.

# Índice general

	Página
<b>1 Introducción</b>	<b>1</b>
<b>2 Momentos y polinomios ortogonales discretos</b>	<b>5</b>
2.1 Polinomios y Momentos Ortogonales Continuos	6
2.1.1 Antecedentes	6
2.1.2 Polinomios ortogonales clásicos	7
2.1.3 Polinomios ortogonales radiales	11
2.2 Polinomios y Momentos Ortogonales Discretos	13
2.2.1 Polinomios y Momentos de Tchebichef	14
2.2.2 Polinomios y Momentos de Krawtchouk	15
2.3 Polinomios ortogonales de rejilla no uniforme	17
2.3.1 Polinomios y Momentos Dual Hahn	20
2.4 Polinomios y Momentos de Shmaliy	23
2.5 Implementación de los polinomios ortogonales discretos	25
2.5.1 Comparación entre estrategias de cálculo de polinomios y momentos ortogonales	26
2.5.2 Polinomios de Tchebichef - Recursión sobre $n$	27
2.5.3 Polinomios de Tchebichef - Recursión sobre $x$	28
2.5.4 Polinomios de Krawtchouk - Recursión sobre $n$	29
2.5.5 Polinomios dual Hahn - Recursión sobre $n$	30
2.5.6 Polinomios de Shmaliy - Recursión sobre $n$	30
<b>3 Clasificación de texturas con polinomios y momentos ortogonales discretos</b>	<b>33</b>
3.1 Firma de textura basada en momentos discretos de Tchebichef	34
3.2 Cálculo de <i>Textons</i>	35
3.3 Esquemas de clasificación propuestos utilizando Polinomios y Momentos Ortogonales Discretos	38
3.3.1 Firmas estadísticas de textura	39
3.3.2 Textons con Polinomios Ortogonales Discretos	40
3.3.3 Firmas de textura $k$ -medias	43
<b>4 Tratamiento de los vectores de características</b>	<b>45</b>
4.1 Selección de características	45
4.1.1 Eliminación recursiva de características	46
4.1.2 Relieff	48
4.2 Análisis Lineal Discriminante	49
4.2.1 Análisis Discriminante Lineal de dos clases	49
4.2.2 Análisis Discriminante Multiclase	51
4.2.3 Análisis Discriminante de Fisher	51
4.2.4 Análisis Discriminante Multiclase Kernel-Fisher	52

<b>5 Experimentos y Resultados</b>	<b>55</b>
5.1 IICBU - 20008	56
5.2 Experimento con firmas estadísticas de textura	60
5.2.1 Análisis de la varianza	64
5.3 Experimento con <i>Textons</i> y Firmas de Textura de $k$ -medias	72
5.4 Estudios secundarios	76
5.4.1 Epitelio y Estroma en Cáncer colorrectal	76
5.4.2 Clasificación de enfermedades hematológicas	79
<b>6 Conclusiones y Trabajo Futuro</b>	<b>83</b>
6.1 Conclusiones	83
6.2 Trabajo Futuro	84
<b>A Patrones Binarios Locales (LBPs)</b>	<b>85</b>
A.1 Contraste invariante a la rotación	87
<b>B Clasificadores</b>	<b>89</b>
B.1 $K$ vecinos más cercanos	89
B.1.1 Desarrollo	89
B.2 Máquinas de Soporte Vectorial	91
B.2.1 Máquina de soporte vectorial para muestras linealmente separables	92
B.2.2 Máquina de soporte vectorial para muestras no linealmente separables	92
B.2.3 Proyección kernel para el caso no linealmente separable	95
B.3 <i>Random Forest</i>	96
B.3.1 Esquema general de <i>Random Forest</i>	97
B.4 Naive Bayes	98
B.4.1 Modelo probabilístico	100
B.4.2 Estimación de parámetros y modelos de eventos	101
<b>Bibliografía</b>	<b>103</b>

# Índice de figuras

	Página
2.1 (a) Polinomios discretos de Tchebichef de orden 0 a 4 con $N = 200$ . (b) Polinomios discretos de Tchebichef bidimensionales de orden 0 a 8, el orden se considera diagonalmente.	16
2.2 (a) Polinomios de Krawtchouk de orden 0 a 4 con $N = 100$ . (b) Polinomios de Krawtchouk bidimensionales de orden 0 al 8.	17
2.3 Polinomios de Krawtchouk bidimensionales con variación en el parámetro $p$ de orden 0.	18
2.4 Esquema Askey de Polinomios ortogonales hipergeométricos y las relaciones entre ellos.	20
2.5 (a) Polinomios dual Hahn de orden 0 a 4 con $N = 60$ y (b) Polinomios dual Hahn bidimensionales de orden 0 al 8.	22
2.6 Polinomios dual Hahn con variación de parámetros. Los polinomios de color negro se obtienen con los parámetros $a = 0$ , $b = 20$ y $c = 0$ . Por otro lado, los polinomios de color azul se calculan con los valores $a = 5$ , $b = 45$ y $c = 5$ . Finalmente, los polinomios de color rojo son el resultado del cálculo con $a = 20$ , $b = 60$ y $c = 20$ . Con la variación de estos parámetros se muestran las propiedades de adaptación a una zona de interés y la habilidad de expansión/compresión propias de los polinomios dual Hahn.	23
2.7 (a) Polinomios discretos de Shmaliiy de orden 0 a 4 con $N = 200$ . (b) Polinomios discretos de Shmaliiy bidimensionales de orden 0 a 8.	25
3.1 Conjunto de polinomios de Tchebichef en el dominio (a) espacial y (b) frecuencial.	34
3.2 El banco LM es una combinación de filtros de borde, barra y punto con múltiples escalas y orientaciones, en total son 48 filtros.	36
3.3 El banco de filtros S es invariante a la rotación y está compuesto por 13 filtros isótropicos, semejantes a los filtros de Gabor.	36
3.4 El banco de filtros MR8 consiste en 2 filtros anisotrópicos y 2 filtros simétricos rotacionalmente. Sin embargo, sólo 8 respuestas son consideradas y corresponden a las máximas respuestas a los filtros anisotrópicos sobre todas las orientaciones.	36
3.5 Proceso para la generación del diccionario de <i>Textons</i> .	37
3.6 Para cada imagen de entrenamiento, calcular el Mapa de <i>Texton</i> y obtener el histograma.	38
3.7 Clasificación de la imagen de validación a partir de la comparación del histograma de mapa de <i>Texton</i> contra los histogramas modelo de cada clase.	38
3.8 Descripción de una imagen a través de ventanas traslapadas.	39
3.9 Esquema general de la clasificación con firmas de textura estadísticas.	40
3.10 Cálculo del diccionario de <i>Textons</i> con Polinomios Ortogonales Discretos.	41
3.11 Clasificación con histogramas de <i>Textons</i> con Polinomios Ortogonales Discretos.	42
3.12 Clasificación con el diccionario de <i>Textons</i> con Polinomios Ortogonales Discretos.	42
3.13 Cálculo de los histogramas de Firmas de Textura.	43
3.14 Clasificación con los histogramas de Firmas de Textura con Momentos Ortogonales Discretos.	44
3.15 Clasificación con el diccionario de Firmas de Textura con Momentos Ortogonales Discretos.	44

5.1	Las dos clases definidas en la base de datos <i>Binucleate</i> son: (a) <i>Binucleate</i> y (b) <i>Non-Binucleate</i> . . . . .	56
5.2	Las edades cronológicas de la base de datos <i>Celegans</i> : (a) 1, (b) 2, (c) 4 y (d) 8 días. . . . .	57
5.3	Las clases de la base de datos <i>CHO</i> son: (a) <i>giantin</i> , (b) <i>hoechst</i> , (c) <i>lamp2</i> , (d) <i>nop4</i> y (e) <i>tubulin</i> . . . . .	58
5.4	Las clases de la base de datos <i>HeLa</i> son: (a) <i>actin</i> , (b) <i>dna</i> , (c) <i>endosome</i> , (d) <i>er</i> , (e) <i>golgia</i> , (f) <i>golgpp</i> , (g) <i>lysosome</i> , (h) <i>microtubules</i> , (i) <i>mitochondria</i> y (j) <i>nucleolus</i> . . . . .	59
5.5	Las imágenes pertenecen a las diferentes edades cronológicas de los ratones sin dieta restrictiva de la base <i>Liver aging</i> : (a) 1, (b) 6, (c) 16 y (d) 24 meses. . . . .	59
5.6	Clases de la base de datos <i>Lymphoma</i> : (a) <i>CLL</i> , (b) <i>FL</i> , y (c) <i>MCL</i> . . . . .	60
5.7	Diagrama general de la validación cruzada por grupos. . . . .	62
5.8	Resultados promedio de exactitud en la clasificación de las familias de momentos en comparación con: (a) el tipo de procesamiento y (b) el tamaño de ventana. . . . .	63
5.9	Procedimiento general para la clasificación basada en firmas de textura estadística y análisis discriminante. . . . .	63
5.10	Las firmas estadísticas de textura son transformadas a través del análisis discriminante. Los valores proyectados corresponden a los $C-1$ valores generadores de mayor magnitud, donde $C$ es el número de clases que contiene cada base de datos. Las proyecciones obtenidas mejoran el agrupamiento entre clases y las separaciones entre clases son maximizadas. . . . .	64
5.11	Resultados promedio de exactitud para la base de datos <i>Binucleate</i> . Esta base de datos contiene solamente dos clases. Esta base de datos tiene la particularidad de que el LDA proyecta y separa las clases de tal manera que la clasificación se vuelve trivial. Sin embargo, cuando la SVM-gsn es utilizada, hay un pequeño error causado por la influencia de los parámetros del kernel Gaussiano. . . . .	65
5.12	Resultados promedio de exactitud para la base de datos <i>Celegans</i> . La base de datos contiene cuatro clases. Los mejores resultados son obtenidos cuando el tamaño de la ventana de análisis es mayor a 20. . . . .	65
5.13	Resultados promedio de exactitud para la base de datos <i>CHO</i> . La base de datos consiste en 5 clases de estructuras celulares de ovario de hámster. En esta base de datos, las firmas de los momentos Shmaliy y Tchebichef muestran un rendimiento uniforme y similar cuando el tamaño de la ventana de análisis es mayor que 30. . . . .	66
5.14	Resultados promedio de exactitud para la base de datos <i>HeLa</i> . La base de datos, como extensión de la base de datos <i>CHO</i> , contiene 10 clases de estructuras celulares de ovario de hámster. Los mejores resultados fueron obtenidos cuando el tamaño de ventana de análisis es mayor que 40. . . . .	66
5.15	Resultados promedio de precisión para la base de datos <i>Liver aging</i> . La base de datos contiene 4 clases diferentes. La base de datos tiene la particularidad de que el rendimiento de los clasificadores es similar entre ellos y la exactitud se incrementa conforme aumenta el tamaño de la ventana. . . . .	67
5.16	Resultados promedio de exactitud para la base de datos <i>Lymphoma</i> . La base de datos contiene 3 tipos de enfermedades hematológicas. Los mejores resultados se obtienen cuando la ventana de análisis es mayor que 40. . . . .	67
5.17	ANOVA de las variables independientes de <i>Celegans</i> . En la imagen izquierda, están los resultados promedio de cada clasificador, se observa que KNN y <i>naive</i> Bayes tienen los mejores resultados, sin embargo, todos los demás clasificadores tienen resultados parecidos. En el centro, se ve que los momentos de Krawtchouk tienen el mejor resultado promedio. Los resultados de cada ventana se muestran en la imagen de la derecha, si la ventana mide $30 \times 30$ o más, tendrá los mejores resultados. . . . .	69

5.18	ANOVA de las variables independientes de <i>CHO</i> . El resultado promedio de los clasificadores se muestran en la imagen de la izquierda, todos los clasificadores tienen resultados parecidos, menos <i>Random Forest</i> . En el centro, están los resultados por momento, Dual Hahn muestra un peor rendimiento que los demás. En la imagen de la izquierda, se observa que, conforme aumenta el tamaño de ventana, los resultados mejoran. . . . .	70
5.19	ANOVA de las variables independientes de <i>HeLa</i> . En la imagen de la izquierda, se observa que los clasificadores SVM lineal y <i>Random Forest</i> tienen los peores resultados promedio. En la imagen del centro, se nota que los momentos Dual Hahn tiene el peor rendimiento. En la imagen de la izquierda, se observa que, conforme aumenta el tamaño de ventana, los resultados mejoran. . . . .	70
5.20	ANOVA de las variables independientes de <i>Liver aging</i> . En la imagen izquierda, los resultados promedio indican que el peor rendimiento pertenece al <i>Random Forest</i> . En la imagen central, Dual Hahn tiene el peor rendimiento, mientras que las demás familias tienen un rendimiento similar entre ellas. En la imagen derecha, se observa que conforme aumenta el tamaño de ventana, el rendimiento mejora. . . . .	70
5.21	ANOVA de las variables independientes de <i>Lymphoma</i> . En la imagen izquierda, se muestra que el rendimiento de los clasificadores es similar. En la imagen central, se muestra que los momentos Dual Hahn tienen, ligeramente, un peor resultado comparándolo con los demás momentos. Los resultados de cada ventana se muestran en la imagen de la derecha, si la ventana mide $40 \times 40$ o más, tendrá los mejores resultados . . . . .	71
5.22	Muestras de cáncer colorrectal en TMA digitalizados, sólo el canal rojo. La primera fila muestra solamente el epitelio del tumor y la segunda fila muestra el estroma extraída desde un bloque de parafina. . . . .	77
5.23	Esquema general de la clasificación del epitelio y estroma del cáncer colorrectal. . . .	78
5.24	Base de datos <i>Lymphoma</i> . Las imágenes fueron transformadas al espacio de color Hematoxilina y Eosina (H+E). La primera fila muestra las imágenes RGB originales, la segunda fila corresponde al canal Hematoxilina y la tercera fila al canal Eosina. (a) CLL, (b) FL y (c) MCL. . . . .	81
5.25	Diagrama general para la clasificación de la base de datos <i>Lymphoma</i> . . . . .	82
A.1	El LBP Original. . . . .	85
A.2	Las vecindades de (8,1), (16,2) y (8,2). Los valores de pixel son interpolados bilinealmente cuando el punto de muestreo no está justo en el centro del pixel. . . . .	86
B.1	Ejemplo del algoritmo de K vecinos. El triángulo verde es la muestra que se debe clasificar. Para $k = 5$ , círculo interno, el triángulo será clasificado como cuadrado (Clase B). Mientras que, si $k = 11$ , círculo externo, la muestra será clasificada como círculo o clase A. . . . .	90
B.2	El margen máximo $m$ es la máxima distancia que hay entre los vectores de soporte más cercanos entre clases. . . . .	91
B.3	(a) Muestras linealmente separables y (b) Muestras no linealmente separables. . . . .	93
B.4	Los puntos $x_i$ y $x_j$ son errores o <i>outliers</i> y se encuentran a una distancia $\xi_i$ y $\xi_j$ , respectivamente, que se interpreta como la variable de error o de holgura. . . . .	94
B.5	Las muestras son proyectadas al espacio del <i>kernel</i> para facilitar su separabilidad y posterior clasificación con una SVM. . . . .	95
B.6	Esquema del algoritmo del <i>Random Forest</i> . Los subconjuntos se forman aleatoriamente a partir de los datos de entrenamiento aplicando <i>bagging</i> y <i>feature bagging</i> . Las muestras seleccionadas a partir del <i>bagging</i> están de color amarillo pálido las características elegidas a partir del <i>feature bagging</i> están en color amarillo intenso. . . . .	99



# Índice de tablas

	Página
2.1 Funciones de los polinomios ortogonales clásicos: Hermite, Laguerre y Jacobi. . . . .	11
2.2 Posibles repeticiones y número de momentos de orden 0 al 3 para los momentos de Zernike. . . . .	13
2.3 Posibles repeticiones y número de momentos de orden 0 al 3 para los momentos pseudo-Zernike y Fourier-Mellin. . . . .	13
2.4 Familias de polinomios ortogonales discretos en una rejilla no uniforme. . . . .	19
2.5 Comparación entre los momentos de Racah y los dual Hahn . . . . .	20
2.6 Resumen de las características de las estrategias de cálculo directo y de recursión de los polinomios ortogonales discretos. . . . .	26
2.7 Valores de $\phi$ , $\zeta$ y $\psi$ para los polinomios de Tchebichef, Krawtchouk y dual Hahn. . . . .	27
5.1 Descripción de las bases de datos de IICBU 2008. . . . .	56
5.2 Valores $p$ del análisis de varianza. Las primeras tres columnas corresponden a cada variable independiente (Ventana, Clasificador y Momento), mientras que, las últimas tres columnas corresponden a las interacciones entre las variables independientes. . . . .	68
5.3 Comparación de resultados de exactitud promedio, en porcentaje, entre las firmas estadísticas de textura con momentos ortogonales discretos y los métodos encontrados en la literatura. . . . .	71
5.4 Resultados de clasificación con el Diccionario de <i>Textons</i> con Polinomios Ortogonales Discretos. . . . .	72
5.5 Resultados de clasificación con los histogramas de <i>Textons</i> con Polinomios Ortogonales Discretos. . . . .	73
5.6 Resultados de clasificación con el diccionario de las Firmas de textura de $k$ -medias. . . . .	74
5.7 Resultados de clasificación con los histogramas de las Firmas de textura de $k$ -medias. . . . .	75
5.8 Resultados de clasificación con el diccionario de las Firmas de textura de $k$ -medias con <i>Relieff</i> y Kernel Fisher. . . . .	75
5.9 Comparación entre todos los resultados de exactitud de clasificación de todos los esquemas propuestos. . . . .	76
5.10 Comparación de los resultados de clasificación para epitelio y estroma. . . . .	79
5.11 Resultados de clasificación de la base <i>Lymphoma</i> con diferentes descriptores. . . . .	82



# Lista de Algoritmos

	<b>Página</b>
4.1 Eliminación Recursiva de Características . . . . .	47
4.2 Algoritmo <i>Relieff</i> . . . . .	48



# 1

## Introducción

El concepto de textura es tan amplio que se le puede encontrar en diversos ámbitos del conocimiento, tanto en las artes como en la ciencia y tecnología. En primer lugar, la palabra *textura* se utilizó en primera instancia para definir la disposición y el orden de los hilos en una tela. Algunos ejemplos de los diferentes significados de *textura* con respecto al campo de conocimiento son:

- **Pintura:** la agrupación de formas y colores que permiten percibir irregularidades en una superficie continua o plana.
- **Artes Visuales:** un elemento de diseño y su aplicación en el arte.
- **Música:** la forma en que se relacionan las voces (ya sean líneas melódicas vocales o instrumentales) que intervienen en una pieza musical.
- **Cosmología:** un defecto topológico teórico en la estructura del espacio-tiempo.
- **Percepción humana:** la disposición en que están colocadas y combinadas entre sí las partículas o elementos de un cuerpo o sustancia. Esta percepción está asociada principalmente a los sentidos de la visión y el tacto.

En el ámbito del procesamiento digital de imágenes, el concepto de textura está directamente relacionado al de la percepción humana porque se define como el conjunto de métricas diseñadas para cuantificar la textura percibida en una imagen. La textura de una imagen nos brinda información sobre el conjunto de patrones en la variación de color o intensidad de gris de los píxeles en el dominio espacial de las imágenes [1]. Debido a que existen diversas métricas para caracterizar la textura, no existe un modelo o concepto preciso que pueda aplicarse de manera global.

El análisis de texturas de imágenes ha sido usado en el campo científico, médico e industrial, ya que es parte de diferentes etapas del procesamiento de imágenes, por ejemplo: segmentación, clasificación, búsqueda y recuperación de imágenes, etcétera. Actualmente, existen muchas propuestas para definir y clasificar texturas, por ejemplo, Xie y Mirmehdi [2], con base en los estudios [3–8], proponen una categorización en cuatro familias: los métodos estadísticos [9–11], los métodos basados en modelos [12–14], los métodos geométricos o estructurales [6, 15] y los métodos basados en el procesamiento de señales. Esta caracterización no es restrictiva porque, a partir de la combinación de ellos, se generan nuevos descriptores de texturas.

Los resultados experimentales sugieren que el cerebro identifica texturas a través del análisis espacio-frecuencia de la imagen de entrada [16]. Los descriptores basados en el procesamiento de señales se extraen comúnmente a partir de la aplicación de bancos de filtros sobre la imagen. Estas características se pueden calcular en el dominio espacial, frecuencial o ambos. Por ejemplo, las técnicas basadas en bancos de filtros [17, 18], Wavelets [19], Gabor [20] [21] y Hermite [22] han sido utilizadas para la identificación de texturas, mostrando buenos resultados de clasificación.

Los momentos, altamente relacionados con las técnicas de procesamiento de señales, son cantidades escalares que se usan para caracterizar una función (imagen) a través del cálculo de atributos significativos de ella. Matemáticamente, el momento  $M_{pq}$  de una imagen  $I(x, y)$  es la proyección de la imagen sobre una base polinomial  $r_{pq}(x, y)$  y se define como  $M_{pq} = \int_x \int_y r_{pq}(x, y) I(x, y) dx dy$ , entonces, el momento es la medida de correlación entre la imagen y la base polinomial. Los momentos han sido ampliamente utilizados para al análisis de imágenes como: reconocimiento de patrones [23, 24], clasificación de objetos [25], comparación de plantillas [26], detección de bordes [27], estimación de posición [28], visión por computadora [29] y compresión de datos [30]. En estas aplicaciones, los momentos geométricos y sus extensiones, momentos radiales y complejos, han sido importantes en la caracterización de formas y estructuras dentro de las imágenes y en la extracción de características invariantes a transformaciones geométricas.

Muchas familias de momentos se han definido en la literatura utilizando diferentes bases polinomiales clasificadas por la ortogonalidad de sus funciones base. Las familias de momentos no ortogonales se han usado para extracción de características de textura [31, 32]. Sin embargo, tienen el inconveniente de generar una representación redundante de información a partir de la imagen [33]. Además, la no ortogonalidad causa que la recuperación de la imagen sea computacionalmente costosa. Teague [34] presentó los momentos con base polinomial ortogonal, con la propiedad de mínima redundancia de información. Pertenecientes a esta clase, los momentos de Zernike y de Legendre han sido utilizados como detectores de características invariantes al ruido y rotación [35–37]. No obstante, las implementaciones prácticas de los momentos ortogonales continuos tienen dos principales debilidades. En primer lugar, para definirlos exactamente se necesita el cálculo de integrales que son aproximadas mediante sumas en el dominio discreto, lo que probablemente genere errores de discretización, poniendo en riesgo la ortogonalidad del conjunto de polinomios. En segundo lugar, los polinomios de Legendre y Zernike, por ejemplo, están definidos en una región específica del espacio, entre  $-1$  y  $1$  para Legendre y el círculo unitario para Zernike, por lo tanto, el espacio coordinado de la imagen debe ser transformado para su procesamiento, pudiendo cambiar la distribución espacial de sus píxeles o bien, realizando el

análisis sobre la imagen incompleta [38, 39].

Los momentos ortogonales discretos no tienen las limitantes antes mencionadas y conservan la mayoría de las propiedades útiles de los momentos ortogonales continuos. Estos momentos se definen a través de familias de polinomios ortogonales en el espacio de la imagen. Como resultado, el error inherente a la discretización de los polinomios continuos no existe, permitiendo una reconstrucción precisa de la imagen [33]. Las familias de polinomios discretos más utilizadas son Tchebichef [33] y Krawtchouk [40]. Sin embargo, las limitantes no dejan de aparecer, el cálculo de momentos de orden alto es susceptible a inestabilidades numéricas causadas por las ecuaciones de recurrencia con las que se calculan los coeficientes polinomiales [41], el resultado se traduce en errores numéricos acumulados por el aumento del orden [42]. Para evitar estos errores, Mukundan [43] define una normalización de los momentos de Tchebichef por un factor dependiente del orden mismo. A partir de este trabajo, diferentes normalizaciones se han propuesto para las familias de Krawtchouk, Hahn [44], Dual Hahn [45], Racah [46], Shmaliy [47], etcétera.

En este trabajo de investigación, se reportan nuevas contribuciones para el uso de Momentos Ortogonales Discretos como descriptores de textura, que a su vez, funcionen para una adecuada clasificación entre clases dentro de una misma base de datos. La propuesta inicial está fundamentada en el trabajo de Marcos y Cristóbal [48], ellos proponen el cálculo de una firma de textura basada en los Momentos Ortogonales de Tchebichef. Esta propuesta inicial también incluye el análisis local de las imágenes a través de ventanas traslapadas y el uso de otras bases polinomiales, como Krawtchouk, Dual Hahn y Shmaliy. Posteriormente, se propone el uso de los polinomios discretos bidimensionales como filtros para generar una descripción y clasificación basadas en el esquema de *Textons*. También se desarrolla un esquema que combina las firmas de textura con el esquema de *Textons*. Todas las propuestas fueron verificadas al variar diferentes parámetros como el tamaño de ventana, y por ende, el orden de los polinomios, la combinación de diferentes bases polinomiales, etc. Además, se utilizan técnicas de selección y transformación de características, con el fin de mejorar los resultados de clasificación. La validación de estas propuestas de descripción y clasificación fue aplicada en bases públicas de datos biomédicos adquiridos con diferentes técnicas de adquisición.



# 2

## Momentos y polinomios ortogonales discretos

Los momentos han sido utilizados en un número extenso de aplicaciones en el campo del procesamiento de imágenes. En 1961, Hu [49] introdujo el concepto de momentos invariantes a la posición, tamaño y orientación. Dudani *et al.* [50] usaron los momentos de Hu de tercer orden para el reconocimiento de aeronaves y los mismos descriptores también fueron utilizados por Casasent y Cheatham [51] para la detección de barcos. Markandey *et al.* [29] desarrollaron técnicas basadas en momentos de orden alto y tensores en robots. Sin embargo, estos momentos, también conocidos como momentos geométricos, no son ortogonales, por ende, si se pretende la reconstrucción de las imágenes a partir de ellos, resulta complicada e imprecisa por la redundancia de información y la definición de un criterio adecuado para la realización de esta tarea [52].

Teague [34] propuso utilizar los polinomios y momentos ortogonales continuos de Zernike y Legendre para la reconstrucción de imágenes. Los momentos de Zernike son capaces de almacenar la información con una redundancia mínima, son invariantes a la rotación y han sido utilizados para reconocimiento y estimación de la posición de objetos tridimensionales, reconocimiento de formas, detección de bordes en objetos tridimensionales, etcétera. Los momentos de Legendre cuentan con muy poca redundancia en la información, poca sensibilidad al ruido y han sido utilizados para el reconocimiento de símbolos árabes escritos a mano. Sin embargo, un problema común de los momentos continuos es el error de discretización porque aumenta conforme al orden de los polinomios. Además, existe la necesidad de proyectar la imagen en el espacio de los polinomios de Zernike y Legendre, situación que puede cambiar la distribución original de los píxeles en la imagen o que su análisis se realice sobre la imagen incompleta.

Recientemente, como solución a estos inconvenientes, se ha propuesto utilizar conjuntos de momentos ortogonales en el dominio discreto. La definición matemática de estos momentos evidencia que no

hay necesidad de alguna aproximación numérica porque la base polinomial es ortogonal en el dominio discreto, coincidiendo así con el espacio coordenado de la imagen. Por otra parte, la reconstrucción de imágenes es precisa y de fácil implementación porque no hay redundancia en la información.

En este capítulo se explica a profundidad qué son los momentos ortogonales. En primer lugar, se definen las propiedades de los momentos ortogonales continuos clásicos (Hermite, Legendre y Jacobi) y radiales (Zernike, pseudo-Zernike y Fourier-Mellin). Después, los momentos ortogonales discretos son definidos y ejemplificados a través de las familias de bases polinomiales de Tchebichef, Krawtchouk y Dual Hahn. Los momentos de Shmaliy son presentados de manera separada porque tienen un origen distinto a las otras familias de momentos discretos. Finalmente, se muestran las implementaciones recursivas de todas las familias discretas mencionadas anteriormente.

### 2.1 Polinomios y Momentos Ortogonales Continuos

En esta sección se presentarán familias de momentos definidas en el espacio continuo. En primer lugar, se presentan antecedentes importantes para la correcta definición de las bases polinomiales. Posteriormente, se definen las condiciones y propiedades de las familias de polinomios ortogonales clásicos. A continuación, se presentan las familias de polinomios ortogonales radiales y su aplicación en la obtención de momentos ortogonales basados en estos polinomios. Después, se definen las familias de polinomios y momentos ortogonales discretos que son las herramientas principales de este trabajo de investigación. Finalmente, se muestran las ecuaciones recursivas utilizadas para la implementación de las familias ortogonales discretas.

#### 2.1.1 ANTECEDENTES

##### ORTOGONALIDAD Y ORTONORMALIDAD

Un sistema de polinomios  $\{P_n(x) : n \in \mathbb{N}\}$  donde  $P_n(x)$  es un polinomio de grado  $n$  es un sistema de polinomios ortogonales con respecto a una medida real positiva  $\phi$  si tenemos la siguiente relación de ortogonalidad:

$$\int_S P_n(x)P_m(x)d\phi(x) = d_n^2\delta_{nm}, \quad (2.1)$$

donde la  $\delta_{nm}$  es una delta de Dirac

$$\delta_{nm} = \begin{cases} +\infty & \text{si } n = m \\ 0 & \text{si } n \neq m, \end{cases} \quad (2.2)$$

$S$  es el soporte de la medida  $\phi$  y  $d_n \in \mathbb{R}$  es la norma de la base polinomial, además, tanto  $\phi$  como  $d_n$  son distintas de cero. Si la norma es  $d_n = 1$ , decimos que el sistema es ortonormal.

La medida  $\phi$  usualmente tiene densidad  $w(x)$  diferenciable con respecto a una medida continua  $dx$

o a una medida discreta con pesos  $w_i$  en los puntos  $x_i$ . La relación anterior se convierte entonces en:

$$\int_S P_n(x)P_m(x)w(x)dx = d_n^2\delta_{nm} \quad (2.3)$$

para el caso continuo y

$$\sum_{i=0}^M P_n(x_i)P_m(x_i)w_i = d_n^2\delta_{nm} \quad (2.4)$$

para el caso discreto en donde  $M$  puede llegar a ser infinito. Cabe resaltar que para el caso discreto es una delta diferente que se denomina delta de Kronecker

$$\delta_{nm} = \begin{cases} 1 & \text{si } n = m \\ 0 & \text{si } n \neq m. \end{cases} \quad (2.5)$$

#### ECUACIÓN DIFERENCIAL LINEAL DE SEGUNDO ORDEN

La ecuación diferencial general lineal de segundo orden en el intervalo definido entre  $-\infty$  a  $\infty$ ,  $Int = (-\infty, \infty)$ , es:

$$p(x)u''(x) + q(x)u'(x) + r(x)u(x) = f(x), \quad (2.6)$$

con su correspondiente ecuación homogénea:

$$p(x)u''(x) + q(x)u'(x) + r(x)u(x) = 0 \quad (2.7)$$

y con el operador diferencial asociado:

$$D = p(x)\frac{d^2}{dx^2} + q(x)\frac{d}{dx} + r(x). \quad (2.8)$$

#### 2.1.2 POLINOMIOS ORTOGONALES CLÁSICOS

Las familias de polinomios ortogonales clásicos se pueden obtener a través de funciones generadoras del operador diferencial  $D$ , visto en la ecuación 2.8. Este operador diferencial  $D$  debe cumplir con la condición de simetría con respecto al peso  $w$ , es decir, si:

$$(Du, v) = (u, Dv), \quad (2.9)$$

donde  $u$  y  $v$  son un par de funciones que se anulan fuera del intervalo cerrado y acotado por definición entre los valores  $a$  y  $b$ ,  $Int = [a, b]$ , con el producto interno definido como:

$$(u, v) = (u, v)_w = \int_a^b u(x)v(x)w(x)dx \quad (2.10)$$

y con el peso  $w > 0$ . Además, el operador diferencial  $D$  simétrico con respecto al peso  $w$  debe tener la siguiente forma:

$$D = p \frac{d^2}{dx^2} + \frac{(pw)'}{w} \frac{d}{dx} + r. \quad (2.11)$$

A partir de estas definiciones y condiciones se precisa la extensión de la condición de simetría a una clase más larga de funciones y para lograrlo es necesario aplicar condiciones de frontera.

Supongamos que las funciones  $w, w', p, p', q$  y  $r$  se extienden como funciones continuas en el intervalo cerrado  $Int = [a, b]$ . Además,  $u, u', v$  y  $v'$  también son continuas en el mismo intervalo. Ahora, si el operador  $D$  es simétrico, entonces:

$$(Du, v) - (u, Dv) = (pwu'v - pwuv') \Big|_a^b. \quad (2.12)$$

Si  $pw$  se anula en ambos puntos de la frontera entonces no necesitamos condiciones adicionales en esa frontera, si no es así, se añaden condiciones sobre las funciones  $u$  y  $v$ . De manera similar, si el intervalo es no acotado y semi infinito,  $Int = [a, \infty)$ , se deben imponer condiciones en  $x = a$ , a menos que  $pw = 0$  en  $x = a$ . Una función  $u$  que no sea idénticamente cero es una función generadora para el operador  $D$  con valor generador  $-\lambda$  si:

$$Du + \lambda u = 0. \quad (2.13)$$

Si  $u_1$  y  $u_2$  son funciones generadoras con diferentes valores generadores  $-\lambda_1$  y  $-\lambda_2$  entonces:

$$-\lambda_1(u_1, u_2) = (Du_1, u_2) = (u_1, Du_2) = -\lambda_2(u_1, u_2), \quad (2.14)$$

además,  $u_1$  y  $u_2$  son ortogonales si  $(u_1, u_2) = 0$ .

Ahora, se supondrá la existencia de polinomios de grados 0, 1 y 2 que son funciones generadoras de  $D$ , es decir, se asume que el espacio de los polinomios de grado  $\leq k, k \in \{0, 1, 2\}$ , está en el dominio de  $D$ .

En primer lugar, se hará la suposición de que el peso  $w$  tiene integral finita:

$$\int_a^b w(x) dx < \infty. \quad (2.15)$$

Aplicando el operador  $D$ , cuya condición de simetría con respecto al peso se mostró en la ecuación 2.11, a la función constante  $u_0(x) = 1$  obtenemos  $Du_0 = r$ , así que  $r$  debe ser una constante y podemos considerar que  $r = 0$ . De esta forma, aplicando  $D$  a  $u_1(x) = x$  obtenemos:

$$Du_1 = \frac{(pw)'}{w} = p' + p \frac{w'}{w}, \quad (2.16)$$

esta expresión debe ser un polinomio de grado máximo 1. Ahora, si se considera  $u_2(x) = \frac{1}{2}x^2$ :

$$Du_2 = p + x \left( p' + p \frac{w'}{w} \right), \quad (2.17)$$

por lo que  $Du_2$  debe ser un polinomio de grado máximo 2, entonces, el polinomio  $p$  es de grado máximo 2.

La condición de simetría requiere que:

$$(Du, v) - (u, Dv) = \int_a^b [pw(u'v - uv')] = 0, \quad (2.18)$$

para toda  $u, v$  definida en el intervalo. Como se ha mencionado anteriormente, una condición necesaria para la simetría es que  $pw \rightarrow 0$  en cada punto final del intervalo. Resumiendo, se buscan soluciones polinomiales  $p_n(x)$  que sean funciones generadoras de un operador diferencial de la forma:

$$D = p \frac{d^2}{dx^2} + q \frac{d}{dx} = \frac{1}{w} \left( pw \frac{d}{dx} \right)', \quad (2.19)$$

donde  $\text{orden}(p) \leq 2$ ,  $\text{orden}(q) \leq 1$ ,  $Dp_n = \lambda_n p_n$  y

$$(pw)' = qw. \quad (2.20)$$

A este tipo de operadores diferenciales se les denomina de tipo hipergeométrico o de tipo Sturm-Liouville. A la ecuación 2.20 se le denomina como la ecuación de Pearson.

Dado que  $p$  es un polinomio de grado máximo 2, después de normalizaciones (mapeos afines de la recta, multiplicación del peso, el operador y los polinomios por constantes) se obtiene que existen 5 tipos de familias de polinomios ortogonales que son funciones generadoras del operador diferencial hipergeométrico  $D$  (ecuación 2.19) dependiendo del grado y las raíces de  $p$ :

1.  $p$  constante: El polinomio  $w'/w$  debe ser de grado máximo 1. Resolviendo la ecuación diferencial, se tiene que  $w(x) = e^h$ , donde  $h$  es un polinomio de grado máximo 2. Las condiciones de frontera sólo se cumplen para  $w(x) = e^{-x^2}$ , cuyo intervalo abierto es  $Int = (-\infty, \infty)$ . Entonces, el polinomio  $q = w'/w = -2x$  es un polinomio de grado 1, por lo tanto, el operador  $D$  en este caso es:

$$D = \frac{d^2}{dx^2} - 2x \frac{d}{dx}. \quad (2.21)$$

Por cada polinomio de grado  $n$ , el operador  $D$  lo lleva a otro polinomio del mismo grado. El valor generador para este caso es  $\lambda_n = -2n$ . Los polinomios ortogonales con respecto al peso  $e^{-x^2}$  son los polinomios de Hermite  $H_n(x)$  y cumplen con  $DH_n = -2nH_n$ . Una vez que el peso  $w(x)$  es normalizado, éste corresponde a la distribución normal o Gaussiana.

2.  $p$  lineal: Si se normaliza a  $p(x) = x$  y se tiene que  $w'/w = b + \alpha/x$ . Al resolver se tiene que, salvo constantes,  $w(x) = x^\alpha e^{bx}$ . Las condiciones de frontera para  $pw = xw = x^{\alpha+1} e^{bx}$  con el intervalo no acotado  $Int = [0, \infty)$  con  $b < 0$  y  $\alpha > -1$ . Con lo cual, tras normalizar,  $w(x) = x^\alpha e^{-x}$ ,  $\alpha > -1$ . El polinomio  $q = (xw)'/w = \alpha + 1 - x$  es un polinomio de grado 1. Por lo tanto, el operador  $D$  es:

$$D = x \frac{d^2}{dx^2} + (\alpha + 1 - x) \frac{d}{dx}. \quad (2.22)$$

Por cada polinomio de grado  $n$ , el operador  $D$  lo lleva a otro polinomio del mismo grado. El valor generador para este caso, igualando coeficientes, es  $\lambda_n = -n$ . Los polinomios ortogonales con respecto al peso  $x^\alpha e^{-x}$  son los polinomios de Laguerre  $L_n^\alpha(x)$  y cumplen con  $DL_n^\alpha = -nL_n^\alpha$ . Una vez que el peso  $w(x)$  es normalizado, éste corresponde a la distribución Gamma y como caso particular, para  $\alpha = 0$ , corresponde a la distribución exponencial.

3.  $p$  cuadrático con raíces reales distintas: Si se normaliza a  $p(x) = 1 - x^2$  y se tiene que  $w'/w = \beta/(1+x) + \alpha/(1-x)$ . Al resolver se tiene que, salvo constantes,  $w(x) = (1-x)^\alpha(1+x)^\beta$ . Las condiciones de frontera para  $pw = (1-x^2)w = (1-x)^{\alpha+1}(1+x)^{\beta+1}$  con el intervalo acotado cerrado  $Int = [-1, 1]$  con  $\alpha, \beta > -1$ . El polinomio  $q = ((1-x^2)w)'/w = -(\alpha + \beta + 2)x + \beta - \alpha$  es un polinomio de grado 1. Por lo tanto, el operador  $D$  es:

$$D = (1-x^2)\frac{d^2}{dx^2} + [-(\alpha + \beta + 2)x + \beta - \alpha]\frac{d}{dx}. \quad (2.23)$$

Por cada polinomio de grado  $n$ , el operador  $D$  lo lleva a otro polinomio del mismo grado. El valor generador para este caso, igualando coeficientes, es  $\lambda_n = -n(n + \alpha + \beta + 1)$ . Los polinomios ortogonales con respecto al peso  $(1-x)^\alpha(1+x)^\beta$  son los polinomios de Jacobi  $J_n^{\alpha,\beta}(x)$  y cumplen con  $DJ_n^{\alpha,\beta} = -n(n + \alpha + \beta + 1)J_n^{\alpha,\beta}$ .

4.  $p$  cuadrático con raíces complejas distintas: se normaliza a  $p(x) = 1 + x^2$  y se tiene que  $w'/w = (-2ax + b)/(1 + x^2)$ . Al resolver se tiene que, salvo constantes,  $w(x) = (1 + x^2)^{-a}e^{btan^{-1}(x)}$ . Las condiciones de frontera y la positividad e integrabilidad del peso  $w$  obligan a que el intervalo de ortogonalidad sea  $Int = (-\infty, \infty)$  y que  $a > 1/2, b \in \mathbb{R}$ . El polinomio  $q = ((1+x^2)w)'/w = 2x(1-a) + b$  es un polinomio de grado 1. Por lo tanto, el operador  $D$  es:

$$D = (1+x^2)\frac{d^2}{dx^2} + [2x(1-a) + b]\frac{d}{dx}. \quad (2.24)$$

Por cada polinomio de grado  $D$ , el operador  $L$  lo lleva a otro polinomio del mismo grado. El valor generador para este caso, igualando coeficientes, es  $\lambda_n = n(n + 1 - 2a)$ . Solamente se pueden construir polinomios bajo la condición  $n < 2a - 1$ , además, bajo esta condición, el valor generador  $\lambda_n$  es negativo. Los polinomios Romanovski se pueden construir bajo estas condiciones.

5.  $p$  cuadrático con raíz doble: se normaliza a  $p(x) = x^2$  y se tiene que  $w'/w = -a/x + b/x^2$ . Al resolver se tiene que, salvo constantes,  $w(x) = x^{-a}e^{-b/x}$ . Las condiciones de frontera y la positividad e integrabilidad del peso  $w$  obligan a que el intervalo de ortogonalidad sea  $Int = [0, \infty)$  y que  $a > -1, b \geq 0$ . El polinomio  $q = ((x^2w)'/w = x(2-a) + b$  es un polinomio de grado 1. Por lo tanto, el operador  $D$  es:

$$D = x^2\frac{d^2}{dx^2} + [x(2-a) + b]\frac{d}{dx}. \quad (2.25)$$

El valor generador para este caso, igualando coeficientes, es  $\lambda_n = n(n + 1 - a)$ . De nuevo, sólo hay un número finito de polinomios ortogonales y la condición es  $n < a - 1$ .

Funciones	Hermite $H_n(x)$	Laguerre $L_n^\alpha(x)$	Jacobi $J_n^{\alpha,\beta}(x)$
$Int = [a, b]$	$(-\infty, \infty)$	$[0, \infty)$	
$w(x)$	$e^{-x^2}$	$x^\alpha e^{-x}$	$(1-x)^\alpha(1+x)^\beta$
$\sigma(x)$	1	$x$	$1-x^2$
$\tau(x)$	$-2x$	$1+\alpha-x$	$-(\alpha+\beta+2)x+\beta-\alpha$
$\lambda_n$	$2n$	$n$	$-n(n+\alpha+\beta+1)$

**Tabla 2.1:** Funciones de los polinomios ortogonales clásicos: Hermite, Laguerre y Jacobi.

Muchos problemas de matemáticas aplicadas y teóricas así como problemas de física conducen a ecuaciones de la forma:

$$\sigma(x)u'' + \tau(x)u' + \lambda u = 0, \quad (2.26)$$

donde  $\lambda$  es una constante y  $\sigma(x)$  y  $\tau(x)$  son polinomios de, al menos, segundo y primer grado, respectivamente. A esta ecuación se le conoce como ecuación diferencial del tipo hipergeométrico y se obtiene a través de las ecuaciones 2.11 y 2.13 e igualando:

$$\begin{aligned} \sigma &= p, \\ \tau &= q = \frac{(pw)'}{w} \end{aligned}$$

y

$$\lambda = r.$$

Un aspecto que hay que tener en cuenta es que  $\lambda$  no sólo es una constante sino que además es el valor generador del operador  $D$ . Por otro lado, la relación que existe entre  $\sigma(x)$  y  $\tau(x)$  en la ecuación 2.26 está dada por la ecuación de Pearson, ecuación 2.20, que indica que  $\tau(x)$  es derivada de  $\sigma(x)$ , siempre y cuando ambas estén multiplicadas por el peso  $w$  [53] [54].

En conclusión tenemos que, salvo normalizaciones, las únicas familias de polinomios ortogonales que son funciones generadoras de un operador diferencial de segundo orden que obedece a la ecuación 2.11, es decir, un operador diferencial simétrico con respecto a un peso  $w$  positivo soportado dentro de la recta real, son las familias de Hermite, Laguerre y Jacobi. A estas familias se les suele denominar polinomios ortogonales clásicos. En la Tabla 2.1 se muestra un resumen de las características de las familias clásicas.

### 2.1.3 POLINOMIOS ORTOGONALES RADIALES

En el reconocimiento de patrones y aplicaciones similares se requieren descriptores robustos que sean invariantes a transformaciones rotacionales [35, 55]. Por esta razón, las primeras aplicaciones de

los momentos ortogonales se definen en forma radial, en coordenadas polares  $(r, \theta)$  y en los cuales, la rotación es transformada en un desplazamiento.

Los momentos radiales más conocidos son: los momentos de Zernike [56], pseudo-Zernike [57] y Fourier-Mellin [58]. Estos momentos y polinomios radiales están definidos en un círculo unitario continuo  $U = \{(r, \theta) | 0 \leq r \leq 1, 0 \leq \theta < 2\pi\}$ . Además de tener la característica de invarianza a la rotación, estos momentos son ortogonales, por lo tanto, tienen poca redundancia de información y buena capacidad de reconstrucción. Consecuentemente, las señales pueden ser reconstruidas usando un conjunto (posiblemente infinito) de estos momentos [59].

Para aplicaciones de procesamiento de imágenes, los momentos radiales ortogonales son discretizados geoméricamente en el círculo unitario y numéricamente en los valores de las funciones polinomiales que definen a los momentos. Los momentos radiales ortogonales discretizados han mostrado su eficiencia en algunas aplicaciones de procesamiento de imágenes, como análisis de imágenes [34, 60], reconocimiento de patrones [61], clasificación de texturas [62], indización de imágenes [63], y la estimación de la orientación de objetivos [64]. También se ha demostrado que son robustos al ruido [60] y se utilizan como base para construir momentos con nuevas propiedades [37] [65].

Sea  $I(r, \theta)$  la imagen en coordenadas polares  $(r, \theta)$  definida en un círculo unitario continuo  $U = \{(r, \theta) | 0 \leq r \leq 1, 0 \leq \theta < 2\pi\}$ . Los momentos radiales ortogonales de orden  $p$  y repetición  $q$  de la imagen  $I(r, \theta)$  se definen como:

$$v_{pq} = n_{pq} \int_0^{2\pi} \int_0^1 R_{pq}(r) e^{-iq\theta} I(r, \theta) r dr d\theta, \quad (2.27)$$

donde  $p \in \mathbb{Z}_+$ ,  $q = -p, \dots, p$ ,  $n_{pq}$  es el factor de normalización,  $R_{pq}$  es un polinomio radial unidimensional, y  $e^{-iq\theta}$  es la parte angular o fase de la función polinomial respectiva. Para calcular los momentos, la imagen  $I(r, \theta)$  debe ser escalada y ajustada, de tal manera que sea contenida en el espacio del círculo unitario  $U$ .

La invarianza a la rotación significa que, aunque se rote la imagen  $I(r, \theta)$ , la magnitud del momento radial no cambiará. Por ejemplo, sea  $v_{pq}^\phi$  el momento radial de la imagen rotada  $I(r, \theta - \phi)$ , entonces  $|v_{pq}^\phi| = |v_{pq}|$  [35].

Los momentos de Zernike, pseudo-Zernike y Fourier-Mellin difieren entre sí en la base polinomial,  $R_{pq}$ , y en los pares factibles de  $p$  y  $q$ . Para los momentos de Zernike:

$$R_{pq}(r) = \sum_{s=0}^{(p-|q|)/2} (-1)^s \frac{(p-s)! r^{p-2s}}{s! \left(\frac{p-2s+|q|}{2}\right)! \left(\frac{p-2s-|q|}{2}\right)!}, \quad (2.28)$$

donde  $p - |q|$  es par y  $|q| \leq p$ .

Para los momentos pseudo-Zernike:

$$R_{pq}(r) = \sum_{s=0}^{p-|q|} (-1)^s \frac{(2p+1-s)! r^{p-s}}{s! (p-s-|q|)! (p-s+|q|+1)!}, \quad (2.29)$$

donde  $|q| \leq p$ .

Orden ( $p$ )	Repetición ( $q$ )	Número de momentos
0	0	1
1	-1, 1	3
2	-2, 0, 2	6
3	-3, -1, 1, 3	10

**Tabla 2.2:** Posibles repeticiones y número de momentos de orden 0 al 3 para los momentos de Zernike.

Orden ( $p$ )	Repetición ( $q$ )	Número de momentos
0	0	1
1	-1, 0, 1	4
2	-2, -1, 0, 1, 2	9
3	-3, -2, -1, 0, 1, 2, 3	16

**Tabla 2.3:** Posibles repeticiones y número de momentos de orden 0 al 3 para los momentos pseudo-Zernike y Fourier-Mellin.

Para los momentos de Fourier-Mellin:

$$R_{pq}(r) = R_p(r) = \sum_{s=0}^p (-1)^s \frac{(p+s+1)! r^s}{(p-s)! s! (s+1)!}, \quad (2.30)$$

donde  $|q| \leq p$ . Las posibles repeticiones y el número de momentos con orden  $p = \{0, 1, 2, 3\}$  se muestran en la Tabla 2.2 para los momentos de Zernike y en la Tabla 2.3 para los momentos pseudo-Zernike y Fourier-Mellin.

Sin embargo, la discretización compromete la exactitud del cálculo de los momentos radiales ortogonales. El error de discretización ha sido analizado y se identificaron dos fuentes principales de error [66, 67]. La primera fuente de error es la aproximación geométrica del círculo unitario a un conjunto finito de píxeles y la otra fuente proviene del muestreo de los momentos radiales ortogonales, ya que, como se puede observar en la ecuación 2.27, la definición está en el dominio continuo.

## 2.2 Polinomios y Momentos Ortogonales Discretos

Si la ecuación diferencial 2.26 es reemplazada por una ecuación en diferencias, se pueden conservar las propiedades principales de los polinomios ortogonales clásicos, pero ahora definidos en variable

discreta. El caso más simple es cuando la ecuación 2.26 es reemplazada por:

$$\begin{aligned} \tilde{\sigma}(x) \frac{1}{h} \left[ \frac{y(x+h) - y(x)}{h} - \frac{y(x) - y(x-h)}{h} \right] + \\ \frac{\tilde{\tau}(x)}{2} \left[ \frac{y(x+h) - y(x)}{h} + \frac{y(x) - y(x-h)}{h} \right] + \lambda y(x) = 0. \end{aligned} \quad (2.31)$$

La ecuación 2.31 se aproxima a la ecuación 2.26 si  $\Delta x = h$ , es decir, cuando  $x$  se define en una rejilla constante. Los polinomios discretos clásicos tales como los Charlier, Meixner, Krawtchouk, Tchebichef y Hahn son soluciones polinomiales a la ecuación 2.31.

Obtenemos una mayor generalización cuando en la ecuación 2.26 se realiza el siguiente cambio de variable  $\Delta x = x(s+h) - x(s)$ , es decir, la ecuación 2.31 ahora está definida en una rejilla no uniforme [53]. Los polinomios ortogonales discretos en una rejilla no uniforme se explican con mayor profundidad en la Sección 2.3 y tienen gran importancia en aplicaciones de sistemas integrales cuánticos, teoría de campos cuánticos y física estadística [68–72].

### 2.2.1 POLINOMIOS Y MOMENTOS DE TCHEBICHEF

Los polinomios discretos de Tchebichef de orden  $n$  se definen como:

$$t_n(x) = (1 - N)_n {}_3F_2(-n, -x, 1 + n; 1, 1 - N; 1), \quad (2.32)$$

donde  $x, n = \{0, 1, 2, \dots, N - 1\}$ . La función hipergeométrica  ${}_3F_2$  se establece como:

$${}_3F_2(a_1, a_2, a_3; b_1, b_2; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{z^k}{k!}, \quad (2.33)$$

donde el símbolo de Pochhammer  $(a)_k$  es:

$$(a)_k = a(a+1) \cdots (a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}, \quad (2.34)$$

donde  $\Gamma(n) = (n-1)!$ . Los polinomios de Tchebichef, satisfacen la propiedad de ortogonalidad de la siguiente forma:

$$\sum_{x=0}^{N-1} t_n(x) t_m(x) = \rho(n, N) \delta_{nm}, \quad (2.35)$$

donde los polinomios de Tchebichef de orden  $n$  y  $m$  son  $t_n$  y  $t_m$ , respectivamente, la  $\delta_{nm}$  es una delta de Kronecker, ecuación 2.5, y el factor de normalización  $\rho(n, N)$  es definido como:

$$\rho(n, N) = \frac{N(N^2 - 1^2)(N^2 - 2^2) \cdots (N^2 - n^2)}{2n + 1} = (2n)! \binom{N+n}{2n+1}. \quad (2.36)$$

La normalización de los momentos garantiza que no habrá variaciones considerables en el rango

dinámico de los valores de los momentos ni inestabilidades numéricas para valores grandes de  $N$  [33]. Por lo tanto, los polinomios normalizados de Tchebichef,  $\hat{t}_n$ , se definen como:

$$\hat{t}_n(x, N) = \frac{t_n(x, N)}{\sqrt{\rho(n, N)}}. \quad (2.37)$$

Entonces, la ecuación 2.37 satisface la propiedad de ortogonalidad de la siguiente manera:

$$\sum_{x=0}^{N-1} \hat{t}_n(x) \hat{t}_m(x) = \delta_{nm}. \quad (2.38)$$

En la Figura 2.1(a) muestra a los polinomios unidimensionales de Tchebichef desde el orden 0 al orden 5, mientras que la Figura 2.1(b) muestra un diagrama con los polinomios bidimensionales de Tchebichef del orden 0 al 8.

Finalmente, los **Momentos de Tchebichef**,  $T_{mn}$ , de orden  $m + n$  de la imagen  $I(x, y)$  se definen como:

$$T_{mn} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \hat{t}_n(x) \hat{t}_m(y) I(x, y), \quad (2.39)$$

donde  $N$  y  $M$  son el tamaño de la imagen en el eje  $X$  y en el eje  $Y$ , respectivamente.

### 2.2.2 POLINOMIOS Y MOMENTOS DE KRAWTCHOUK

Los polinomios de Krawtchouk, propuestos en [73], están asociados a la distribución binomial. La definición de los polinomios clásicos de Krawtchouk de orden  $n$  es:

$$K_n(x; p, N) = \sum_{k=0}^N a_{k,n,p} x^k = {}_2F_1(-n, -x; -N; \frac{1}{p}), \quad (2.40)$$

donde  $x, n = \{0, 1, 2, \dots, N, N > 0\}$ ,  $p \in (0, 1)$ .  ${}_2F_1$  es la función hipergeométrica:

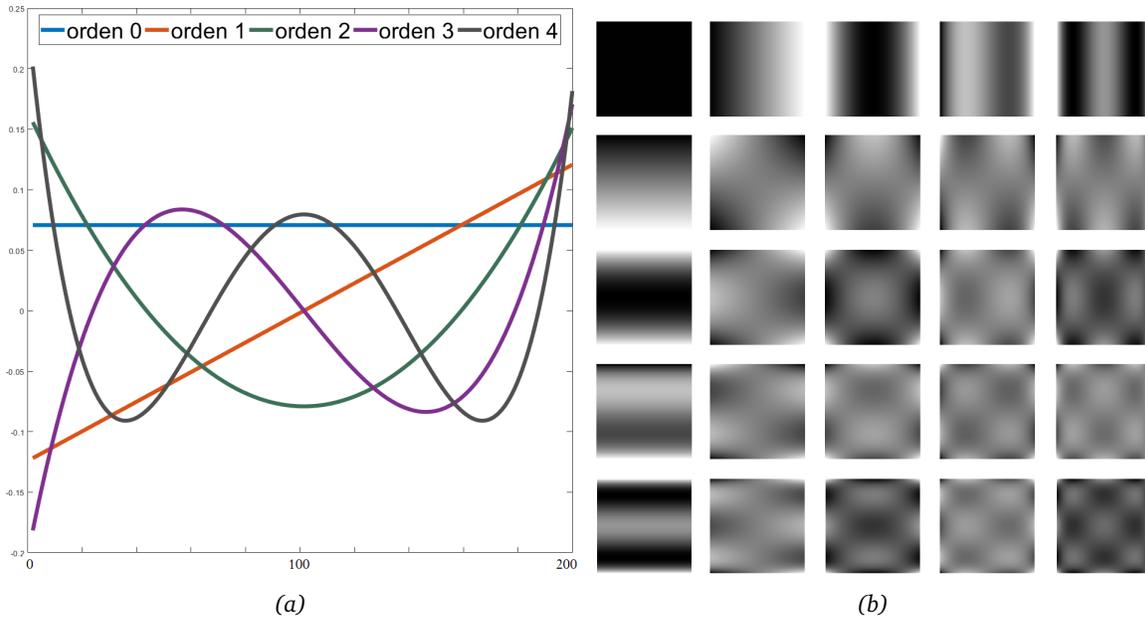
$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} \quad (2.41)$$

y  $(a)_k$  es el símbolo de Pochhammer definido en la ecuación 2.34. Los polinomios de Krawtchouk forman un conjunto completo de funciones base discretas con una función de peso:

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \quad (2.42)$$

y ese conjunto satisface la condición de ortogonalidad:

$$\sum_{x=0}^N w(x; p, N) K_n(x; p, N) K_m(x; p, N) = \rho(n; p, N) \delta_{nm}, \quad (2.43)$$



**Figura 2.1:** (a) Polinomios discretos de Tchebichef de orden 0 a 4 con  $N = 200$ . (b) Polinomios discretos de Tchebichef bidimensionales de orden 0 a 8, el orden se considera diagonalmente.

donde  $n, m = \{1, 2, \dots, N\}$ ,  $\delta_{nm}$  es una delta de Kronecker (definida en la ecuación 2.5) y el factor de normalización  $\rho$  es:

$$\rho(n; p, N) = (-1)^n \left( \frac{1-p}{p} \right)^n \frac{n!}{(-N)_n}. \quad (2.44)$$

El método más utilizado para disminuir las fluctuaciones numéricas para el cómputo de los polinomios es la normalización, tal y como vimos en los polinomios de Tchebichef en la Sección 2.2.1. Sin embargo, en el caso de los polinomios de Krawtchouk, la normalización no es suficiente para garantizar la estabilidad de la implementación. Entonces, además de incluir el factor de normalización (ecuación 2.44), la raíz cuadrada de la función de peso (ecuación 2.42) se introduce como factor de ponderación. Por lo tanto, los polinomios discretos de Krawtchouk ponderados,  $\hat{K}_n$ , [40] son definidos de la siguiente manera:

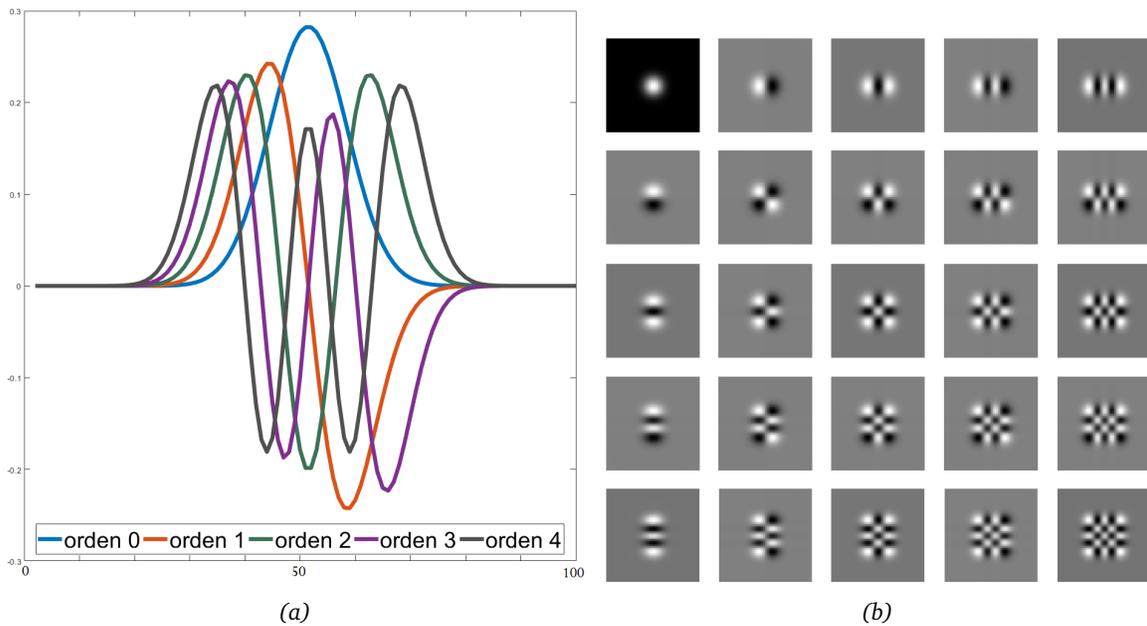
$$\hat{K}_n(x; p, N) = K_n(x; p, N) \sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}}. \quad (2.45)$$

Ahora, la condición de ortogonalidad se satisface de la siguiente forma:

$$\sum_{x=0}^N \hat{K}_n(x; p, N) \hat{K}_m(x; p, N) = \delta_{nm}. \quad (2.46)$$

Finalmente, los valores de los polinomios normalizados de Krawtchouk están entre  $-1$  y  $1$ . En la Figura 2.2(a) se muestran los polinomios discretos en una dimensión de Krawtchouk desde el orden 0 al orden 4. En la Figura 2.2(b) se observan los polinomios discretos bidimensionales de Krawtchouk del orden 0 al 8.

Los **Momentos de Krawtchouk**,  $Q_{mn}$ , de orden  $m+n$  de la imagen  $I(x, y)$  se calculan con la siguiente



**Figura 2.2:** (a) Polinomios de Krawtchouk de orden 0 a 4 con  $N = 100$ . (b) Polinomios de Krawtchouk bidimensionales de orden 0 al 8.

expresión:

$$Q_{mn} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \hat{K}_n(x; p_1, N-1) \hat{K}_m(y; p_2, M-1) I(x, y), \quad (2.47)$$

donde  $N$  y  $M$  son el tamaño de la imagen en el eje  $X$  y  $Y$ , respectivamente.

Los momentos de Krawtchouk tienen la ventaja de ser capaces de extraer características locales de cualquier región de interés dentro de la imagen al variar el parámetro  $p$ . En la Figura 2.3, están las gráficas de los polinomios ponderados de Krawtchouk bajo diferentes condiciones. En ellas se observa que si  $p$ , que parte de un valor de 0.5, se desvía  $\Delta p$ , es decir, si  $p = 0.5 + \Delta p$ , los polinomios ponderados de Krawtchouk son desplazados en aproximadamente  $N\Delta p$ . La dirección del corrimiento depende del signo de  $\Delta p$ , de tal manera que los polinomios se desplazan en dirección  $+x$  cuando  $\Delta p$  es positivo y viceversa. Esta propiedad es crucial para la extracción de características en regiones de interés porque indica que podemos posicionarnos en esa región al variar los parámetros  $p_1$  y  $p_2$ .

### 2.3 Polinomios ortogonales de rejilla no uniforme

Los polinomios ortogonales de rejilla no uniforme contienen más parámetros debido al dominio en el que están definidos, estos parámetros otorgan mayor flexibilidad para su aplicación en la descripción de imágenes. Además, tampoco hay necesidad de normalización espacial porque no hay error de discretización. Las familias de Tchebichef y Krawtchouk son casos especiales de estos polinomios, que son más generales y manipulables [53].

Los polinomios discretos ortogonales de rejilla no uniforme se definen usando una malla variable y

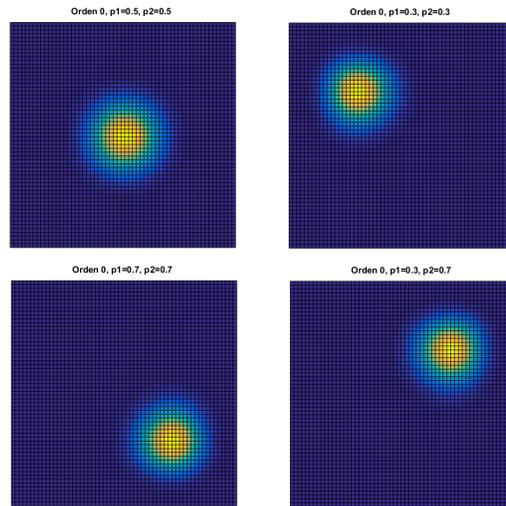


Figura 2.3: Polinomios de Krawtchouk bidimensionales con variación en el parámetro  $p$  de orden 0.

reemplazando la ecuación 2.26 por la siguiente ecuación en diferencias hipergeométrica de segundo orden para  $x = x(s)$ :

$$\tilde{\sigma}(x(s)) \frac{\Delta}{\Delta x \left(s - \frac{1}{2}\right)} \left[ \frac{\nabla y(s)}{\nabla x(s)} \right] + \frac{\tilde{\tau}(x(s))}{2} \left[ \frac{\Delta y(s)}{\Delta x(s)} + \frac{\nabla y(s)}{\nabla x(s)} \right] + \lambda y(s) = 0, \quad (2.48)$$

donde

$$\nabla g(s) = g(s) - g(s-1), \quad \Delta g(s) = g(s+1) - g(s) \quad (2.49)$$

son las diferencias finitas anterior y posterior,  $\lambda$  es una constante y  $\tilde{\sigma}(x)$  y  $\tilde{\tau}(x)$  son polinomios de, al menos, segundo y primer grado, respectivamente.

Entonces, resulta conveniente reescribir la ecuación 2.48 de la siguiente manera:

$$\sigma(s) \frac{\Delta}{\Delta x \left(s - \frac{1}{2}\right)} \left[ \frac{\nabla y(s)}{\nabla x(s)} \right] + \tau(s) \left[ \frac{\Delta y(s)}{\Delta x(s)} \right] + \lambda y(s) = 0 \quad (2.50)$$

con

$$\sigma(s) = \tilde{\sigma}(x(s)) - \frac{1}{2} \tilde{\tau}(x(s)) \left[ \Delta x \left(s - \frac{1}{2}\right) \right], \quad (2.51)$$

donde  $\tau(s) = \tilde{\tau}(x(s))$ .

Las soluciones polinomiales  $y_n(x(s)) \equiv P_n(s)$  de la ecuación 2.50 son determinadas únicamente a través de la diferencia análoga de la fórmula de Rodrigues, hasta un factor de normalización  $B_n$  [53]. Entonces, estas soluciones quedan definidas de la siguiente manera:

$$\begin{aligned} P_n(s) &= \frac{B_n}{\rho(s)} \nabla_n^{(n)} [\rho_n(s)], \\ &= \frac{B_n}{\rho(s)} \frac{\nabla}{\nabla x_1(s)} \cdots \frac{\nabla}{\nabla x_{n-1}(s)} \frac{\nabla}{\nabla x_n(s)} [\rho_n(s)], \end{aligned} \quad (2.52)$$

Nombre	Notación	Rejilla	$s_{min}$	$s_{max}$	$\rho(s)$
Dual Hahn	$w_n^{(c)}(s, a, b)$	$x(s) = s(s + 1)$	$a$	$b$	$\frac{\Gamma(a+s+1)\Gamma(c+s+1)}{\Gamma(s-a+1)\Gamma(b-s)\Gamma(b+s+1)\Gamma(s-c+1)}$
Racah	$u_n^{(\alpha, \beta)}(s, a, b)$	$x(s) = s(s + 1)$	$a$	$b$	$\frac{\Gamma(a+s+1)\Gamma(s-a+\beta+1)\Gamma(b+\alpha-s)\Gamma(b+\alpha+s+1)}{\Gamma(a-\beta+s+1)\Gamma(s-a+1)\Gamma(b-s)\Gamma(b+s+1)}$
q-Krawtchouk	$k_n^{(p)}(s, N, q)$	$x(s) = q^{2s}$	$0$	$N$	$\frac{q^{s(s+1)}\mu^s[N!]}{\Gamma_q(s+1)\Gamma_q(N+1-s)}, \mu = p/(1-p)$
q-Meixner	$m_n^{(r, \mu)}(s, q)$	$x(s) = q^{2s}$	$0$	$\infty$	$\frac{\mu^s \Gamma_q(r+s)}{\Gamma_q(r)\Gamma_q(s+1)}$
q-Hahn	$w_n^{(c)}(s, a, b)_q$	$x(s) = [s]_q[s + 1]_q$	$a$	$b$	$\frac{q^{-s(s+1)}[s+a]_q![s+c]_q!}{[s-a]_q![s+c]_q![s+b]_q![b-s-1]_q!}$

**Tabla 2.4:** Familias de polinomios ortogonales discretos en una rejilla no uniforme.

donde

$$x_n(s) = x\left(s + \frac{n}{2}\right) \tag{2.53}$$

y

$$\rho_n(s) = \rho(n + s) \prod_{k=1}^n \sigma(s + k). \tag{2.54}$$

Se sabe que para algunos tipos especiales de rejillas, las soluciones de la ecuación 2.50 son polinomios ortogonales de variable discreta y satisfacen la siguiente propiedad de ortogonalidad:

$$\sum_{s=a}^{b-1} P_n(s)P_m(s)\rho(s) \left[ \Delta x \left( s - \frac{1}{2} \right) \right] = \delta_{nm}d_n^2, \tag{2.55}$$

donde  $d_n^2$  es el cuadrado de la norma de los polinomios ortogonales y  $\rho(s)$  es una función de peso no negativa, por ejemplo:

$$\rho(s) \left[ \Delta x \left( s - \frac{1}{2} \right) \right] > 0, \quad a \leq s \leq b - 1 \tag{2.56}$$

dentro de un conjunto real entre  $a$  y  $b$  tal que

$$\frac{\Delta}{\Delta x \left( s - \frac{1}{2} \right)} [\sigma(x)\rho(x)] = \tau(x)\rho(x). \tag{2.57}$$

Algunos de los polinomios ortogonales discretos más importantes son mencionados en la Tabla 2.4. De entre ellos, los polinomios dual Hahn y los de Racah son relativamente simples por su función de peso y la rejilla en la que están definidos. Además, estos polinomios tienen un dominio finito por definición que los hace aptos para trabajar con imágenes de  $N \times N$  píxeles. En la Tabla 2.5 se hace una comparación entre los polinomios de Racah y los dual Hahn. Los polinomios de Racah sirven como funciones base para definir nuevos momentos porque todas las demás familias en el esquema

Momento	Polinomios	Simetría	Parámetros	Reconstrucción	Compresión
Dual Hahn	Simple: ${}_2F_3$	sobre $n$ sobre $n + a$	$(a, b, c)$	Muy buena	Buena
Racah	Compleja: ${}_3F_4$	y $s$ , sólo si $a = \alpha = \beta$	$(a, b, \alpha, \beta)$	Buena	Muy buena

Tabla 2.5: Comparación entre los momentos de Racah y los dual Hahn

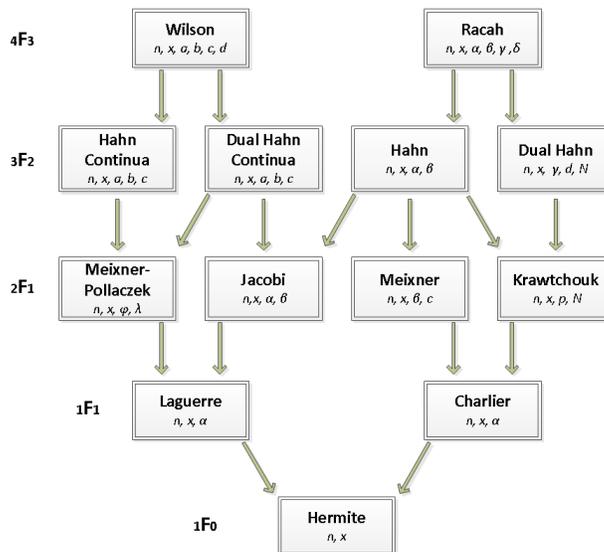


Figura 2.4: Esquema Askey de Polinomios ortogonales hipergeométricos y las relaciones entre ellos.

Askey (Figura 2.4) se pueden obtener a partir de ellos al hacer diferentes consideraciones. La Figura 2.4 muestra las relaciones entre las familias de polinomios. Muchas otras relaciones entre polinomios ortogonales hipergeométricos y sus  $q$ -análogas se pueden encontrar en [74].

### 2.3.1 POLINOMIOS Y MOMENTOS DUAL HAHN

Los polinomios clásicos dual Hahn  $w_n(s, a, b, c)$  están definidos en la rejilla no uniforme  $x(s) = s(s + 1)$  y son una solución a la ecuación 2.50 publicada en [53]:

$$\sigma(s) = (s - a)(s + b)(s - c),$$

$$\tau(s) = ab - ac + bc - a + b - c - 1 - x(s)$$

y

$$\lambda = n.$$

La función de peso  $\rho(s)$  se define como:

$$\rho(s) = \frac{\Gamma(a+s+1)\Gamma(c+s+1)}{\Gamma(s-a+1)\Gamma(b-s)\Gamma(b+s+1)\Gamma(s-c+1)}, \quad (2.58)$$

donde los parámetros  $a$ ,  $b$  y  $c$  tienen las siguientes restricciones:

$$-\frac{1}{2} < a < b, \quad |c| < 1 + a \quad y \quad b = a + N. \quad (2.59)$$

Cabe mencionar que si se define una rejilla uniforme, por ejemplo:  $x(s) = s$ , para resolver la ecuación 2.50 con los parámetros  $a$ ,  $b$  y  $c$  definidos como  $a = (\alpha + \beta)/2$ ,  $b = a + N$  y  $c = (\beta - \alpha)/2$ , los polinomios dual Hahn se convierten en los polinomios Hahn [53]. A partir de este punto, si  $\alpha = 0$  y  $\beta = 0$ , los polinomios Hahn son reducidos a los polinomios de Tchebichef. Además, si definimos  $\beta = pt$  y  $\alpha = (1 - p)t$  en los polinomios de Hahn y si  $t \rightarrow \infty$ , se obtienen los polinomios de Krawtchouk [74].

Los polinomios dual Hahn de orden  $n$  se definen de la siguiente manera [71]:

$$w_n(s, a, b, c) = \frac{(a-b+1)_n(a+c+1)_n}{n!} {}_3F_2(-n, a-s, a+s+1; a-b+1, a+c+1; 1), \quad (2.60)$$

en la cual  $n = \{0, 1, \dots, N-1\}$ ,  $s = \{a, a+1, \dots, b-1\}$ ,  $(u)_k$  es el símbolo de Pochhammer (definido en la ecuación 2.34) y  ${}_3F_2$  es la función general hipergeométrica dada por:

$${}_3F_2(a_1, a_2, a_3; b_1, b_2; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{z^k}{k!}. \quad (2.61)$$

Los polinomios dual Hahn satisfacen la propiedad de ortogonalidad:

$$\sum_{s=a}^{b-1} w_n(s, a, b, c) w_m(s, a, b, c) \rho(s) \left[ \Delta x \left( s - \frac{1}{2} \right) \right] = \delta_{mn} d_n^2, \quad (2.62)$$

donde  $n, m = \{0, 1, \dots, N-1\}$ ,  $\delta_{mn}$  es una delta de Kronecker, con la función de peso  $\rho(s)$  definida en la ecuación 2.58 y la norma cuadrada  $d_n^2$  dada por:

$$d_n^2 = \frac{\Gamma(a+c+n+1)}{n!(b-a-n-1)!\Gamma(b-c-n)}, \quad n = 0, 1, \dots, N-1. \quad (2.63)$$

El conjunto de polinomios dual Hahn no sirve para definir momentos útiles para el procesamiento de imágenes porque el rango de valores se incrementa rápidamente a medida que el orden aumenta. Para evitar este problema, se propusieron los polinomios dual Hahn normalizados [45].

Los polinomios dual Hahn son escalados utilizando una norma cuadrada y la función de peso. Por lo tanto, el conjunto de polinomios dual Hahn normalizados,  $\hat{w}_n$ , es definido como:

$$\hat{w}_n(s, a, b, c) = w_n(s, a, b, c) \sqrt{\frac{\rho(s)}{d_n^2} \left[ \Delta x \left( s - \frac{1}{2} \right) \right]}, \quad n = 0, 1, \dots, N-1. \quad (2.64)$$

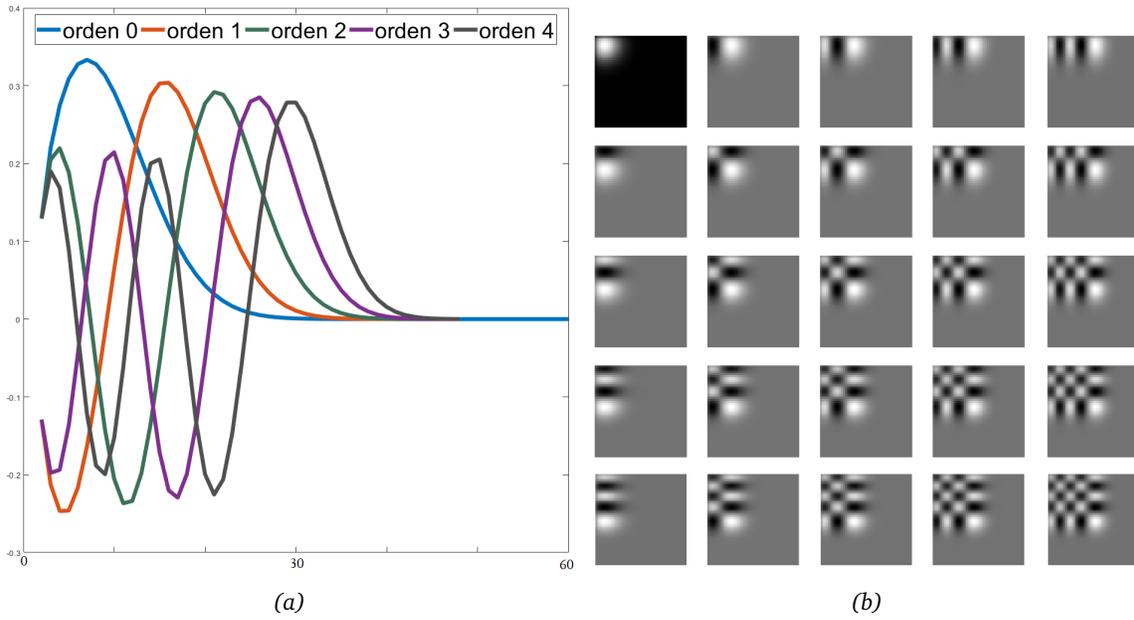


Figura 2.5: (a) Polinomios dual Hahn de orden 0 a 4 con  $N = 60$  y (b) Polinomios dual Hahn bidimensionales de orden 0 al 8.

Para este caso, la condición de ortogonalidad es:

$$\sum_{s=a}^{b-1} \hat{w}_n(s, a, b, c) \hat{w}_m(s, a, b, c) = \delta_{mn}. \quad (2.65)$$

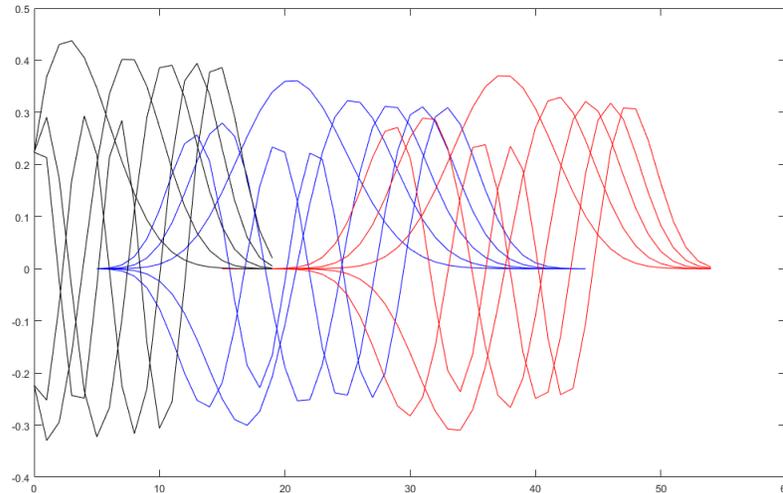
Los valores de los polinomios dual Hahn normalizados están contenidos en el rango  $[-1, 1]$ . En la Figura 2.5(a) se muestra a los polinomios discretos dual Hahn unidimensionales desde el orden 0 al orden 4. En la Figura 2.5(b), se tiene un diagrama con los polinomios discretos bidimensionales dual Hahn del orden 0 al 8.

Finalmente, los **Momentos dual Hahn**,  $W_{mn}$ , de orden  $m + n$  de la imagen  $I(s, t)$  son calculados de la siguiente forma:

$$W_{mn} = \sum_{s=a}^{b-1} \sum_{t=a}^{b-1} \hat{w}_n(s, a, b, c) \hat{w}_m(t, a, b, c) I(s, t), \quad (2.66)$$

donde  $n, m = \{0, 1, \dots, N - 1\}$  y  $N$  es el tamaño de la imagen cuadrada.

Estos polinomios discretos también tienen la capacidad de adaptarse a una zona de interés, como los polinomios de Krawtchouk. Además, cuentan con la habilidad de expansión y compresión dependiendo de los valores de  $a, b$  y  $c$ . En la Figura 2.6 hay tres conjuntos de polinomios unidimensionales dual Hahn, el primer conjunto de color negro se obtiene con  $a = 0, b = 20$  y  $c = 0$ , el segundo conjunto de color azul cuenta con  $a = 5, b = 45$  y  $c = 5$  y el tercero en color rojo se calcula con  $a = 20, b = 60$  y  $c = 20$ . Comparando al conjunto azul con el rojo, se observa la capacidad de corrimiento de los polinomios. Ahora, si se compara al conjunto negro con cualquiera de los otros dos, se muestra la capacidad de compresión que pudiera servir para enfocarlos en zonas de interés más angostas o gruesas, dependiendo del caso.



**Figura 2.6:** Polinomios dual Hahn con variación de parámetros. Los polinomios de color negro se obtienen con los parámetros  $a = 0$ ,  $b = 20$  y  $c = 0$ . Por otro lado, los polinomios de color azul se calculan con los valores  $a = 5$ ,  $b = 45$  y  $c = 5$ . Finalmente, los polinomios de color rojo son el resultado del cálculo con  $a = 20$ ,  $b = 60$  y  $c = 20$ . Con la variación de estos parámetros se muestran las propiedades de adaptación a una zona de interés y la habilidad de expansión/compresión propias de los polinomios dual Hahn.

## 2.4 Polinomios y Momentos de Shmaliy

Los momentos de Tchebichef están basados en los polinomios discretos del mismo nombre y para su expansión dependen sólo de un parámetro, la longitud de las  $N$  muestras, mientras que, los polinomios Dual Hahn y Krawtchouk dependen de tres y dos parámetros, respectivamente. Shmaliy [75] desarrolló los polinomios discretos de respuesta finita al impulso no sesgado (UFIRs) de un parámetro para filtros FIR [76] basados en FIRs predictivos de un paso desarrollados por Heinonen y Neuvo [77]. Después, Morales-Mendoza *et al.* muestran en [78] que los UFIRs conforman una nueva familia de polinomios ortogonales discretos que también dependen de la longitud  $N$  de los datos, como único parámetro para su expansión. Los autores concluyen que esta familia puede ser utilizada en procesamiento en tiempo real por su cálculo rápido.

Los polinomios Shmaliy de grado  $p$  se definen como:

$$h_p(n, N) = \sum_{i=0}^p a_{ip}(N) n^i, \quad (2.67)$$

donde  $p = \{0, 1, \dots, N - 1\}$  y los coeficientes  $a_{ip}$  son:

$$a_{ip}(N) = (-1)^i \frac{M_{(i+1),1}^{(p)}(N)}{|H_p(N)|}, \quad (2.68)$$

donde  $|H_p(N)|$  es el determinante y  $M_{(i+1),1}^{(p)}(N)$  es el menor de la matriz de Hankel  $H_p(N)$

$$H_p(N) = \begin{bmatrix} c_0 & c_1 & \cdots & c_p \\ c_1 & c_2 & \cdots & c_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_p & c_{p+1} & \cdots & c_{2p} \end{bmatrix}.$$

Los elementos de la matriz de Hankel son series de potencias:

$$c_k(N) = \sum_{i=0}^{N-1} i^k = \frac{1}{k+1}(B_{k+1}(N) - B_{k+1}), \quad (2.69)$$

donde  $B_k(x)$  es el polinomio de Bernoulli y  $B_k = B_k(0)$  es el número de Bernoulli. Sustituyendo la ecuación 2.68 en la ecuación 2.67, obtenemos un polinomio de variable discreta  $n$  es:

$$h_p(n, N) = \sum_{i=0}^p (-1)^i \frac{M_{(i+1),1}^{(p)}(N)}{|H_p(N)|} n^i. \quad (2.70)$$

La definición hipergeométrica de los polinomios Shmaliy [47] corresponden a la siguiente ecuación:

$$h_p(n, N) = \frac{(-1)^p (p+1)(n-p)_p (N-n)_p}{p!(N)_{p+1}} {}_3F_2(-p, n+1, 1-N+n; n-p, 1-N-p+n; 1). \quad (2.71)$$

Los polinomios discretos  $h_p(n, N)$  son ortogonales en  $n \in [0, N-1]$  y satisfacen la siguiente relación:

$$\sum_{n=0}^{N-1} w(n, N) h_p(n, N) h_q(n, N) = \rho(p, N) \delta_{pq}, \quad (2.72)$$

donde  $p, q = 0, 1, \dots, N-1$  y  $\delta_{pq}$  es la delta de Kronecker. La norma  $\rho(p, N)$  de  $h_p(n, N)$  es:

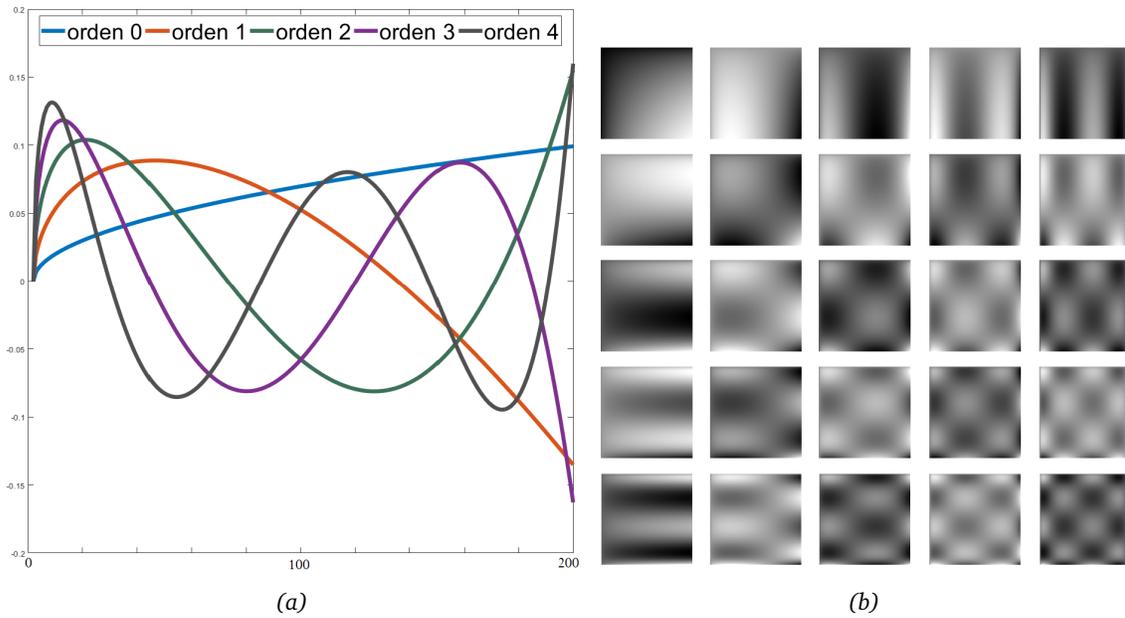
$$\rho(p, N) = \frac{p+1}{N(N+1)} \prod_{i=0}^p \frac{N-i-1}{N+i} = \frac{(p+1)(N-p-1)_p}{N(N)_{p+1}}, \quad (2.73)$$

donde  $(a)_k$  es el símbolo de Pochhammer. La función de peso  $w(n, N)$  es una función lineal discreta con respecto a  $n$ :

$$w(n, N) = \frac{2n}{N(N-1)}. \quad (2.74)$$

Para obtener la estabilidad numérica en el cálculo de los polinomios, es necesario considerar una ponderación. Por lo tanto, los polinomios ponderados discretos de Shmaliy,  $\hat{h}_p$ , son:

$$\hat{h}_p(n, N) = h_p(n, N) \sqrt{\frac{w(n, N)}{\rho(p, N)}}. \quad (2.75)$$



**Figura 2.7:** (a) Polinomios discretos de Shmaliy de orden 0 a 4 con  $N = 200$ . (b) Polinomios discretos de Shmaliy bidimensionales de orden 0 a 8.

Para los polinomios ponderados, la condición de ortogonalidad ahora se escribe como:

$$\sum_{n=0}^{N-1} \hat{h}_p(n, N) \hat{h}_q(n, N) = \delta_{pq}. \quad (2.76)$$

En la Figura 2.7(a) muestra a los polinomios unidimensionales de Shmaliy desde el orden 0 al orden 4, mientras que, la Figura 2.7(b) muestra un diagrama con los polinomios bidimensionales de Shmaliy del orden 0 al 8.

Finalmente, los **Momentos de Shmaliy**,  $S_{mn}$ , de orden  $m + n$  de la imagen  $I(x, y)$  se definen como:

$$S_{mn} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \hat{h}_n(n, N) \hat{h}_m(m, N) I(x, y), \quad (2.77)$$

donde  $N$  y  $M$  son el tamaño de la imagen en el eje  $X$  y en el eje  $Y$ , respectivamente.

## 2.5 Implementación de los polinomios ortogonales discretos

De acuerdo con Papakostas *et al.* [79], las principales direcciones de investigación en el campo de los momentos ortogonales aplicados en imágenes son las siguientes:

1. El desarrollo de nuevos algoritmos que aceleran el tiempo de cálculo general de los momentos.
2. La mejora de la precisión del momento al reducir errores de cuantización y aproximación.
3. La incorporación de características invariantes a la traslación, rotación y escala en el cálculo de

Estrategia	Ventajas	Desventajas
Cálculo directo	<ul style="list-style-type: none"> <li>● Cálculo rápido del momento de un orden en particular.</li> <li>● Aplicable a cualquier familia de momentos.</li> </ul>	<ul style="list-style-type: none"> <li>● Cálculo lento de un conjunto completo de momentos.</li> <li>● Inestabilidad numérica para momentos de orden alto.</li> <li>● Alta complejidad.</li> </ul>
Recursión	<ul style="list-style-type: none"> <li>● Cálculo rápido de un conjunto completo de momentos.</li> <li>● Baja complejidad.</li> <li>● Estabilidad numérica para momentos de orden alto.</li> </ul>	<ul style="list-style-type: none"> <li>● Dependencia de las propiedades de la familia de polinomios.</li> <li>● Necesidad de una mayor cantidad e memoria.</li> <li>● Cálculo lento del momento de un orden en particular.</li> <li>● Sensibilidad a la acumulación de errores numéricos.</li> </ul>

**Tabla 2.6:** Resumen de las características de las estrategias de cálculo directo y de recursión de los polinomios ortogonales discretos.

los momentos.

En lo que respecta a la primera dirección, Papakostas *et al.* proponen cuatro categorías principales basadas en la estrategia que siguen para el cálculo de los momentos, caracterizadas por diferentes tipos de procesamiento adicionales y específicos para cada categoría. Las estrategias para el cálculo de los momentos son: directa, de recursión, de partición y de bloque de corte.

En este trabajo de investigación, la implementación que se utilizó para el cálculo de los polinomios ortogonales discretos está basada en la estrategia de recursión. Por esta razón, en esta sección se hace mención de algunas características de esta estrategia y también se le compara con respecto a la estrategia de cálculo directo. Además, se muestran las fórmulas de recursión implementadas para las diferentes familias de polinomios ortogonales discretos utilizadas en este trabajo de investigación.

### 2.5.1 COMPARACIÓN ENTRE ESTRATEGIAS DE CÁLCULO DE POLINOMIOS Y MOMENTOS ORTOGONALES

La estrategia de cálculo directo fue la primera en ser utilizada, ya que se basa en las definiciones hipergeométricas de cada familia de momentos ortogonales. El cálculo directo de los momentos es un procedimiento muy lento ya que, para cada pixel, un conjunto de polinomios debe ser evaluado.

Por otra parte, la estrategia recursiva se caracteriza por el aprovechamiento de las propiedades

Polinomios	$\Lambda$	$\zeta$	$\psi$
Tchebichef	$\frac{n+1}{2(2n+1)}$	$\frac{N-1}{2}$	$\frac{n(N^2-n^2)}{2(2n+1)}$
Krawtchouk	$n + 1$	$n + p(N - 2n)$	$pq(N - n + 1)$
dual Hahn	$n + 1$	$ab - ac + bc + (b - a - c - 1)(2n + 1) - 2n^2$	$(a + c + n)(b - a - n)(b - c - n)$

**Tabla 2.7:** Valores de  $\phi$ ,  $\zeta$  y  $\psi$  para los polinomios de Tchebichef, Krawtchouk y dual Hahn.

recursivas de las definiciones hipergeométricas de los polinomios [53]. Esta propiedad acelera el cálculo general significativamente, especialmente en los casos en que se requiere un conjunto de momentos hasta un orden específico. Con relación a la velocidad de cálculo, este tipo de algoritmos parece ser una elección perfecta, sin embargo, su naturaleza recursiva permite la aparición de errores en la cuantización [42, 43] o en la precisión [80]. Dichos errores se acumulan recursión tras recursión y finalmente pueden arrojar resultados poco confiables. En comparación con el método de cálculo directo, la estrategia recursiva necesita mayor cantidad de memoria porque este método permite el cálculo previo de algunas variables que se utilizan en cada recursión, sin embargo, el cálculo es menos demandante para el procesador porque elimina el sobre flujo numérico causado por las funciones hipergeométricas y las funciones gamma. En la Tabla 2.6 se muestra un resumen de las ventajas y desventajas de las estrategias de cálculo directo y de recursión en términos de su complejidad, necesidades de preprocesamiento, estabilidad numérica, etcétera.

A pesar de este comportamiento inestable bajo ciertas condiciones, esta estrategia se puede considerar como la más eficiente y se ha aplicado con éxito en casi todas las familias de momentos ortogonales discretos [33, 40, 45]. La relación de recursión utilizada para la implementación de los polinomios ortogonales discretos obedece a la siguiente relación de tres términos:

$$xy_n(x) = \Lambda_n y_{n+1}(x) + \zeta_n y_n(x) + \psi y_{n-1}(x). \tag{2.78}$$

Los valores de  $\phi$ ,  $\zeta$  y  $\psi$  para los polinomios de Tchebichef, Krawtchouk y dual Hahn se muestran en la Tabla 2.7. Los polinomios de Shmaliiy obedecen a una relación de recursión diferente que es explicada en la Sección 2.5.6.

### 2.5.2 POLINOMIOS DE TCHEBICHEF - RECURSIÓN SOBRE N

Para un número positivo y entero  $N$ , un valor  $x$  y el orden del polinomio  $n$ , ambos con el rango  $[0, N - 1]$ , se pueden definir los polinomios normalizados de Tchebichef a partir de las siguientes

funciones recursivas:

$$\tilde{t}_0(x) = \frac{1}{\sqrt{N}} \quad (2.79)$$

y

$$\tilde{t}_1(x) = (2x + 1 - N) \sqrt{\frac{3}{N(N^2 - 1)}}. \quad (2.80)$$

La ecuación de recursividad esta dada por:

$$\tilde{t}_n(x) = \alpha_1 x t_{n-1}(x) + \alpha_2 t_{n-1}(x) + \alpha_3 t_{n-2}(x), \quad (2.81)$$

donde:

$$\alpha_1 = \frac{2}{n} \sqrt{\frac{4n^2 - 1}{N^2 - n^2}}, \quad (2.82)$$

$$\alpha_2 = \frac{1 - N}{n} \sqrt{\frac{4n^2 - 1}{N^2 - n^2}} \quad (2.83)$$

y

$$\alpha_3 = \frac{n - 1}{n} \sqrt{\frac{2n + 1}{2n - 1}} \sqrt{\frac{N^2 - (n - 1)^2}{N^2 - n^2}}. \quad (2.84)$$

### 2.5.3 POLINOMIOS DE TCHEBICHEF - RECURSIÓN SOBRE X

Se sabe que:

$$\tilde{t}_n(0) = \frac{(1 - N)(2 - N) \dots (n - N)}{\beta(n, N)}, \quad (2.85)$$

donde  $n = 0, 1, \dots, N - 1$ , y

$$\beta(n, N) = \sqrt{\frac{N(N^2 - 1)(N^2 - 2^2) \dots (N^2 - n^2)}{2n + 1}}.$$

Obtenemos:

$$\tilde{t}_n(0) = -\sqrt{\frac{N - n}{N + n}} \sqrt{\frac{2n + 1}{2n - 1}} \tilde{t}_{n-1}(0), \quad n = 1, \dots, N - 1. \quad (2.86)$$

Además, tenemos:

$$\tilde{t}_n(1) = \left\{ 1 + \frac{n(1 + n)}{1 - N} \right\} \tilde{t}_n(0). \quad (2.87)$$

Entonces, las ecuaciones 2.86 y 2.87 sirven para obtener los valores iniciales de la siguiente ecuación de recursión en  $x$ :

$$\tilde{t}_n(x) = \gamma_1 \tilde{t}_n(x - 1) + \gamma_2 \tilde{t}_n(x - 2), \quad (2.88)$$

donde  $n = 1, 2, \dots, N - 1$ ,  $x = 2, 3, \dots, N$  y

$$\gamma_1 = \frac{-n(n + 1) - (2x - 1)(x - N - 1) - x}{x(N - x)} \quad (2.89)$$

y

$$\gamma_2 = \frac{(x-1)(x-N-1)}{x(N-x)}. \quad (2.90)$$

Cabe resaltar que la recursión puede terminar en  $x = N/2$ , siempre y cuando  $N$  sea par, ya que podemos hacer uso de la condición de simetría

$$\tilde{t}_n(N-1-x) = (-1)^n \tilde{t}_n(x) \quad (2.91)$$

para calcular los valores polinomiales cuando el rango de  $x$  sea  $[N/2, N-1]$ . La recursividad en  $x$  reduce el tiempo de cómputo y la cantidad de errores acumulados, es decir, es más estable que el cálculo sobre el orden  $n$ .

#### 2.5.4 POLINOMIOS DE KRAWTCHOUK - RECURSIÓN SOBRE N

La recursividad en términos del grado  $n$  de los polinomios normalizados de Krawtchouk se define como:

$$p(N-n)\hat{K}_{n+1}(x; p, N) = A(Np - 2np + n - x)\hat{K}_n(x; p, N) - Bn(1-p)\hat{K}_{n-1}(x; p, N) \quad (2.92)$$

donde:

$$A = \sqrt{\frac{p(N-n)}{(1-p)(n+1)}} \quad (2.93)$$

y

$$B = \sqrt{\frac{p^2(N-n)(N-n+1)}{(1-p)^2(n+1)n}}, \quad (2.94)$$

con

$$\hat{K}_0(x; p, N) = \sqrt{\frac{w(x; p, N)}{\rho(0; p, N)}}, \quad (2.95)$$

y

$$\hat{K}_1(x; p, N) = \left(1 - \frac{x}{Np}\right) \sqrt{\frac{w(x; p, N)}{\rho(1; p, N)}}. \quad (2.96)$$

La función de peso en la ecuación 2.42 puede ser calculada recursivamente de la siguiente manera:

$$w(x+1; p, N) = \left(\frac{N-x}{x+1}\right) \frac{p}{1-p} w(x; p, N), \quad (2.97)$$

con  $w(0; p, N) = (1-p)^N = e^{N \ln(1-p)}$ .

El tiempo de procesamiento de los momentos de Krawtchouk para el caso especial en el que  $p_1 = p_2 = 0.5$  puede ser reducido utilizando la propiedad de simetría. La relación de simetría en los

polinomios normalizados de Krawtchouk se define como:

$$\hat{K}_n(x; p, N - 1) = (-1)^n \hat{K}_n(N - 1 - x; p, N - 1). \quad (2.98)$$

### 2.5.5 POLINOMIOS DUAL HAHN - RECURSIÓN SOBRE N

La fórmula de recurrencia para los polinomios dual Hahn sobre el orden  $n$  es:

$$\hat{W}_0^{(c)}(s, a, b) = \sqrt{\frac{\rho(s)}{d_0^2}}(2s + 1), \quad (2.99)$$

$$\hat{W}_1^{(c)}(s, a, b) = -\frac{\rho(s+1)(s+1-a)(s+1+b)(s+1-c) - \rho(s)(s-a)(s+b)(s-c)}{\rho(s)(2s+1)} \sqrt{\frac{\rho(s)}{d_1^2}}(2s+1) \quad (2.100)$$

y

$$\hat{W}_{n+1}^{(c)}(s, a, b) = A \frac{d_n}{d_{n+1}} \hat{W}_n^{(c)}(s, a, b) - B \frac{d_{n-1}}{d_{n+1}} \hat{W}_{n-1}^{(c)}(s, a, b), \quad (2.101)$$

donde

$$A = \frac{1}{n+1}(s(s+1) - ab + ac - bc - (b-a-c-1)(2n+1) + 2n^2) \quad (2.102)$$

y

$$B = \frac{1}{n+1}(a+c+n)(b-a-n)(b-c-n). \quad (2.103)$$

### 2.5.6 POLINOMIOS DE SHMALIY - RECURSIÓN SOBRE N

La relación de recurrencia de tres términos para cualquier familia de polinomios ortogonales mónicos reales  $P_p(x)$  es:

$$-nP_p(n) = b_p P_{p+1}(n) + \gamma_p P_p(n) + c_p P_{p-1}(n), \quad n \geq 1, \quad (2.104)$$

donde  $b_p, c_p \neq 0$  y  $c_p/b_{p-1} > 0$ . Entonces, para los polinomios de Shmaliy  $h_p(n, N)$ , la relación mostrada en la ecuación 2.104 es:

$$h_p(n, N) = (\alpha_p + n\beta_p)h_{p-1}(n, N) - \zeta_n h_{p-2}(n, N). \quad (2.105)$$

Si consideramos que :

$$\frac{d_{p-1}^2(N)}{d_{p-2}^2(N)} = \frac{p(N-p)}{(p-1)(N+p-1)},$$

entonces, los coeficientes  $\beta_p$ ,  $\alpha_p$  y  $\zeta_p$  se plantean de la siguiente manera:

$$\beta_p = \frac{a_{pp}(N)}{a_{(p-1)(p-1)}(N)} = -\frac{2(2p+1)}{p(N+p)},$$

$$\alpha_p = \beta_p \left[ \frac{a_{p-1}(N)}{a_{pp}(N)} - \frac{a_{(p-1)(p-2)}(N)}{a_{(p-1)(p-1)}(N)} \right] = \frac{2p(2N-1)}{(2p-1)(N+p)}$$

y

$$\zeta_p = \beta_p \frac{a_{(p-2)(p-2)}(N)}{a_{(p-1)(p-1)}(N)} \frac{d_{p-1}^2(N)}{d_{p-2}^2(N)} = \frac{(2p+1)(N-p)}{(2p-1)(N+p)}.$$

Finalmente, al usar las funciones anteriores, la ecuación 2.105 queda establecida como:

$$h_p(n) = 2 \frac{p^2(2N-1) - n(4p^2-1)}{p(2p-1)(N+p)} h_{p-1}(n) - \frac{(2p+1)(N-p)}{(2p-1)(N+p)} h_{p-2}(n), \quad (2.106)$$

donde  $p \geq 1$ ,  $h_{-1}(n) = 0$  y  $h_0(n) = 1/N$ .



# 3

## Clasificación de texturas con polinomios y momentos ortogonales discretos

El enfoque principal de este trabajo es la descripción y clasificación de texturas, del cual, existen pocos estudios que emplean momentos ortogonales aplicados en bases de datos biomédicos. Por ejemplo, Li y Meng usan los coeficientes de Tchebichef de bajo orden para caracterizar textura y color en imágenes de endoscopía [81]. También, el estudio de Wu [82] emplea una combinación de coeficientes de diferentes familias de momentos para caracterizar regiones cuadradas en imágenes de tomografía de hígado. Por otro lado, el filtrado con los polinomios de Tchebichef en ventanas de  $5 \times 5$  para generar una nueva imagen codificada útil para clasificar diferentes tejidos fue reportado por Mukundan [83]. Y en la aplicación de Marcos y Cristóbal [48] se reporta el cálculo de una firma de textura a través de los momentos discretos de Tchebichef aplicada en bases de datos públicas.

En todos los casos, lo más importante es el uso de los coeficientes calculados a partir de los momentos, es decir, la generación de los vectores de características que servirán para calcular la regla de clasificación. En los primeros tres trabajos [81–83], las propuestas sólo utilizan un pequeño número de coeficientes de bajo orden para clasificar las bases de prueba. En cambio, en el estudio de Víctor y Cristóbal [48], el vector de características de tamaño  $2N - 1$ , donde  $N$  es el tamaño de la imagen cuadrada, es considerablemente mayor que en las propuestas anteriores al incluir los coeficientes de frecuencias altas, teniendo así un descriptor, en principio, más completo.

Este capítulo profundiza en este último estudio, por el hecho de construir un vector de características a partir de los momentos bidimensionales. Además, se explica qué son los *Textons* [84], metodología que utiliza bancos de filtros para la descripción de texturas y que, con una adecuación, los polinomios ortogonales discretos son incluidos en ese esquema. Posteriormente, se proponen diferentes métodos

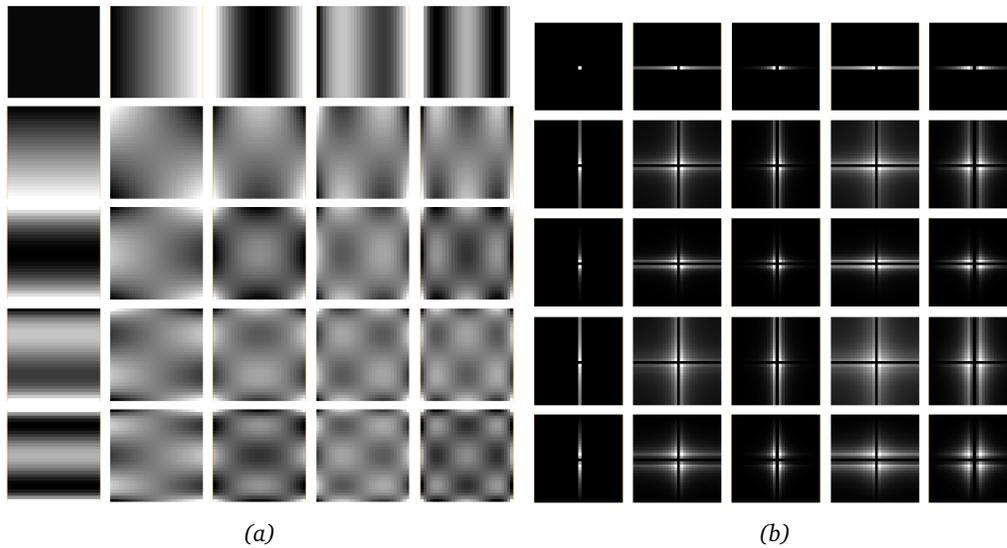


Figura 3.1: Conjunto de polinomios de Tchebichef en el dominio (a) espacial y (b) frecuencial.

de clasificación basados en estos dos trabajos. Una descripción profunda de todos los clasificadores utilizados es este trabajo de investigación se encuentra en el Apéndice B.

### 3.1 Firma de textura basada en momentos discretos de Tchebichef

Los momentos ortogonales, descritos en la Sección 2, son cantidades escalares que caracterizan una función de interés. De acuerdo con la ecuación 2.39, el valor del momento bidimensional  $T_{mn}$  puede ser interpretado como la correlación entre la imagen  $I(x, y)$  y la base polinomial bidimensional  $r_{mn}(x, y) = \hat{t}_n(x)\hat{t}_m(y)$  que se obtiene a partir de los polinomios discretos normalizados de Tchebichef.

Esta base polinomial tiene un comportamiento oscilatorio, parecido al de una función seno. La Figura 3.1 muestra el conjunto de bases de Tchebichef en el dominio espacial y frecuencial. Como se puede observar, conforme al orden del conjunto polinomial aumenta, la energía del conjunto tiende concentrarse en las frecuencias altas [85].

La base polinomial  $r_{mn}(x, y)$  actúa como un filtro para el cálculo de  $T_{mn}$ . La magnitud  $T_{mn}$  será mayor para imágenes cuyo comportamiento oscilatorio sea similar al de  $r_{mn}(x, y)$  en ambas direcciones. Esta característica es interesante para el análisis de textura de imágenes. Por lo tanto, se obtiene una descripción de las propiedades de textura evaluando la dependencia de la magnitud del momento de orden  $s$ , que se relaciona con el contenido frecuencial de la base polinomial. Finalmente, se construye la firma de textura  $M(s)$  de la siguiente forma:

$$M(s) = \sum_{s=m+n} |T_{mn}| \quad (3.1)$$

con  $s = 0, 1, \dots, 2N - 2$ .

Marcos y Cristóbal [48] realizaron la validación de los momentos ortogonales discretos de Tche-

bichet como descriptores de textura al compararlos con otros métodos estándar como la Matriz de co-ocurrencia de Haralick [9], Filtrado Gaussiano [86, 87] y Patrones Binarios Locales [88]. Los experimentos demostraron que los momentos de Tchebichef capturan información esencial de la textura teniendo resultados comparables, e incluso mejores, que los métodos convencionales.

Estas firmas de texturas se pueden obtener para cada una de las familias de momentos de la Sección 2 y su aplicación se muestra en la Sección 5.

### 3.2 Cálculo de *Textons*

Los *Textons* son estructuras fundamentales de textura dentro de cualquier imagen y, por lo tanto, constituyen los elementos básicos de la percepción visual. Haciendo una analogía con la Física, las bases de las imágenes son como los protones, neutrones y electrones, entonces, los *Textons* serían los átomos [84, 89]. Los *Textons* pueden representar las diversas relaciones entre píxeles que existen en una región, resultando esenciales para el análisis de textura.

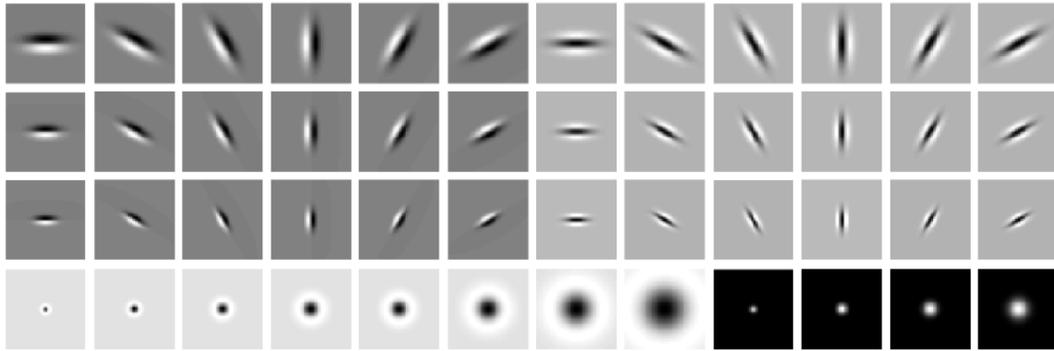
Los *Textons* también son definidos como las respuestas representativas que se obtienen después de convolucionar una imagen con un conjunto o banco de filtros. El cálculo de *Textons* se realiza de la siguiente manera:

1. **Obtención de los datos de entrenamiento:** Los datos de entrenamiento deben tener toda la información que necesita ser clasificada. Esa información es recolectada y clasificada por expertos.
2. **Elección del banco de filtros:** El banco de filtros tiene como propósito la obtención de características relevantes a través de las respuestas generadas por los mismos. Los tres bancos de filtros más utilizados en la clasificación de textura basada en *Textons* son:
  - (a) *El banco Leung-Malik (LM):* El banco LM consiste en 48 filtros con múltiples escalas y orientaciones. Está conformado por la primera y segunda derivadas de Gaussiana a 6 orientaciones y 3 escalas distintas, también incluye 8 filtros Laplacianos de Gaussianas (LoG) y 4 Gaussianas [84]. El conjunto lo vemos en la Figura 3.2.
  - (b) *El banco Schmid (S):* El banco S consiste de 13 filtros invariantes a la rotación, definidos por la siguiente ecuación:

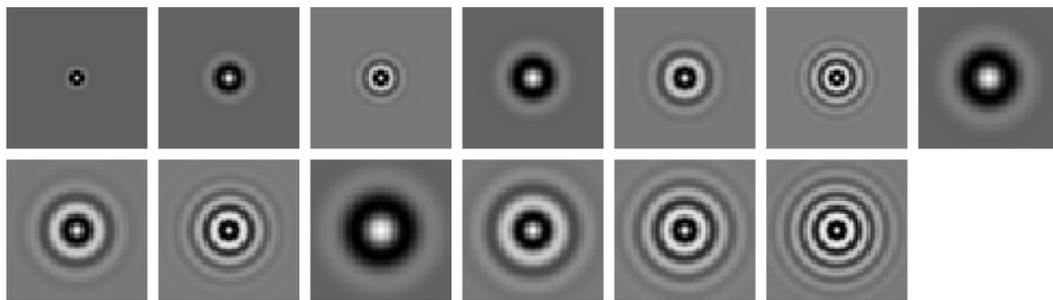
$$F(r, \sigma, \tau) = F_0(\sigma, \tau) + \cos(\pi\tau r/\sigma)e^{-\frac{r^2}{2\sigma^2}}, \quad (3.2)$$

donde  $r$  representa las coordenadas  $x$  y  $y$  del píxel en coordenadas polares,  $\sigma$  es la escala y  $\tau$  es la frecuencia, es decir, el número de ciclos de una función armónica dentro de la envolvente Gaussiana del filtro.  $F_0(\sigma, \tau)$  es adicionada para obtener la componente cero en DC con los pares  $(\sigma, \tau)$ , tomando 13 diferentes valores [84]. Además, se observa en la Figura 3.3 que los filtros tienen simetría rotacional.

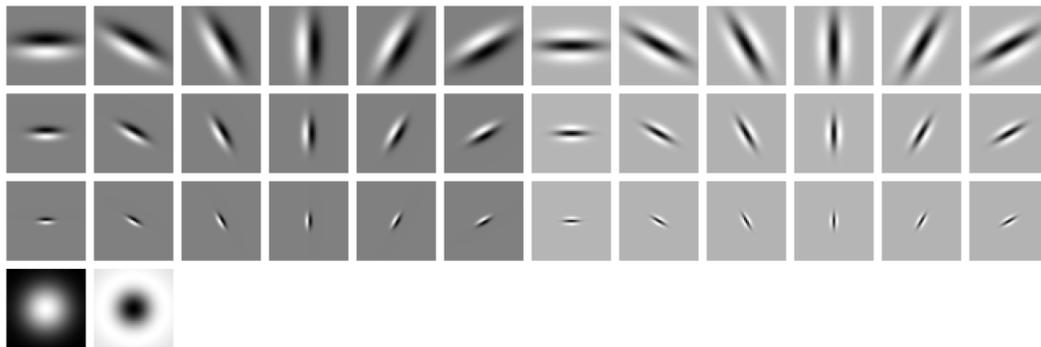
- (c) *El banco de Máxima Respuesta (MR):* El banco MR8 cuenta con 38 filtros con múltiples orientaciones, que constan de filtros Gaussianos y Laplacianos de Gaussianas. La máxima



**Figura 3.2:** El banco LM es una combinación de filtros de borde, barra y punto con múltiples escalas y orientaciones, en total son 48 filtros.



**Figura 3.3:** El banco de filtros S es invariante a la rotación y está compuesto por 13 filtros isotrópicos, semejantes a los filtros de Gabor.



**Figura 3.4:** El banco de filtros MR8 consiste en 2 filtros anisotrópicos y 2 filtros simétricos rotacionalmente. Sin embargo, sólo 8 respuestas son consideradas y corresponden a las máximas respuestas a los filtros anisotrópicos sobre todas las orientaciones.

respuesta a través de todas las orientaciones reduce el número de respuestas de 38 (6 orientaciones a 3 escalas para 2 filtros orientados, más 2 filtros isotrópicos) a 8 (3 escalas para dos filtros, más 2 isotrópicos) [84]. En la Figura 3.4 se muestra el banco de filtros MR. El banco MR4 es un subconjunto del banco MR8 donde los filtros de bordes y de barra se definen en una escala fija ( $\sigma_x = 4, \sigma_y = 12$ ).

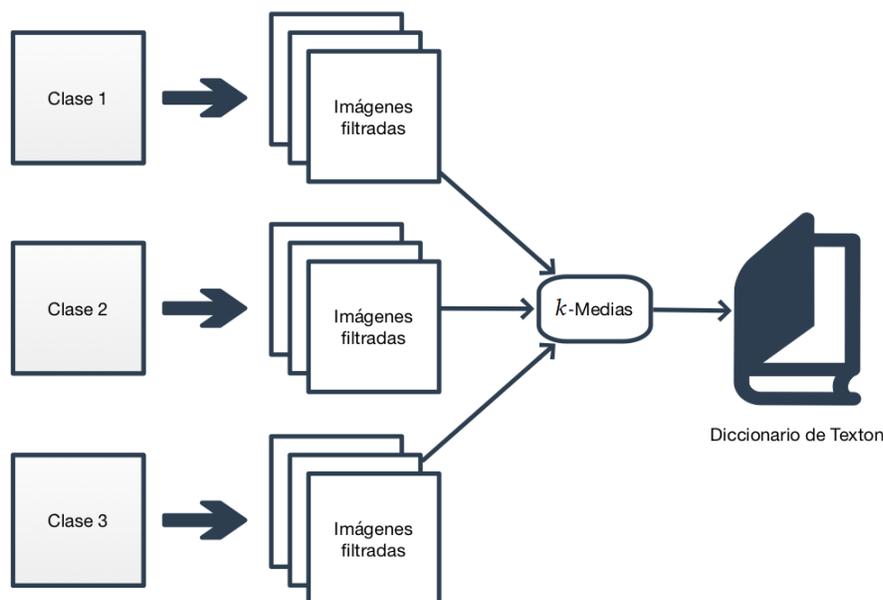


Figura 3.5: Proceso para la generación del diccionario de *Textons*.

3. **Agrupar las respuestas de los filtros:** Para cada clase,  $k$  vectores representativos son seleccionados y llamados *Textons*. Para calcularlos, cada imagen de la clase es filtrada a través de los bancos. Tenemos un vector respuesta por cada pixel y del mismo tamaño que el número de filtros  $n$ . Este vector respuesta se representa de la siguiente manera:

$$V_{pi} = [FR_1 \ FR_2 \ FR_3 \ \dots \ FR_n], \quad (3.3)$$

donde  $p_i$  representa el índice del pixel y  $FR$  es la respuesta de ese pixel. Los vectores respuesta resultantes son divididos en  $k$  agrupamientos usando el algoritmo de  $k$ -medias. Los vectores respuesta que corresponden a los  $k$  centros de agrupamiento componen los *Textons* de una clase de textura en particular.

Una vez que se calculan los *Textons*, la metodología siguiente es aplicada para la clasificación de texturas:

- **Diccionario de *Textons*:** El diccionario de *Textons* se construye combinando los *Textons* de todas las clases. Cada imagen aporta  $k$  *Textons* a través del agrupamiento de  $k$ -medias. La Figura 3.5 muestra el desarrollo de este punto.
- **Construcción de Modelos:** Usando el diccionario descrito en el punto anterior, se construye el histograma de cada imagen asignando cada pixel al *Texton* más parecido, a través de una medida de similitud entre el vector respuesta del pixel y el diccionario. La Figura 3.6 muestra la construcción de los mapas de *Texton* y el cálculo de su histograma.
- **Clasificación:** Los dos pasos anteriores se realizan para el conjunto de entrenamiento, al final tenemos que el diccionario de *Textons* se transformó en un diccionario de histogramas con sus

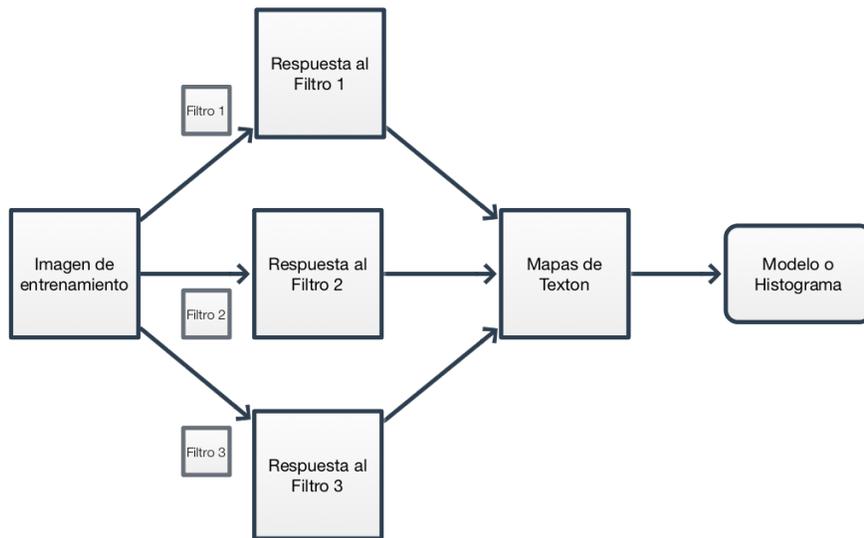


Figura 3.6: Para cada imagen de entrenamiento, calcular el Mapa de Texton y obtener el histograma.

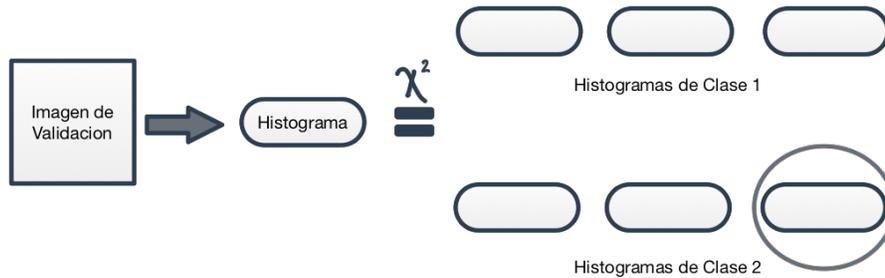


Figura 3.7: Clasificación de la imagen de validación a partir de la comparación del histograma de mapa de Texton contra los histogramas modelo de cada clase.

respectivas etiquetas de clase. Entonces, las imágenes del conjunto de validación son filtradas y el vector respuesta de cada pixel se compara contra el diccionario de *Textons* calculado en el entrenamiento. Se calcula el histograma de cada imagen y se compara contra el diccionario de histogramas para asignarle una etiqueta de clase. La clasificación a través de los histogramas generados a partir de los mapas de *Texton* están en la Figura 3.7.

### 3.3 Esquemas de clasificación propuestos utilizando Polinomios y Momentos Ortogonales Discretos

La clasificación de texturas involucra dos principales tareas: extracción y tratamiento de características y el clasificador. Como se ha mencionado antes, este trabajo de investigación está enfocado en la extracción de características. Los tres esquemas de clasificación propuestos están basados en las firmas de textura, Sección 3.1, y en los *Textons*, Sección 3.2. El primero genera una descripción estadística de la imagen a partir del análisis local por ventanas. El segundo esquema reemplaza los bancos de filtros estándar de *Textons* por los polinomios ortogonales discretos bidimensionales. Y el tercero, combina el

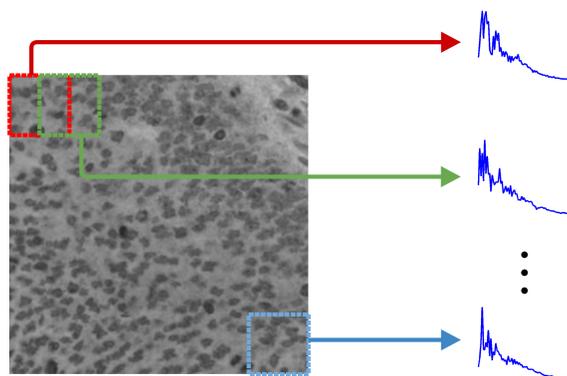


Figura 3.8: Descripción de una imagen a través de ventanas traslapadas.

análisis local por ventanas y el esquema de clasificación de los *Textons*. En todos los casos, se requiere de una etapa de entrenamiento y otra de validación para la evaluación de cada esquema.

#### 3.3.1 FIRMAS ESTADÍSTICAS DE TEXTURA

En [48], los autores proponen el cálculo de un solo vector de características para la imagen completa, esto implica que se calculan momentos de orden alto y, de acuerdo con [43], pueden producir errores debido a la inestabilidad de las oscilaciones. Este problema de inestabilidad también lo encontramos en los polinomios de Krawtchouk, Dual Hahn y Shmaliy. Ésta es una de las razones más importantes por la cual no se han utilizado en mayor proporción en la descripción de texturas en la literatura. Sin embargo, nosotros proponemos una modificación basada en ventanas cuadradas traslapadas que recorren la imagen evitando el cálculo de momentos de orden alto, asegurando así la estabilidad de la descripción.

Con las ventanas cuadradas traslapadas tenemos una sobredescripción de la imagen porque obtenemos firmas de textura para cada una de ellas. El traslape propuesto entre ventanas es del 50%; en la Figura 3.8 se muestra la descripción de cada ventana traslapada. El tamaño de ventana está limitado por la estabilidad del cálculo.

Después de calcular las firmas de textura correspondientes a cada ventana, se obtiene un conjunto de vectores  $M_i(s)$ , donde  $i$  es la posición o el número de la ventana, comenzando de izquierda a derecha y continuando de arriba hacia abajo. Este esquema es útil cuando el tamaño de las imágenes de la base de datos no es el mismo. La firma estadística de textura  $t$  se construye de la siguiente manera:

$$t = [\mu(M_i(0)), \sigma(M_i(0)), \kappa(M_i(0)), \dots, \mu(M_i(2N - 2)), \sigma(M_i(2N - 2)), \kappa(M_i(2N - 2))], \quad (3.4)$$

donde  $\mu$ ,  $\sigma$  y  $\kappa$  son el promedio, la desviación estándar y la curtosis, respectivamente. La ecuación 3.4 representa la propuesta para describir las características de textura de una imagen.

El esquema general de clasificación indica que se calcula una firma estadística de textura para cada imagen del conjunto de entrenamiento. Una vez calculadas las características, entran al clasificador,

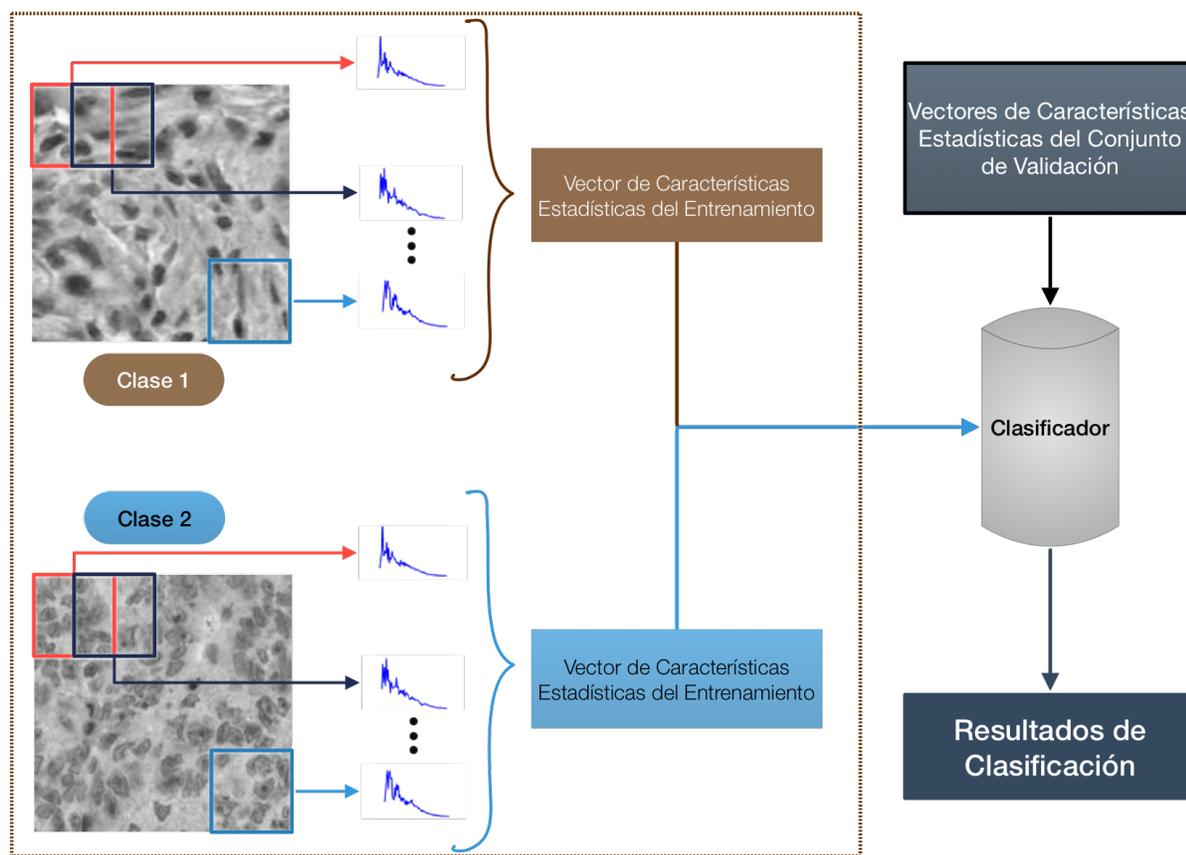


Figura 3.9: Esquema general de la clasificación con firmas de textura estadísticas.

que se encargará de generar una regla con la que posteriormente las características calculadas para el conjunto de validación serán clasificadas. En la Figura 3.9 se muestra el esquema general de la clasificación con firmas de textura estadísticas.

### 3.3.2 TEXTONS CON POLINOMIOS ORTOGONALES DISCRETOS

Una propuesta de esquema de clasificación está basada en la inclusión de los polinomios ortogonales discretos bidimensionales como banco de filtros en el cálculo de los *Textons*, descrito en la Sección 3.2. En las Figuras 2.1(b), 2.2(b), 2.5(b) y 2.7(b) se muestran los primeros órdenes de los polinomios bidimensionales de todas las bases mencionadas en la Sección 2.

El número de filtros del banco de filtros de estos polinomios es  $N^2$ , donde  $N$  es el orden máximo de los polinomios. Por ejemplo, si  $N = 50$  entonces el banco tendrá 2500 filtros. Es evidente que el tamaño del banco es muy superior comparado al tamaño que generan los bancos de Leung-Malik, Schmid y Máxima respuesta, vistos en la Sección 3.2. Con esta longitud del vector, se presenta el problema de dimensionalidad y, muy probablemente, una sobredescripción, contrastando con una de las ideas fundamentales en la clasificación de texturas que es el uso de vectores de menor longitud que entreguen los mejores resultados.

### 3.3 Esquemas de clasificación propuestos utilizando Polinomios y Momentos Ortogonales Discretos

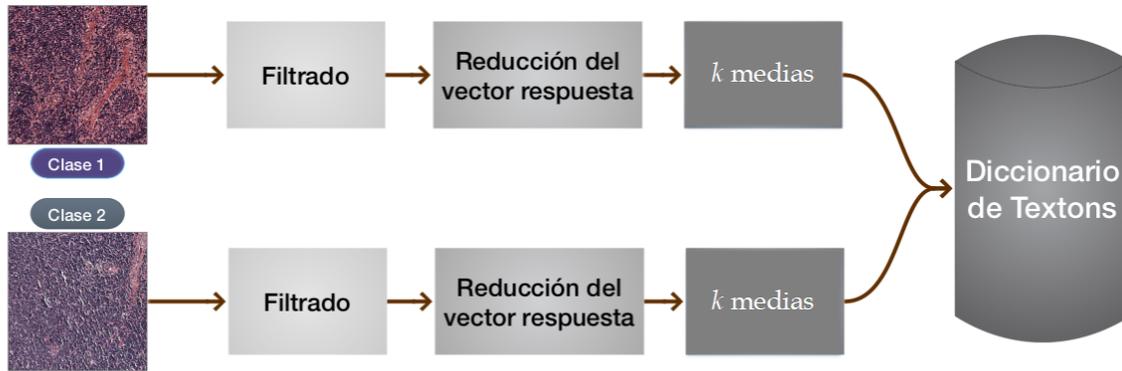


Figura 3.10: Cálculo del diccionario de Textons con Polinomios Ortogonales Discretos.

El vector respuesta de cada pixel que se obtiene a partir de la convolución entre la imagen  $I(x, y)$  y el banco de filtros  $r_{mn}(x, y)$  se define como:

$$VR_{m,n} = [FR_{0,0} \ FR_{1,0} \ FR_{2,0} \ \dots \ FR_{M-1,0} \ FR_{0,1} \ FR_{1,1} \ \dots \ FR_{M-1,1} \ \dots \ FR_{M-1,N-1}], \quad (3.5)$$

donde  $FR$  es la respuesta de ese pixel al filtro  $r_{m,n}$ ,  $m$  y  $n$  son el orden del polinomio unidimensional en la dimensión  $x$  y  $y$ , respectivamente,  $M$  y  $N$  son la longitud de los polinomios en  $x$  y  $y$ . Entonces, se proponen las siguientes maneras para reducir el tamaño del vector respuesta:

1. **La máxima respuesta de cada orden:** Dado el orden de los polinomios bidimensionales  $s$ , el vector respuesta reducido es:

$$M(s) = \max_{s=m+n} |VR_{m,n}| \quad (3.6)$$

con  $s = 0, 1, \dots, M + N - 2$ .

2. **La suma de las respuestas de cada orden:** Dado el orden de los polinomios bidimensionales  $s$ , el vector respuesta reducido es:

$$S(s) = \sum_{s=m+n} |VR_{m,n}| \quad (3.7)$$

con  $s = 0, 1, \dots, M + N - 2$ .

Los nuevos vectores respuesta tienen la misma longitud  $M + N - 2$ , que es considerablemente menor al tamaño original.

Cada imagen aporta el mismo número de vectores respuesta que de pixeles dentro de la misma. Entonces, al igual que en los Textons originales, seleccionamos  $k$  vectores que se obtienen a través del agrupamiento con  $k$ -medias para integrar un diccionario de Textons. La generación del diccionario se muestra en la Figura 3.10. Posteriormente, se generan los mapas de Texton para cada imagen de entrenamiento. Finalmente, se calculan los histogramas de los mapas de Texton de entrenamiento para conformar el conjunto de características de clasificación.

En cuanto a la validación, se calculan los mapas de Texton de cada imagen de este conjunto. Después, se obtienen los histogramas y, a través de una medida de similitud, se encuentra el histograma

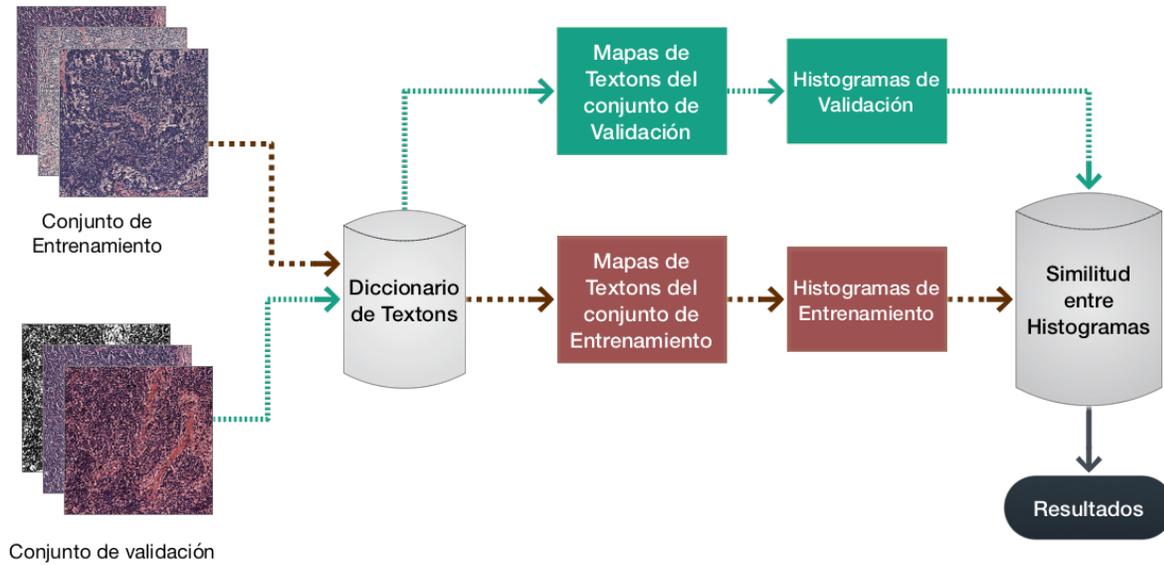


Figura 3.11: Clasificación con histogramas de Textons con Polinomios Ortogonales Discretos.

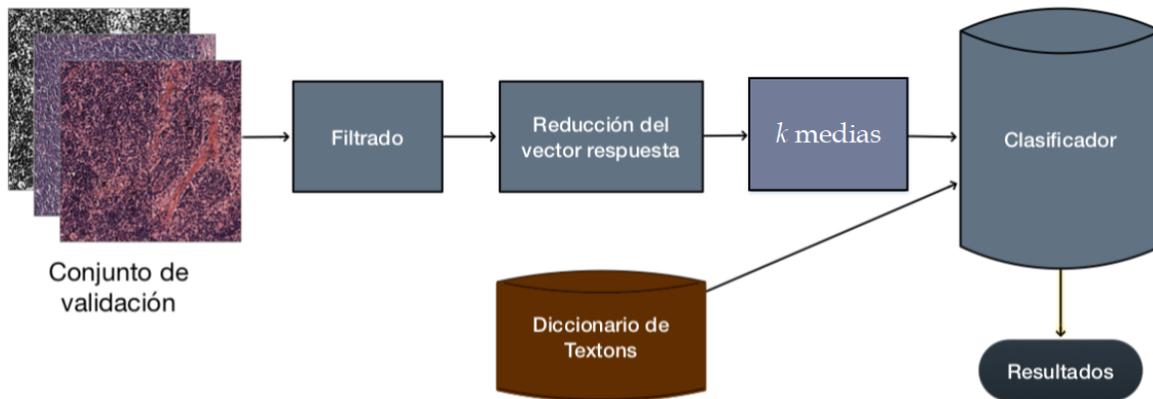


Figura 3.12: Clasificación con el diccionario de Textons con Polinomios Ortogonales Discretos.

más parecido dentro del conjunto de clasificación y se asigna la etiqueta de clase correspondiente. La clasificación a través de la similitud entre histogramas de *Textons* de polinomios ortogonales discretos está en la Figura 3.11.

Otra opción más sencilla es usar al propio diccionario de *Textons* como conjunto de características de clasificación. Después, se calculan los  $k$  vectores para cada imagen del conjunto de características y los vectores son asignados a una clase a través de un clasificador (*Random Forest*, *KN* vecinos, Máquinas de Soporte Vectorial, etcétera.). Ahora se tienen  $k$  etiquetas de clase para cada imagen, entonces, se elige como etiqueta final a la que aparezca el mayor número de ocasiones. En la Figura 3.12 se observa la clasificación con el diccionario de *Textons* con polinomios ortogonales discretos.

### 3.3 Esquemas de clasificación propuestos utilizando Polinomios y Momentos Ortogonales Discretos

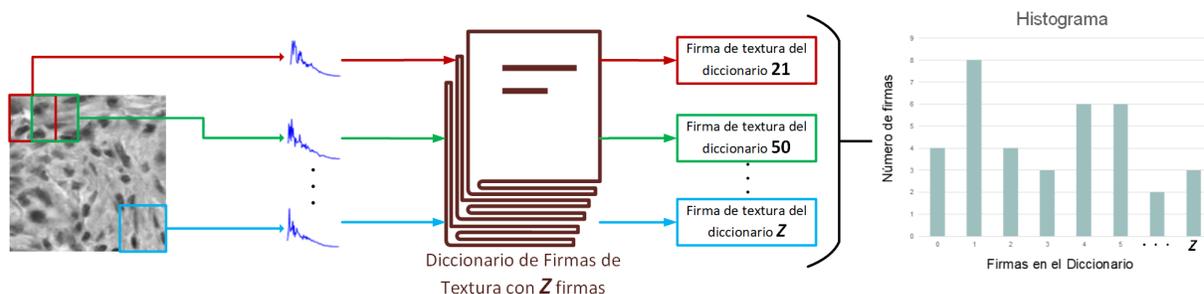


Figura 3.13: Cálculo de los histogramas de Firmas de Textura.

#### 3.3.3 FIRMAS DE TEXTURA $k$ -MEDIAS

Esta propuesta está basada en las firmas de texturas (Sección 3.1) dentro de un esquema de clasificación como el que utilizan los *Textons* (Sección 3.2).

Para el conjunto de entrenamiento, se calculan las firmas de textura de las ventanas traslapadas que se encuentran dentro de cada imagen. En lugar de calcular parámetros estadísticos (promedio, desviación estándar y curtosis), vistos en la Sección 3.1, se calculan  $k$  vectores a través de  $k$ -medias. Estos  $k$  vectores por cada imagen de entrenamiento compondrán un nuevo diccionario de firmas de textura.

De manera similar al cálculo de los mapas de *Textons*, se construye un mapa de firmas de textura para cada imagen, es decir, en lugar de tener que comparar el vector respuesta de cada pixel con el diccionario de *Textons*, ahora la firma de textura de cada ventana es *buscada* en el diccionario de firmas a través de una medida de similitud. Ahora que se sabe a que firma (palabra) del diccionario se asemeja más cada firma de textura de nuestra imagen analizada, se procede a crear el mapa de firmas de textura para la imagen y se contruye un histograma a partir de este mapa. En la Figura 3.13 se observa con mayor detalle el procedimiento necesario para la generación del histograma a partir de las firmas de textura.

Por otra parte, se calcula el mapa de firmas de textura para cada imagen del conjunto de validación a través de la comparación con el diccionario de firmas de textura del entrenamiento. La clase que se asigna a cada imagen corresponde a la etiqueta del pseudo histograma con mayor similitud. En la Figura 3.14, se observa el proceso que tiene la clasificación con histogramas de Firmas de Textura con los Momentos Ortogonales Discretos.

De nueva cuenta, se propone una opción más sencilla que utiliza al diccionario de firmas como conjunto de características de clasificación. Se calculan los  $k$  vectores de cada imagen del conjunto de validación a través de las firmas de textura de cada ventana dentro de cada una de las imágenes. Cada uno de esos  $k$  vectores entra a un clasificador que entrega una etiqueta de clase para cada vector. Finalmente, la etiqueta final es la de la clase con mayor número de repeticiones. El proceso para la clasificación a través del diccionario de Firmas de Textura con Momentos Ortogonales Discretos está en la Figura 3.15.

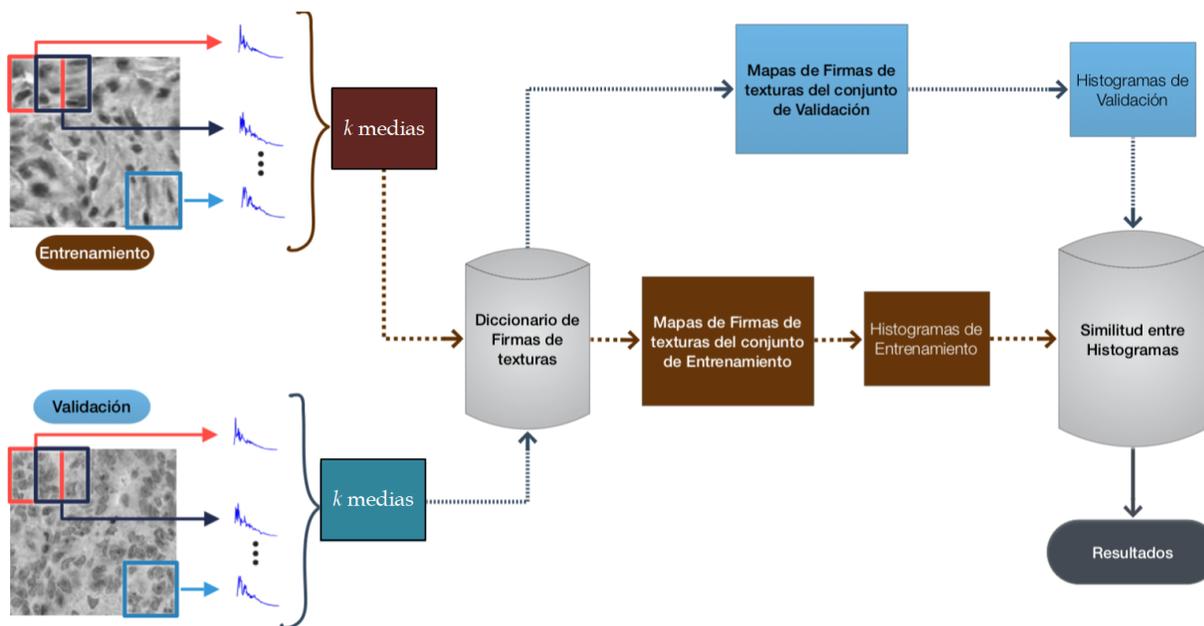


Figura 3.14: Clasificación con los histogramas de Firmas de Textura con Momentos Ortogonales Discretos.

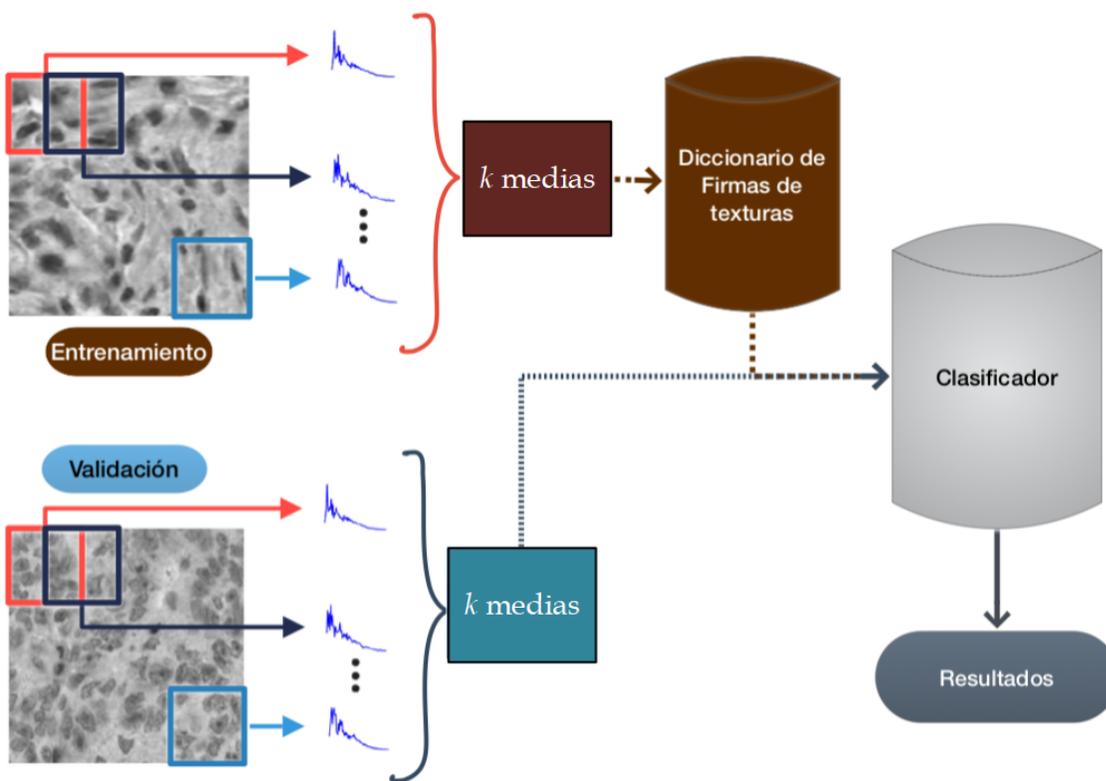


Figura 3.15: Clasificación con el diccionario de Firmas de Textura con Momentos Ortogonales Discretos.

# 4

## Tratamiento de los vectores de características

La estimación de la calidad de las características es un tema importante en el área del Aprendizaje de Máquina, en donde las tareas más importantes son la clasificación y regresión. Estas tareas requieren de una selección adecuada de parámetros o características para ofrecer los mejores resultados.

En los problemas de aprendizaje, hay cientos o miles de características potenciales que pueden describir a un objeto de entrada. La mayoría de los métodos de aprendizaje no funcionan adecuadamente bajo estas circunstancias porque, desde el punto de vista estadístico, la descripción con muchas características irrelevantes y ruidosas provoca problemas en la clasificación.

Los vectores de características que se proponen en la Sección 3 tienen una dimensión alta que depende del tamaño de la ventana traslapada de análisis. Por tanto, debido a las razones antes mencionadas, la reducción de esta dimensionalidad es considerada para mejorar el costo computacional y los resultados de clasificación. Las técnicas que se describen en esta sección son: selección de características, para encontrar las características más importantes para la clasificación, y reducción de dimensionalidad a través del análisis discriminante, que busca combinaciones lineales de las características que describan de mejor manera los datos.

### 4.1 Selección de características

La selección de características busca un subconjunto que idealmente es necesario y suficiente para describir un objetivo concreto. Para elegir estas características se necesita un método eficiente que cuantifique su relevancia en la descripción de un objeto en particular.

Existen muchas medidas para estimar la importancia o calidad de las características. Si el problema

es de clasificación, algunos ejemplos de estas medidas son: ganancia de información [90], índice Gini [91], medida de distancia [92], medida J [93], longitud mínima de descripción (MDL) [94] y las pruebas estadísticas: chi cuadrada,  $\chi^2$ , y la prueba  $G$  [95]. No obstante, todas estas medidas asumen que los atributos o características son independientes condicionalmente a la clase a la cual pertenecen, por lo tanto, no funcionan adecuadamente cuando hay dependencia entre los atributos y tienen pocas posibilidades de entregar una buena hipótesis sobre la selección de los parámetros más importantes.

Además de las medidas mencionadas anteriormente, usando la terminología de John *et al.* [96], la mayoría de los enfoques para reducir el número de características se pueden ubicar en dos categorías principales:

- **Métodos envolventes:** Estos métodos evalúan modelos múltiples usando procedimientos que agregan y/o eliminan características para encontrar la combinación óptima que maximiza el rendimiento del modelo. En esencia, los métodos envolventes son algoritmos de búsqueda que consideran a las características como las entradas y utilizan el rendimiento del modelo como la salida que será optimizada. Algunos métodos envolventes son: eliminación recursiva de características [97], algoritmos genéticos [98] y recocido simulado [99, 100].
- **Métodos de filtrado:** Estos métodos evalúan la relevancia de las características fuera de los modelos predictivos y, posteriormente, estiman el modelo solamente con las características que superan cierto criterio. Por ejemplo, para los problemas de clasificación, cada característica podría evaluarse individualmente para verificar si existe una relación plausible entre él y las clases observadas. Solo los predictores con relaciones importantes se incluirían en el modelo de clasificación. Saeys *et al.* [101] presentaron una compilación de diversas técnicas basadas en este método de selección y propusieron un esquema que combina varias de estas técnicas que mejora la selección de características en el campo de la bioinformática.

Ambos enfoques tienen ventajas e inconvenientes. Los métodos de filtrado suelen ser más eficaces desde el punto de vista informático que los métodos envolventes, pero el criterio de selección no está directamente relacionado con el modelo. Además, la mayoría de los métodos de filtrado evalúan cada característica por separado y, en consecuencia, pueden seleccionarse características redundantes o correlacionadas, y no se pueden cuantificar las interacciones importantes entre variables. La desventaja de los métodos envolventes es que muchos modelos son evaluados, los cuales pueden requerir de ajuste de parámetros y, por ende, un aumento en el tiempo de cálculo. También existe un mayor riesgo de sobre ajuste con los métodos envolventes.

### 4.1.1 ELIMINACIÓN RECURSIVA DE CARACTERÍSTICAS

En primer lugar, este proceso ajusta el modelo de clasificación utilizando todas las características del conjunto de entrenamiento. Cada característica se clasifica usando su importancia para el modelo. Sea  $S$  una secuencia ordenada que son valores candidatos para el número de características que serán conservadas ( $S_1 > S_2, \dots$ ). En cada iteración de la selección, se conservan las características  $S_i$  que

tuvieron los mejores resultados de clasificación, el modelo es reajustado y se evalúa el rendimiento. Se determina el valor de  $S_i$  con el mejor rendimiento y las mejores características  $S_i$  que se utilizan para ajustar el modelo final. El Algoritmo 4.1 tiene una definición más completa.

El algoritmo tiene un paso opcional (línea 9), en donde las clasificaciones se vuelven a calcular en el modelo con el número de características reducidas. Svetnik *et al.* [97] demostraron que, para los modelos de *Random Forest*, hubo una disminución en el rendimiento cuando las clasificaciones fueron recalculadas en cada paso. Sin embargo, en otros casos cuando las clasificaciones iniciales no son buenas (por ejemplo, los modelos lineales con predictores altamente colineales), recalculando las clasificaciones puede mejorar ligeramente el rendimiento.

---

**Algoritmo 4.1:** Eliminación Recursiva de Características

---

```

1 Entrenar el modelo con el conjunto completo de entrenamiento
2 Calcular el rendimiento del modelo de clasificación
3 Calcular la importancia de las variables
4 for Cada tamaño de subconjunto  $S_i$ ,  $i = 1 \dots S$  do
5     Mantener las características  $S_i$  más importantes
6     (Opcional) Pre procesamiento de los datos
7     Ajustar el modelo con el conjunto de entrenamiento usando las  $S_i$ 
        características
8     Calcular el rendimiento del modelo
9     (Opcional) Recalcular la importancia de cada predictor
10 end
11 Calcular el rendimiento sobre  $S_i$ 
12 Determinar el número apropiado de características
13 Usar el modelo correspondiente al  $S_i$  óptimo

```

---

Se puede presentar un problema de sobre ajuste por parte del predictor si es que éste se concentra en ciertas características del conjunto de entrenamiento que no se encuentren en el conjunto de validación. Por ejemplo, el algoritmo de clasificación *Random Forest*, descrito en el Apéndice B.3, evita este problema al seleccionar aleatoriamente un cierto número de características para cada árbol de decisión, esta aleatoriedad evita que los árboles estén correlacionados entre sí y que no se le de una relevancia excesiva a algunas características en particular.

El sesgo de selección [102] es un problema potencial de sobre ajuste en el que la envolvente puede enfocarse en los matices del conjunto de entrenamiento que no se encuentran en muestras o características subsecuentes. Por ejemplo, supongamos que se recolectó un gran número de características no relevantes y una de ellas se correlacionó aleatoriamente con el resultado. El algoritmo daría una buena posición a esta característica y el error de predicción (en el mismo conjunto de datos) se reduciría. Tomaría un proceso diferente de prueba y validación para descubrir que esta característica no es relevante.

En el algoritmo actual, los datos de entrenamiento se utilizan para tres propósitos básicos: selección

del conjunto de características, ajuste del modelo y evaluación del rendimiento. A menos que el número de muestras sea grande, especialmente en relación con el número de características, un conjunto de entrenamiento estático puede no ser capaz de satisfacer estas necesidades.

#### 4.1.2 RELIEFF

El algoritmo *Relieff* [103] se muestra en el Algoritmo 4.2. este método también pertenece a los métodos envolventes, es una extensión del algoritmo *Relief* [104, 105], y sus ventajas son que no está limitado a problemas de dos clases, es más robusto y puede trabajar con datos incompletos y con ruido. De manera similar a *Relief*, *Relieff* elige aleatoriamente el vector  $R_i$ , busca a los  $k$  vecinos más cercanos de la misma clase, llamados *nearest hits*  $H_j$ , y a los  $k$  vecinos más cercanos de cada una de las otras clases, llamados *nearest misses*  $V_j(C)$ . La contribución de los errores para cada clase se pondera con la probabilidad *a priori* de la clase  $C$ ,  $P(C)$ , que es estimada a partir del conjunto de entrenamiento. Entonces, la estimación de pesos  $W[A]$  es actualizada dependiendo de  $R_i$ ,  $H_j$  y  $V_j(C)$ , la fórmula de la actualización se escribe como:

$$W[A] = W[A] - \sum_{i=1}^k \frac{\text{diff}(A, R, H_i)}{n \times k} + \sum_{C \neq \text{class}(R)} \left[ \frac{P(C)}{1 - P(\text{class}(R))} \sum_{i=1}^k \frac{\text{diff}(A, R, V_j(C))}{n \times k} \right]. \quad (4.1)$$

La idea es que el algoritmo pueda estimar la capacidad de los atributos para separar cada par de clases independientemente de cuáles sean las dos clases más cercanas entre sí. La normalización de las probabilidades *a priori* es necesaria ya que los  $k$  *near misses* de cada clase diferente tenderían a exagerar la influencia de las clases con un número pequeño de casos. El proceso completo se repite  $n$  veces, donde  $n$  es un parámetro definido por el usuario

---

**Algoritmo 4.2:** Algoritmo *Relieff*

---

**Input:** Vector de características  $R_i$  y las etiquetas de la clase

**Output:** La estimación de pesos  $W$  de la importancia de cada característica

```

1 Inicializar  $W[A] = 0.0$ 
2 for  $i = 1$  to  $n$  do
3     Selección aleatoria de  $R_i$ 
4     Encontrar a  $H_j$ 
5     for cada clase  $C \neq \text{class}(R_i)$  do
6         Encontrar a  $V_i(C)$ 
7         for  $A = 1$  to  $a$  do
8              $W[A] = W[A] - \sum_{i=1}^k \frac{\text{diff}(A, R, H_i)}{n \times k} + \sum_{C \neq \text{class}(R)} \left[ \frac{P(C)}{1 - P(\text{class}(R))} \sum_{i=1}^k \frac{\text{diff}(A, R, V_j(C))}{n \times k} \right]$ 
9         end
10    end
11 end

```

---

## 4.2 Análisis Lineal Discriminante

El Análisis Lineal Discriminante (LDA, por sus siglas en inglés), también conocido como análisis normal discriminante o análisis de función discriminante, es una generalización del discriminante lineal de Fisher, un método utilizado en estadística, reconocimiento de patrones y aprendizaje de máquina para encontrar combinaciones lineales entre características que describen dos o más clases en bases de datos. La combinación lineal resultante puede ser utilizada como clasificador o, como en la mayoría de las ocasiones, para reducir la dimensionalidad antes de una clasificación posterior.

El LDA está muy relacionado con el análisis de varianza (ANOVA) y con el análisis de regresión, el cual también intenta expresar una variable dependiente como una combinación lineal de otras características o medidas [106, 107]. Sin embargo, ANOVA utiliza variables categóricas independientes y una variable dependiente continua, mientras que el análisis discriminante tiene variables independientes continuas y una variable dependiente categórica (la etiqueta de clase) [108]. La regresión logística se parece más al LDA que al ANOVA, ya que también expresan o asocian una variable categórica a través de las variables independientes continuas. Estos métodos se utilizan más en aplicaciones donde no es posible suponer que las variables independientes están distribuidas normalmente, lo cual es una suposición fundamental del método LDA.

El LDA también está relacionado con el análisis de componentes principales (PCA, por sus siglas en inglés) y el análisis de factores, ya que ambos buscan combinaciones lineales de variables que expresen los datos de una manera más sencilla [109]. El LDA intenta modelar explícitamente la diferencia entre las clases de la base de datos. Por otro lado, PCA no tiene en cuenta ninguna información referente a la clase y el análisis factorial construye combinaciones de características basadas en diferencias en lugar de similitudes. El LDA también es diferente del análisis factorial porque no es una técnica de interdependencia: se debe hacer una distinción entre variables independientes y variables dependientes (también llamadas variables de criterio).

El LDA es considerado cuando las variables independientes están definidas en el dominio continuo. Cuando estas variables son categóricas, la técnica equivalente es el Análisis de Correspondencia Discriminante [110].

### 4.2.1 ANÁLISIS DISCRIMINANTE LINEAL DE DOS CLASES

Para este análisis, se toma en cuenta una base de datos de entrenamiento que contiene vectores de características  $t$ , cada vector con una etiqueta de clase  $l \in \{0, 1\}$ . El LDA asume que las funciones de densidad de probabilidad  $p(t|l = 0)$  y  $p(t|l = 1)$  tienen una distribución normal con sus respectivos parámetros de media y covarianza  $(\mu_0, \Sigma_0)$  y  $(\mu_1, \Sigma_1)$ . Bajo esta suposición, la solución bayesiana óptima es predecir puntos como si fueran de la segunda clase si el logaritmo de las razones de verosimilitud es mayor que algún umbral  $U$ , de modo que:

$$(t - \mu_0)^T \Sigma_0^{-1} (t - \mu_0) + \ln|\Sigma_0| - (t - \mu_1)^T \Sigma_1^{-1} (t - \mu_1) - \ln|\Sigma_1| > U. \quad (4.2)$$

Sin considerar otras suposiciones, el clasificador resultante se le conoce como Análisis Discriminante Cuadrático (QDA, por sus siglas en inglés). En cambio, el LDA realiza adicionalmente la suposición de homocedasticidad (es decir, que las covarianzas de clase son idénticas, tal que  $\Sigma_0 = \Sigma_1 = \Sigma$ ) y que las covarianzas tienen rango completo. En este caso, varios términos se cancelan:

$$t^T \Sigma_0^{-1} t = t^T \Sigma_1^{-1} t$$

y

$$t^T \Sigma_i^{-1} \mu_i = \mu_i^T \Sigma_i^{-1} t,$$

en donde  $\Sigma_i$  es una matriz Hermitiana, y el criterio de decisión anterior se convierte en un umbral en el producto punto:

$$w \cdot t > u,$$

para la constante de umbralización  $u$ , donde:

$$w = \Sigma^{-1} (\mu_1 - \mu_0)$$

y

$$u = \frac{1}{2} (U - \mu_0^T \Sigma^{-1} U).$$

Esto significa que el criterio de una entrada  $x$  que está en una clase  $y$  es solamente una función de la combinación lineal de las características del conjunto de entrenamiento.

Las suposiciones del LDA son las mismas que para el ANOVA. El análisis es muy sensible a los valores atípicos (*outliers*) y el número de muestras debe ser mayor que el número de características [111]. Otras suposiciones son:

- **Variables con distribución normal:** las variables independientes son normales para cada nivel de la variable de agrupación [111].
- **Homogeneidad de la varianza / covarianza (homocedasticidad):** las variaciones entre las variables del grupo son las mismas en todos los niveles de predictores. El análisis discriminante lineal se utiliza cuando las covarianzas son iguales, y que el análisis discriminante cuadrático se puede usar cuando las covarianzas no son iguales.
- **Multicolinealidad:** los resultados de clasificación pueden disminuir si es que existe una mayor correlación entre los vectores de características.

El LDA tiene cierta robustez frente a ligeras violaciones de estas suposiciones [112], y también se ha demostrado que puede ser confiable cuando se usan variables dicotómicas (donde a menudo se encuentran variables que no tienen distribución normal) [113].

El resultado del análisis discriminante canónico encuentra los  $C - 1$  ejes (en donde  $C$  es el número

de clases) que mejor separan las clases. Estas funciones lineales no están correlacionadas y definen un espacio  $C - 1$  óptimo a través del conjunto de datos que mejor separa las clases.

#### 4.2.2 ANÁLISIS DISCRIMINANTE MULTICLASE

En el caso donde hay más de dos clases, el análisis discriminante de Fisher puede extenderse para encontrar un subespacio que parece contener toda la variabilidad de clase [114]. Suponiendo que cada una de las clases tiene una media  $\mu_i$  y la misma covarianza  $\Sigma$ . Entonces, la dispersión entre la variabilidad de clase puede definirse por la covarianza de las muestras de las medias de las clases:

$$\Sigma_b = \frac{1}{C} \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T, \quad (4.3)$$

donde  $\mu$  es la media de las medias de cada clase. En este caso, la separación entre clases en una dirección  $w$  se define como:

$$S = \frac{w^T \Sigma_b w}{w^T \Sigma w}, \quad (4.4)$$

ésto significa que cuando la separación  $w$  es un vector generador de  $\Sigma^{-1} \Sigma_b$ , la separación será igual al valor generador correspondiente.

Si  $\Sigma^{-1} \Sigma_b$  es diagonalizable, la variabilidad entre las características estará contenida en el subespacio abarcado por los vectores generadores correspondientes a los  $C - 1$  valores generadores más grandes, puesto que  $\Sigma_b$  es de rango  $C - 1$ , al menos. Estos vectores generadores se utilizan principalmente en la reducción de dimensionalidad de las características, como en el PCA.

#### 4.2.3 ANÁLISIS DISCRIMINANTE DE FISHER

El Análisis Discriminante de Fisher (FDA) es uno de los métodos de reducción de dimensionalidad más usados, fue desarrollado por Ronald Fisher [106] para problemas con dos clases. Contrario al PCA, el FDA proyecta los vectores de características de tal manera que conserva la geometría de los mismos.

Sean  $\Theta^1 = \{\vartheta_1^1, \vartheta_2^1, \dots, \vartheta_{l_1}^1\}, \dots, \Theta^C = \{\vartheta_1^C, \vartheta_2^C, \dots, \vartheta_{l_C}^C\}$  las muestras de  $C$  clases, entonces la proyección de Fisher es dada por el vector  $w$  que maximiza:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}. \quad (4.5)$$

La matriz de dispersión inter clases  $S_B$  se define como:

$$S_B = \sum_{i=1}^C l_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (4.6)$$

donde  $\mu_i = \frac{1}{l_i} \sum_{j=1}^{l_i} \vartheta_j^i$  y  $\mu$  es el promedio general.

Por otro lado, la matriz de dispersión intra clases  $S_W$  es:

$$S_W = \sum_{i=1}^C S_i, \quad (4.7)$$

donde  $S_i = \sum_{j=1}^{l_i} (\vartheta_j^i - \mu_i)(\vartheta_j^i - \mu_i)^T$ .

Finalmente, la matriz de proyección óptima  $w^*$  la forman los  $C - 1$  valores generadores más importantes de  $S_W^{-1} S_B$ .

#### 4.2.4 ANÁLISIS DISCRIMINANTE MULTICLASE KERNEL-FISHER

El FDA tiene una limitación importante porque asume que las características tienen un comportamiento Gaussiano. Para superar este inconveniente, Scholkopf y Mullert [115] proponen una generalización no lineal al mapear los datos originales a algún otro espacio y se realiza el FDA dentro de él. Entonces, el objetivo es encontrar  $w^* \in \zeta$  que maximiza:

$$J(w) = \frac{w^T S_B^\Phi w}{w^T S_W^\Phi w}, \quad (4.8)$$

donde  $S_B^\Phi$  y  $S_W^\Phi$  son las matrices de dispersión inter clase e intra clase, respectivamente, en el espacio  $\zeta$  y  $\Phi$  es la función que mapea los datos de entrada al espacio producto interno  $\zeta$  de mayor dimensión (posiblemente infinito).

Una desventaja de este método es que no siempre es posible calcular la función  $\Phi$ , así que, el algoritmo fue reformulado en términos de productos internos y se propuso el Análisis Discriminante Kernel Fisher (KFDA) para problemas de dos clases. La idea principal radica en resolver las limitaciones del FDA calculando productos punto de los datos mapeados, sin necesidad de la función  $\Phi$ . En primer lugar, es necesario definir una matriz kernel de un producto interno,  $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$ , en lugar de calcular una función de mapeo explícita. Entonces,  $K(m, n) = k(X_m, X_n)$  donde  $X = \cup_{i=1}^C X^i$ . Los tipos de kernel más populares son: kernel lineal, kernel polinomial, kernel sigmoidal y kernel Gaussiano, y se definen en el Apéndice B.2.3, ya que las máquinas de soporte vectorial utilizan estas funciones para proyectar los datos. Entonces:

$$w^T S_B^\Phi w = \alpha P \alpha^T \quad (4.9)$$

y

$$w^T S_W^\Phi w = \alpha Q \alpha^T. \quad (4.10)$$

La matriz de dispersión entre clases  $P$  se define como:

$$P = \sum_{j=1}^C l_j (\mu_j - \mu)(\mu_j - \mu)^T, \quad (4.11)$$

donde  $\mu_j = \frac{1}{l_j} \sum_{n \in X^j} K(m, n)$  y  $\mu = \frac{1}{l} \sum_{n \in X} K(m, n)$ .

La matriz de dispersión intra clases  $Q$  se define como

$$Q = KK^T - \sum_{j=1}^C l_j \mu_j \mu_j^T, \quad (4.12)$$

tal vez sea necesario modificar a  $Q = Q + r\mathbf{I}$  para que su inversa exista.

Finalmente,  $\alpha^*$  se construye con los  $C - 1$  valores generadores más grandes de  $Q^{-1}P$  y la proyección queda como:

$$y = K\alpha^*. \quad (4.13)$$

Este método reduce la longitud el vector de características a uno de  $C - 1$ .

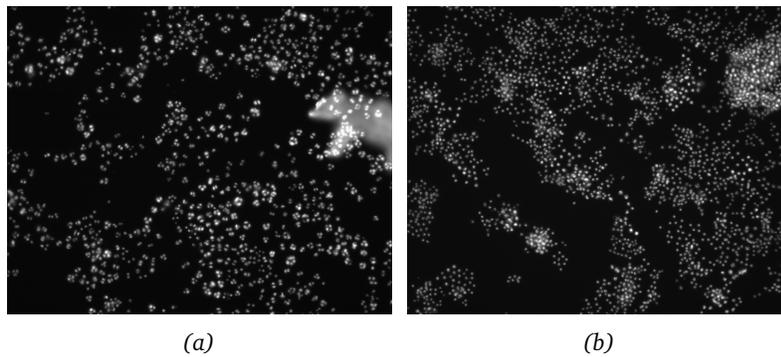


# 5

## Experimentos y Resultados

En esta sección, las propuestas descritas en la Sección 3 son aplicadas a bases de datos biomédicas. Se eligen este tipo de bases de datos por la relevancia que representa la aplicación médica porque un diagnóstico mejor y más rápido propiciará la elección adecuada del tratamiento a seguir. Por esa razón, el interés ha ido aumentando considerablemente por parte de los médicos en esta área del procesamiento de imágenes. Debido al avance tecnológico, principalmente en cuanto a la adquisición de imágenes y su resolución, se pueden notar patrones de textura que probablemente ayuden a distinguir entre diferentes tipos de tejidos.

En primer lugar, se presentan las bases de datos biomédicas utilizadas para la validación de los métodos propuestos en este trabajo de investigación. Las bases de datos pertenecen a al conjunto de bases pública llamado IICBU-2018. Después, se plantea el desarrollo del experimento basado en firmas de textura de momentos ortogonales discretos para la clasificación de estas bases de datos. Cabe mencionar que los resultados de este primer desarrollo experimental fueron publicados en el artículo titulado *A Comparative Study on Discrete Shmaliy Moments and Their Texture-Based Applications* [116] en la revista *Mathematical Problems in Engineering*. Posteriormente, se comparan los resultados entre las diferentes propuestas basadas en *Textons* y firmas de textura *k*-medias, mencionadas en la Sección 3. Finalmente, se mencionan los experimentos realizados durante este trabajo de investigación y que culminaron en su presentación y publicación en congresos internacionales.



**Figura 5.1:** Las dos clases definidas en la base de datos Binucleate son: (a) Binucleate y (b) Non-Binucleate.

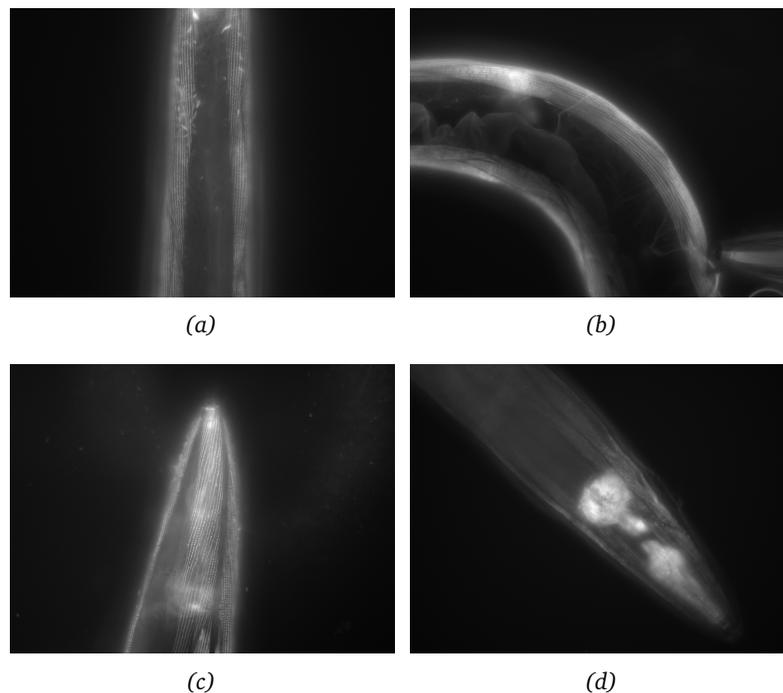
Base de datos	Clases	Imágenes	Formato	Microscopía
Binucleate	2	40	1280 × 1024 16-bit	Fluorescencia
Celegans	4	252	1600 × 1200 16-bit	Fluorescencia
CHO	5	340	512 × 382 16-bit	Fluorescencia
HeLa	10	860	382 × 382 16-bit	Fluorescencia
Liver aging	4	850	1388 × 1040 32-bit color	Campo claro
Lymphoma	3	375	1388 × 1040 32-bit color	Campo claro

**Tabla 5.1:** Descripción de las bases de datos de IICBU 2008.

## 5.1 IICBU - 20008

El conjunto de bases de datos pública **IICBU - 2008** fue presentada por Shamir [117] y está compuesta por 11 conjuntos de imágenes que representan un amplio rango de problemas en imagenología biológica. El formato o tipo de archivo en todos los conjuntos es TIFF (*Tagged Image File Format*), de 8, 16 o 32 bits, y el tamaño de las imágenes varía desde  $25 \times 25$  hasta  $1600 \times 1200$  píxeles. Las bases de datos incluidas están enumeradas en la Tabla 5.1. A continuación, se presenta una breve introducción de cada una de las bases de datos que fueron utilizadas en la validación de los métodos de clasificación propuestos:

- **Binucleate:** Esta base de datos fue adquirida con el objetivo de clasificar fenotipo celular binuclear y células normales. El fenotipo binuclear está asociado a fallas en la división celular y es de principal interés de los especialistas cuando buscan genes que tienen efectos en la división celular.



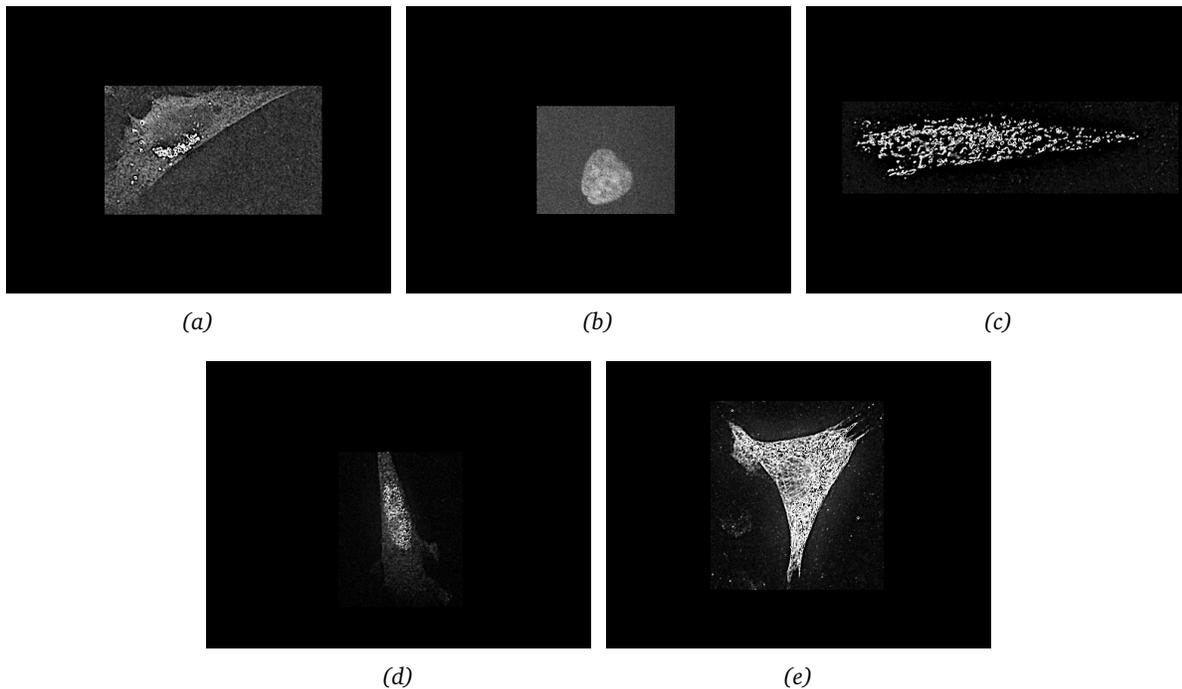
**Figura 5.2:** Las edades cronológicas de la base de datos *Celegans*: (a) 1, (b) 2, (c) 4 y (d) 8 días.

Muchos agentes utilizados en las quimioterapias también alteran la división celular. Las células incluidas en esta base de datos pertenecen a la *Drosophila melanogaster* y son mostradas en la Figura 5.1.

- ***Celegans*:** El nombre completo de esta base de datos es *Caenorhabditis elegans muscle aging* y su propósito principal es estimar la edad de los nematodos utilizando imágenes de tejido muscular. Las imágenes fueron capturadas con un microscopio de luz a diferentes edades. Las imágenes se asignaron a la edad cronológica respectiva, en lugar de la edad morfológica. Estas edades no están correlacionadas por completo porque el desarrollo físico entre individuos de la misma especie no es el mismo aunque tengan la misma edad cronológica.

*Celegans* tiene un objetivo diferente al resto de las bases de datos incluidas en IICBU-2008. El objetivo principal es encontrar una variable que correlacione las edades morfológicas y cronológicas. Las edades varían en el rango de 1, 2, 4 y 8 días. Una muestra de cada edad cronológica se presenta en la Figura 5.2.

- ***CHO* y *HeLa*:** El conjunto de datos Chinese Hamster Ovary (*CHO*) es el segundo conjunto de datos más antiguo de IICBU 2008 publicado originalmente en 1998. *CHO* es un conjunto de datos que se centra en la identificación de organelos subcelulares y estructuras celulares. En la Figura 5.3 se muestra una imagen por clase de la base de datos *CHO*. *HeLa* se publicó posteriormente en 2001 bajo la misma línea de investigación que *CHO*, pero proporciona más estructuras para la identificación. También contiene un conjunto de vectores de características numéricas que se calcularon para caracterizar los datos *HeLa*. En la Figura 5.4 se muestra una imagen por clase de la base de datos *HeLa*.



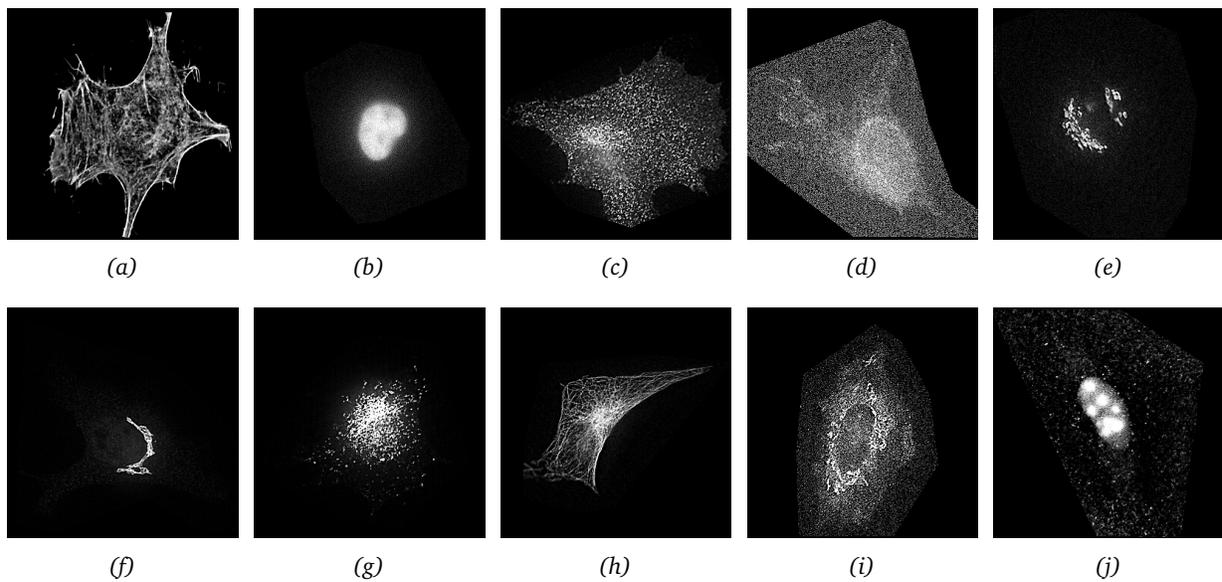
**Figura 5.3:** Las clases de la base de datos CHO son: (a) *giantin*, (b) *hoechst*, (c) *lamp2*, (d) *nop4* y (e) *tubulin*.

Uno de los puntos originales principales de investigación de estos conjuntos de datos es la capacidad de localizar proteínas en una célula. La ubicación de una proteína en relación con otras proteínas o estructuras funcionales conocidas, es un fuerte indicador del propósito de la proteína. Incluso se pueden diferenciar proteínas muy similares, ya que los procesos a menudo se localizan en una parte de la célula. Ambos conjuntos de datos se han investigado durante la última década y han sido evaluados por numerosos algoritmos [117–119].

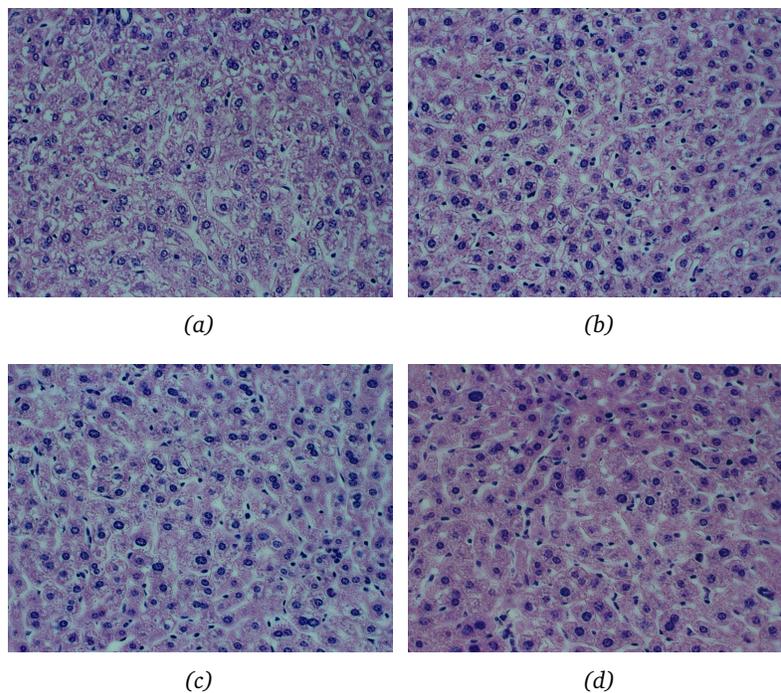
- **Liver aging:** El Atlas del Proyecto de Expresión Genética en la Edad de Ratones (AGEMAP) es un estudio hecho por el *National Institute of Aging* y el *National Institutes of Health* de los EEUU, que involucra 48 ratones machos y hembras, de cuatro edades diferentes (1, 6, 12 y 24 meses), con dietas *ad-libitum* (a placer) y con restricción calórica. Los hígados fueron extraídos de ratones sacrificados, fueron seccionados y entintados con Hematoxilina/Eosina (H+E). Las imágenes fueron adquiridas a través de microscopía de campo claro. En la Figura 5.5 se muestra un ejemplo de cada edad que compone la base de datos.

Las imágenes AGEMAP se pueden analizar en múltiples ejes de diferenciación: edad, sexo, dieta o individualmente para construir una variedad de problemas de clasificación. Los clasificadores pueden ser entrenados en hígados de la misma edad y dieta para predecir el sexo del ratón. También pueden ser entrenados para predecir la dieta del ratón. La edad es una variable continua, por lo que se debe tener cuidado al entrenar clasificadores para predecir la edad del ratón, ya que la edad cronológica no corresponde necesariamente a la edad morfológica.

La variabilidad de tinción es un desafío importante que impide el análisis cuantitativo de las



**Figura 5.4:** Las clases de la base de datos HeLa son: (a) actin, (b) dna, (c) endosome, (d) er, (e) golgia, (f) golgpp, (g) lysosome, (h) microtubules, (i) mitochondria y (j) nucleolus.



**Figura 5.5:** Las imágenes pertenecen a las diferentes edades cronológicas de los ratones sin dieta restrictiva de la base Liver aging: (a) 1, (b) 6, (c) 16 y (d) 24 meses.

muestras con H+E. Todas las muestras de AGEMAP fueron preparadas por el mismo experto patólogo, siguiendo un protocolo establecido y adecuado para su adquisición. Por lo tanto, la variabilidad de la tinción en este conjunto de datos es significativamente menor que las variaciones de tinción en conjuntos de imágenes como el de *Lymphoma*, que no se pudieron adquirir en condiciones restrictivas similares.

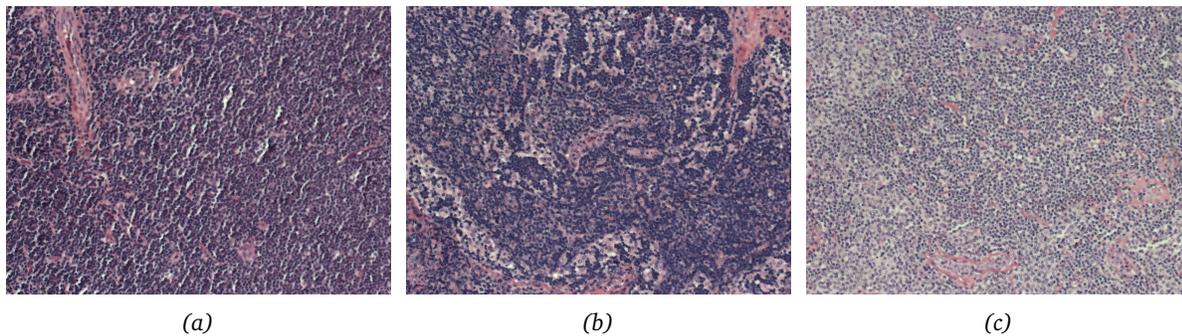


Figura 5.6: Clases de la base de datos *Lymphoma*: (a) CLL, (b) FL, y (c) MCL.

- **Lymphoma:** Las enfermedades hematológicas incluyen diferentes tipos de leucemia y linfomas. Aunque estas enfermedades no son habituales, han despertado interés médico en los últimos años. Por esa razón, este conjunto de datos es una de las bases de datos más populares porque se ha utilizado en varios estudios para clasificación [120, 121] y registro [122]. El conjunto de datos de imágenes está teñido con Hematoxilina/Eosina (H+E). La base de datos de *Lymphoma* está compuesta por tres tipos de enfermedades hematológicas: leucemia linfocítica crónica (LLC) con 133 muestras, linfoma folicular (FL) con 139 y linfoma de células del manto (MCL) con 122. La Figura 5.6 contiene muestras de cada clase del conjunto de datos del *Lymphoma*.

## 5.2 Experimento con firmas estadísticas de textura

Las firmas estadísticas de textura, presentadas en la Sección 3.3.1, es el método de descripción y clasificación de textura que, como se verá más adelante en este capítulo, mejores resultados obtuvo. Además, estos resultados son los que se publicaron en la revista *Mathematical Problems in Engineering* [116]. Antes de comenzar con la descripción de las imágenes a partir de los momentos discretos, las imágenes pasan por una etapa de preprocesamiento simple: las bases de datos a color RGB (*Liver aging* y *Lymphoma*) son transformadas a escala de grises ( $I_{gris}$ ) a través de la siguiente suma ponderada:

$$I_{gris} = 0.2989 * R + 0.5870 * G + 0.1140 * B, \quad (5.1)$$

donde  $R$ ,  $G$  y  $B$  son los canales rojo, verde y azul respectivamente.

Para el análisis local, todas las imágenes de las bases de datos son divididas en ventanas de análisis cuadradas y traslapadas. El tamaño de las ventanas varía desde los  $20 \times 20$  hasta los  $50 \times 50$  píxeles, el intervalo del tamaño de ventana es de  $10 \times 10$  píxeles y el traslape es del 50%. Este proceso es mostrado en la Figura 3.8. Otros parámetros fijos son: para los momentos de Krawchouk,  $p_1 = p_2 = 0.5$  (ecuación 2.47), y para los momentos Dual Hahn,  $a = c = 0$  y  $b$  es igual al tamaño de la ventana de análisis local (ecuación 2.66). El motivo de la elección de estos parámetros para los polinomios de Krawtchouk se debe a que en las ventanas traslapadas no es necesario analizar zonas en particular, por lo tanto, elegimos los parámetros que ubican a los polinomios bidimensionales justo en el centro de la ventana de análisis. El efecto de la variación de estos parámetros se muestra en la Figura 2.2. En cuanto a

los parámetros elegidos para los polinomios dual Hahn, no necesitamos analizar zonas particulares dentro de la ventana de análisis y tampoco es necesario expandir o comprimir a los polinomios, sino que se ajusten al tamaño de la ventana de análisis. Estas características de los polinomios dual Hahn se muestran en la Figura 2.6.

Las ventanas traslapadas son procesadas como se muestra en la Figura 3.9. Las firmas de textura, ecuación 3.1, son calculadas para cada ventana traslapada. La propuesta en este trabajo de investigación es el cálculo de las firmas estadísticas de textura. La firma estadística de textura es calculada a partir del promedio, desviación estándar y curtosis de todas las firmas de textura correspondientes a cada imagen, tal y como se muestra en la ecuación 3.4.

Las variables que se tomaron en cuenta para calcular los resultados de clasificación a partir de las firmas estadísticas de textura son:

- **Familias de Momentos Discretos:** Las firmas de textura se calculan a partir de las bases polinomiales discretas de Shmaliy, Tchebichef, Krawtchouk y Dual Hahn.
- **Tamaño de ventana de análisis:** los tamaños de las ventanas de análisis para calcular las firmas estadísticas de textura son  $20 \times 20$ ,  $30 \times 30$ ,  $40 \times 40$  y  $50 \times 50$  píxeles. Este tamaño está limitado por que el cálculo de los polinomios dual Hahn es inestable a partir de ese tamaño. Esa inestabilidad se debe a la acumulación del error en las recursiones, tal y como se menciona en la Sección 2.5.1. Por lo tanto, el tamaño de la ventana se limita para todas las familias de polinomios para comparar los resultados de las cuatro bases polinomiales bajo las mismas condiciones.
- **Procesamiento de las firmas de textura:** Antes de que las firmas de textura estadística entren a los clasificadores se les aplica el siguiente tratamiento: Análisis Discriminante Lineal (LDA), eliminación recursiva de características y la combinación de ambas, eliminación recursiva y LDA. Evidentemente, las pruebas también se realizaron sobre las firmas estadísticas sin tratamiento alguno.
- **Clasificadores:** Los algoritmos de clasificación utilizados son:  $K$  vecinos más cercanos (kNN), máquina de soporte vectorial con kernel lineal (SVM lineal) y con kernel Gaussiano (SVM gauss), *Random Forest* y *naive Bayes*. Los diferentes clasificadores están explicados en el Apéndice B.

La validación de la clasificación de las bases de datos se hizo a través de una validación cruzada de 5 grupos. La validación cruzada por grupos propone la división de los datos en  $X$  grupos con aproximadamente el mismo número de elementos, de tal forma que se entrena con  $X - 1$  grupos, dejando uno para validación. Finalmente, se repite este procedimiento de tal manera que cada grupo sea el grupo de validación en cada prueba. Los resultados finales de precisión en la clasificación son el promedio obtenido en los experimentos. En la Figura 5.7 se muestra un esquema que explica este tipo de validación.

En total, se calculan resultados para 4 familias de momentos, 4 tamaños de ventanas, 4 tipos de procesamiento de las firmas y 5 clasificadores distintos, es decir, 320 resultados diferentes. Resulta complicado mostrar todos los resultados obtenidos, así que deben ser acotados. No se puede prescindir

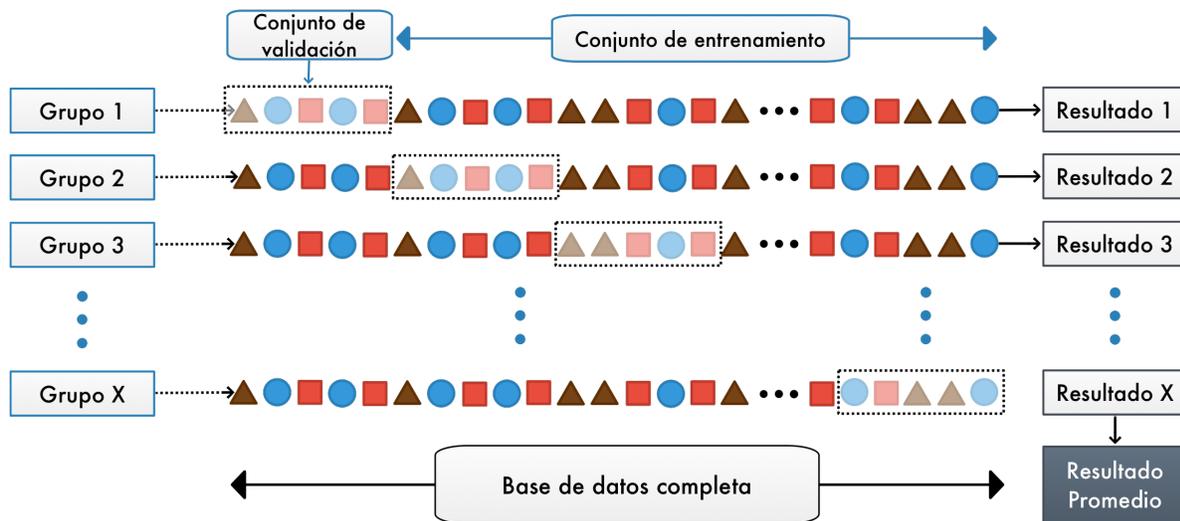


Figura 5.7: Diagrama general de la validación cruzada por grupos.

de las familias de momentos puesto que son de interés máximo en este trabajo de investigación. Tampoco se prescindirá de los resultados de los diferentes clasificadores porque resulta interesante observar el comportamiento de las características calculadas bajo diferentes esquemas de clasificación. Por lo tanto, quedan los diferentes tamaños de ventana y los tipos de procesamiento.

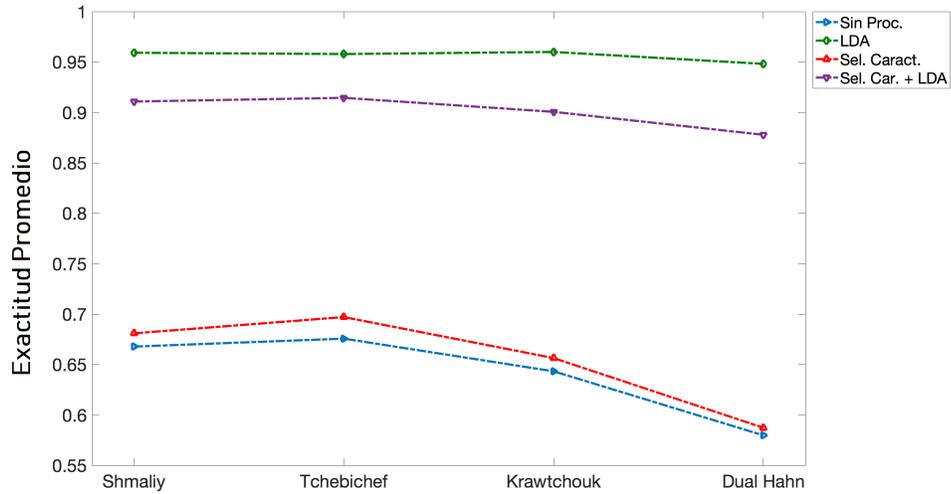
La métrica elegida para mostrar los resultados de las clasificaciones es la exactitud. La exactitud se define como:

$$Exactitud = \frac{\sum_{i=0}^C VP_i}{PT}, \quad (5.2)$$

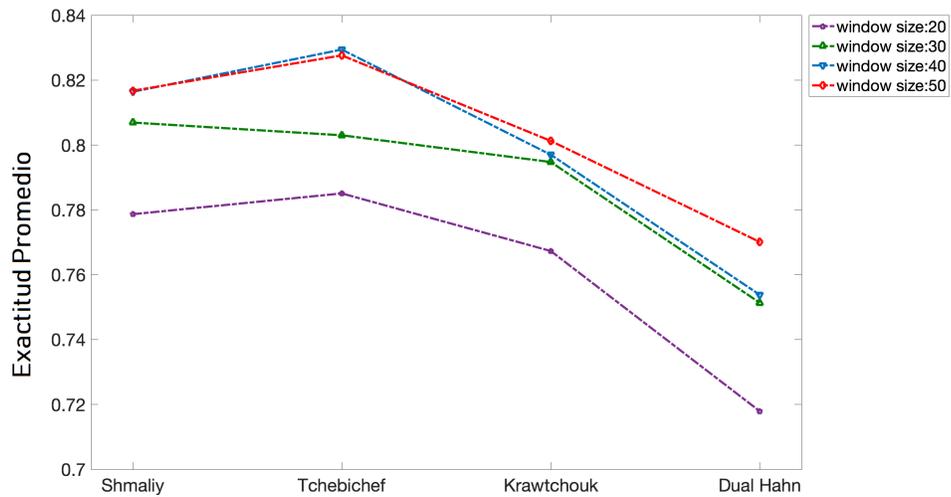
donde  $VP_i$  es el número de verdaderos positivos (muestras clasificadas correctamente) obtenidos para cada clase  $C$  y  $PT$  es el número total de muestras que tiene el conjunto de validación. La exactitud se puede presentar en términos de porcentaje o en valores normalizados entre 0 y 1.

En la Figura 5.8(a) se muestra el resultado promedio de exactitud en la clasificación, en todas las bases de datos, del tipo de procesamiento por cada familia de momentos y en la Figura 5.8(b) se presenta el resultado promedio del tamaño de ventana contra cada familia de momentos. En cuanto al tipo de procesamiento, se observa que para todas las familias de momentos, el análisis discriminante lineal tiene los mejores resultados y, además, presenta la mayor uniformidad entre ellos. Mientras que, en cuanto al tamaño de ventana, se observa que los tamaños  $40 \times 40$  y  $50 \times 50$  tienen los mejores resultados, sin embargo, es complicado discernir entre una y otra. Por lo tanto, los resultados mostrados sólo pertenecerán a los obtenidos con el análisis discriminante, descartando así a la selección de características y su combinación. En la Figura 5.9 se muestra la secuencia de pasos principales para el experimento de clasificación.

Una vez elegido el LDA como etapa de procesamiento para las firmas estadística de textura, las firmas resultantes son centradas: promedio cero y desviación estándar unitaria. Es necesario tener en cuenta que estas firmas tienen una longitud de  $C - 1$  características, donde  $C$  es el número de clases que contiene cada base de datos. En la Figura 5.10 se muestran las firmas resultantes para cada



(a)



(b)

Figura 5.8: Resultados promedio de exactitud en la clasificación de las familias de momentos en comparación con: (a) el tipo de procesamiento y (b) el tamaño de ventana.

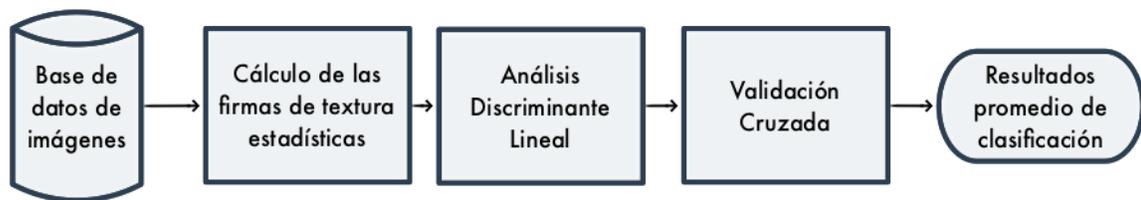
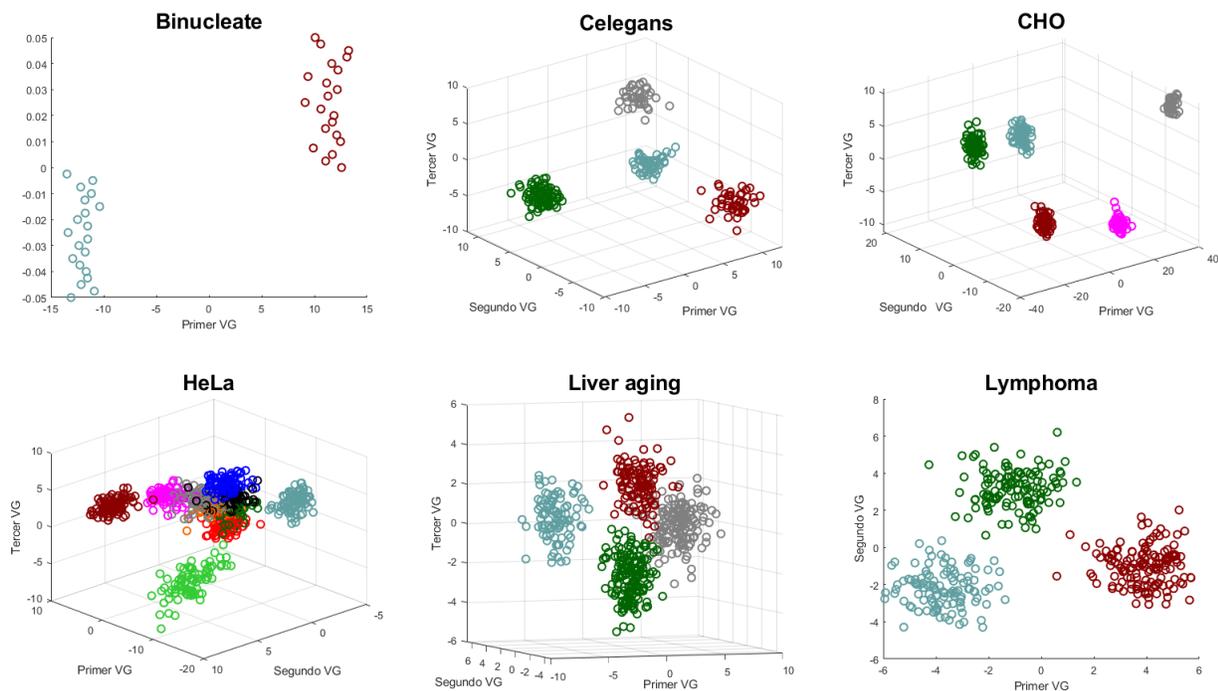


Figura 5.9: Procedimiento general para la clasificación basada en firmas de textura estadística y análisis discriminante.



**Figura 5.10:** Las firmas estadísticas de textura son transformadas a través del análisis discriminante. Los valores proyectados corresponden a los  $C - 1$  valores generadores de mayor magnitud, donde  $C$  es el número de clases que contiene cada base de datos. Las proyecciones obtenidas mejoran el agrupamiento entre clases y las separaciones entre clases son maximizadas.

base de datos. Los ejes de las subfiguras indican el valor generador al cual hacen referencia. Cabe resaltar que las bases de datos *CHO* y *HeLa* están mostradas parcialmente porque su longitud excede la representación tridimensional con 4 y 9 características, respectivamente.

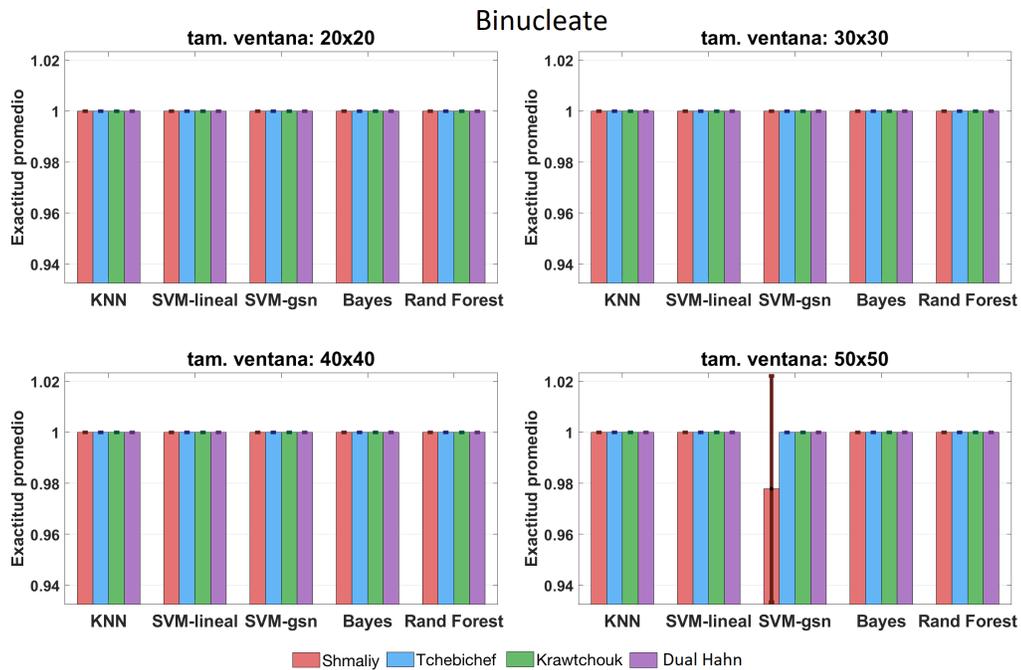
Los resultados promedio de exactitud para cada base de datos son mostrados en las siguientes figuras: *Binucleate* en la Figura 5.11, *Celegans* en la Figura 5.12, *CHO* en la Figura 5.13, *HeLa* en la Figura 5.14, *Liver aging* en la Figura 5.15 y *Lymphoma* en la Figura 5.16.

Cada subfigura de la Figura 5.11 a la 5.16 muestra los resultados para un tamaño de ventana específico. Las barras de color rojo corresponden a los momentos Shmaliy, las de color azul a los momentos de Tchebichef, las verdes a los momentos de Krawtchouk y las moradas a los momentos Dual Hahn. Las líneas verticales que aparecen al centro de las barras representan la desviación estándar del promedio de los resultados de la validación cruzada.

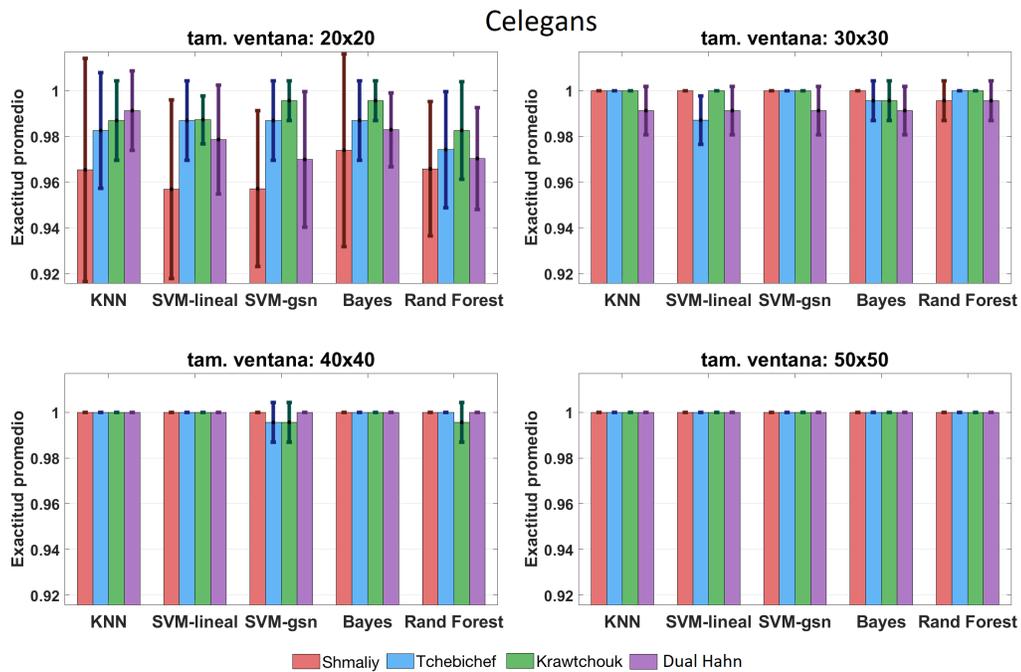
### 5.2.1 ANÁLISIS DE LA VARIANZA

El análisis de la varianza factorial [123], o ANOVA factorial, es realizado con el objetivo de encontrar si la elección del tamaño de ventana de análisis, el clasificador y/o el momento discreto es estadísticamente relevante porque los resultados de la clasificación son muy similares entre sí.

El ANOVA compara los promedios entre grupos e indica si alguno de esos grupos tiene cierta diferencia estadística significativa en comparación a los demás grupos. ANOVA prueba la hipótesis nula,



**Figura 5.11:** Resultados promedio de exactitud para la base de datos Binucleate. Esta base de datos contiene solamente dos clases. Esta base de datos tiene la particularidad de que el LDA proyecta y separa las clases de tal manera que la clasificación se vuelve trivial. Sin embargo, cuando la SVM-gsn es utilizada, hay un pequeño error causado por la influencia de los parámetros del kernel Gaussiano.



**Figura 5.12:** Resultados promedio de exactitud para la base de datos Celegans. La base de datos contiene cuatro clases. Los mejores resultados son obtenidos cuando el tamaño de la ventana de análisis es mayor a 20.

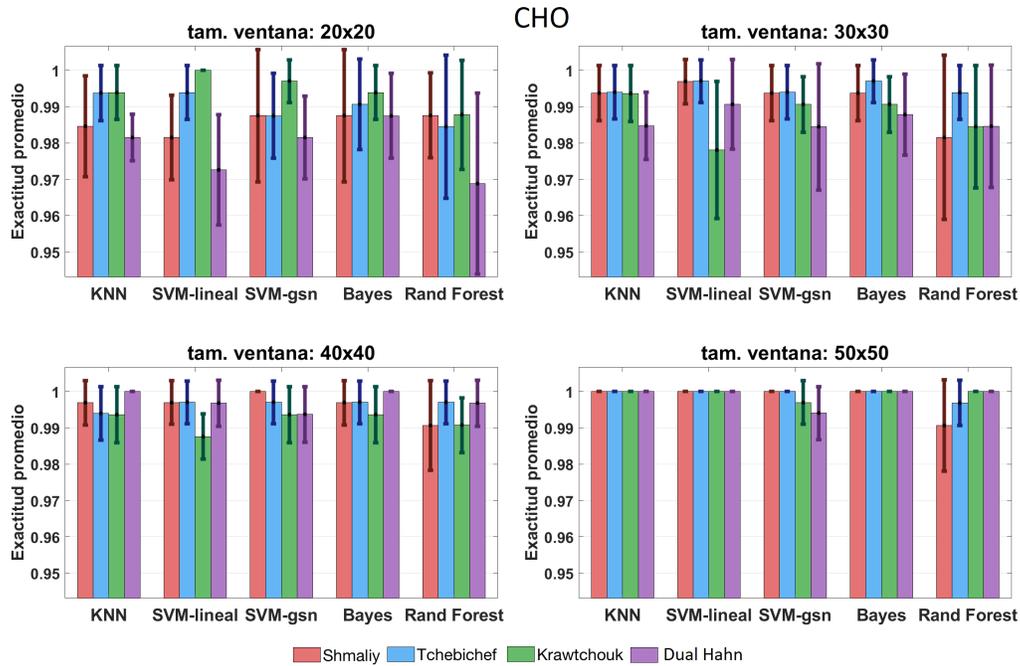


Figura 5.13: Resultados promedio de exactitud para la base de datos CHO. La base de datos consiste en 5 clases de estructuras celulares de ovario de hámster. En esta base de datos, las firmas de los momentos Shmaliy y Tchebichef muestran un rendimiento uniforme y similar cuando el tamaño de la ventana de análisis es mayor que 30.

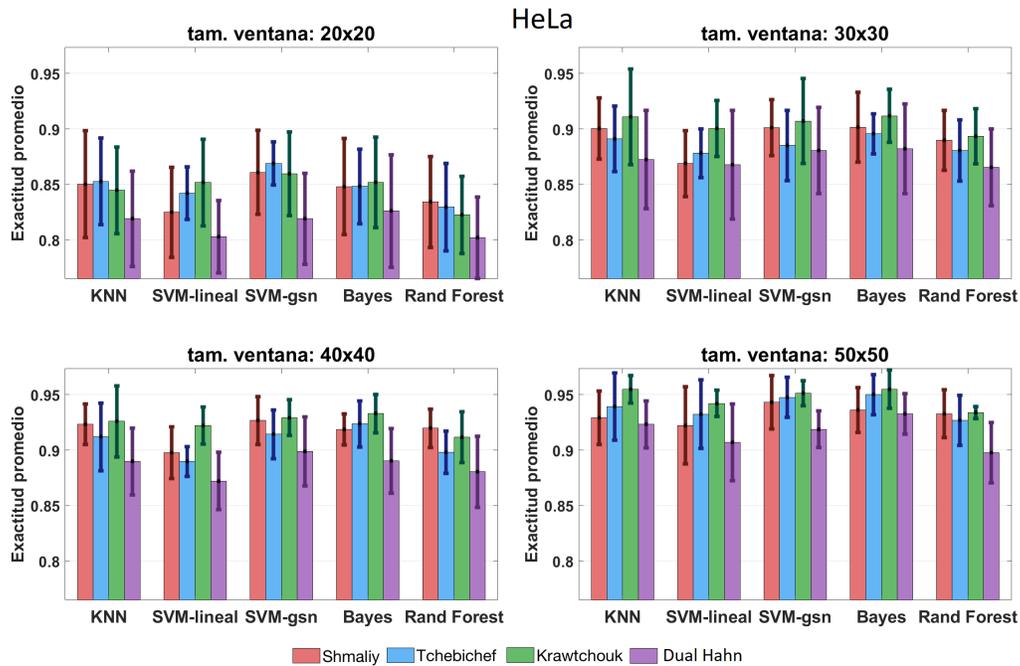
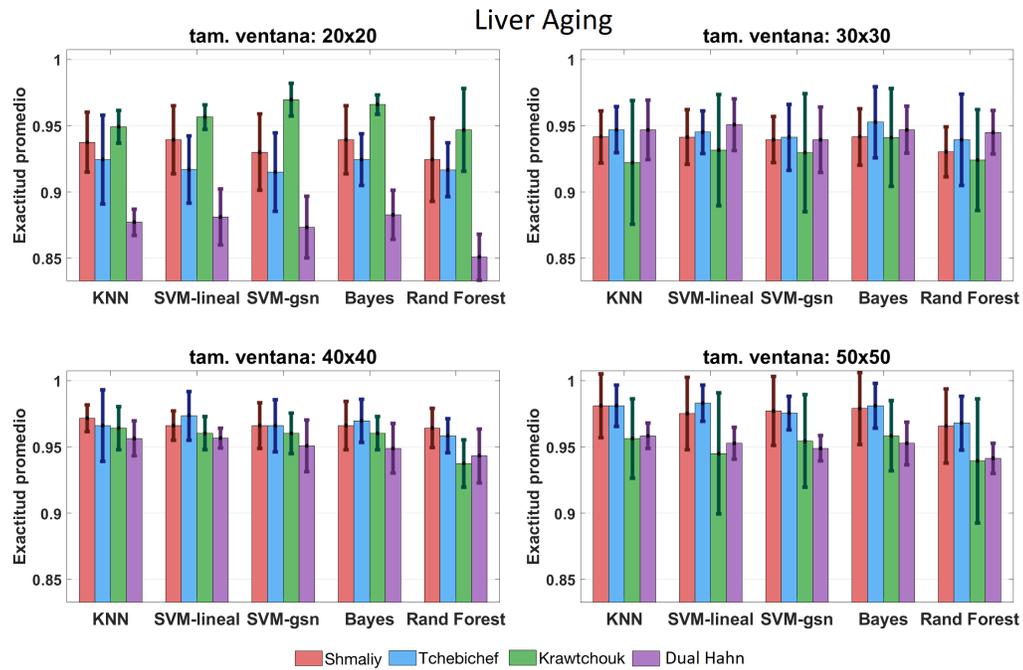
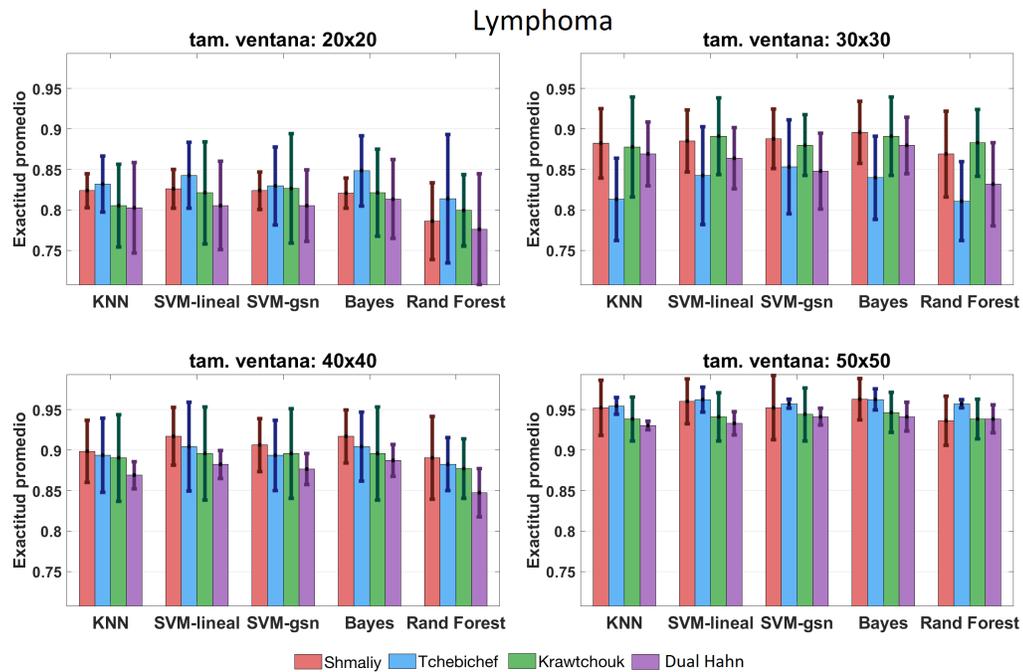


Figura 5.14: Resultados promedio de exactitud para la base de datos HeLa. La base de datos, como extensión de la base de datos CHO, contiene 10 clases de estructuras celulares de ovario de hámster. Los mejores resultados fueron obtenidos cuando el tamaño de ventana de análisis es mayor que 40.



**Figura 5.15:** Resultados promedio de precisión para la base de datos Liver aging. La base de datos contiene 4 clases diferentes. La base de datos tiene la particularidad de que el rendimiento de los clasificadores es similar entre ellos y la exactitud se incrementa conforme aumenta el tamaño de la ventana.



**Figura 5.16:** Resultados promedio de exactitud para la base de datos Lymphoma. La base de datos contiene 3 tipos de enfermedades hematológicas. Los mejores resultados se obtienen cuando la ventana de análisis es mayor que 40.

Base de datos	Ventana	Clasificador	Momento	Vent. & Clas.	Vent. & Mom.	Clas. & Mom.
Celegans	0	72.41	2.35	87.89	0	99.96
CHO	0	3.16	1.06	99.74	0	93.95
HeLa	0	0	0	99.67	73.34	95.48
Liver aging	0	0.7	0	99.8	0	99.37
Lymphoma	0	0.54	0.36	99.87	0.1	99.98

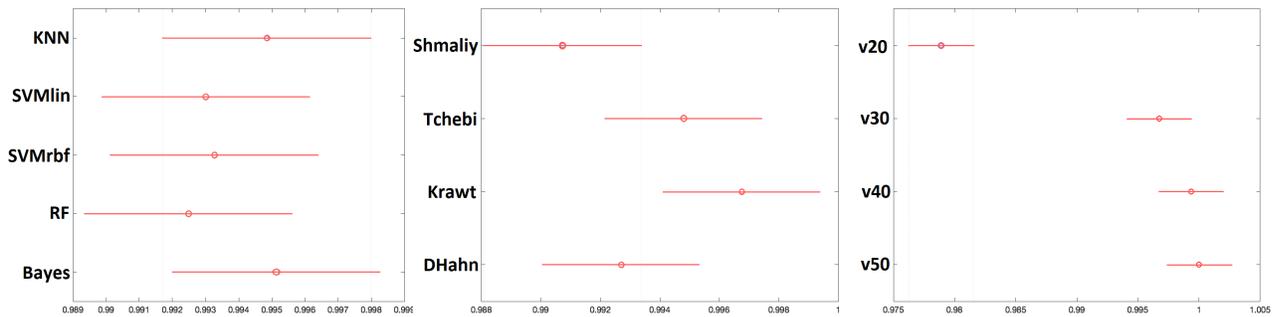
**Tabla 5.2:** Valores  $p$  del análisis de varianza. Las primeras tres columnas corresponden a cada variable independiente (Ventana, Clasificador y Momento), mientras que, las últimas tres columnas corresponden a las interacciones entre las variables independientes.

$H_0 = \mu_1 = \mu_2 = \dots = \mu_K$ , donde  $\mu$  es el promedio del grupo  $K$ . En otras palabras, la hipótesis nula es que ningún grupo tiene relevancia estadística. Si esta hipótesis es rechazada, la hipótesis alternativa,  $H_1$ , es aceptada e implica que al menos dos promedios de grupos son significativamente diferentes uno de otro.

En este experimento, la exactitud es la variable dependiente. El tamaño de ventana de análisis, el clasificador y el momento discreto son las tres variables nominales o independientes. Además, el ANOVA puede examinar la interacción entre las variables independientes porque las interacciones muestran que las diferencias entre los grupos no son uniformes en las categorías de las variables independientes. El resultado que arroja el ANOVA es el valor  $p$ , un valor  $p$  es calculado para cada variable nominal y las interacciones entre ellas. En general, si el valor  $p$  es menor que el 5 %, la hipótesis nula  $H_0$  es rechazada. La Tabla 5.2 muestra los valores  $p$ , en porcentajes, de las variables independientes y sus interacciones.

Con base en la Tabla 5.2 se observa que las variables independientes y la interacción ventana/momento son, en general, estadísticamente significativas para todas las bases de datos. Las gráficas del ANOVA de las variables independientes para la base de datos *Celegans* están en la Figura 5.17, para *CHO* en la Figura 5.18, para *HeLa* en la Figura 5.19, para *Liver aging* en la Figura 5.20 y para *Lymphoma* en la Figura 5.21. Cabe mencionar que la base de datos *Binucleate* no fue considerada para el análisis ANOVA porque, para todas las combinaciones de parámetros, los resultados de clasificación fueron muy altos. Esta base de datos es la más fácil de clasificar, tan fácil que su evaluación se vuelve trivial.

En primer lugar, la variable independiente más importante que observamos en la Tabla 5.2 es el tamaño de la ventana de análisis porque para todas las bases de datos se obtiene el valor  $p = 0$ . Se percibe, en general, que conforme aumenta el tamaño de la ventana, mejores son los resultados de clasificación y que, definitivamente, la ventana de tamaño  $20 \times 20$  es la que tiene el peor rendimiento. En particular, para *Celegans* se observa que a partir de la ventana de  $30 \times 30$  se obtiene los mejores resultados y con mayor uniformidad. Los mejores resultados se obtienen a partir de la ventana de  $40 \times 40$  para *CHO*, *Liver aging* y *Lymphoma*, mientras que para *HeLa* se obtienen solamente con la



**Figura 5.17:** ANOVA de las variables independientes de *Celegans*. En la imagen izquierda, están los resultados promedio de cada clasificador; se observa que KNN y naive Bayes tienen los mejores resultados, sin embargo, todos los demás clasificadores tienen resultados parecidos. En el centro, se ve que los momentos de Krawtchouk tienen el mejor resultado promedio. Los resultados de cada ventana se muestran en la imagen de la derecha, si la ventana mide  $30 \times 30$  o más, tendrá los mejores resultados.

ventana  $50 \times 50$ .

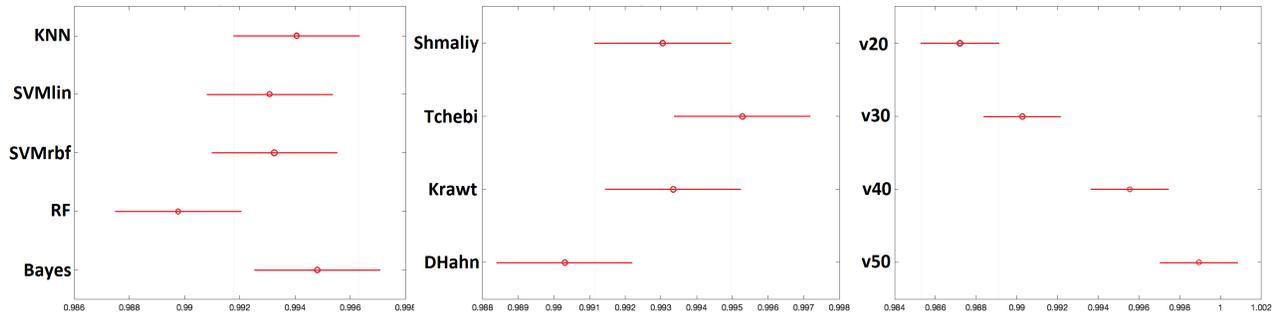
La siguiente variable más importante es la elección del momento discreto. En general, no se observa que algún momento en particular sea el mejor de todos, en cambio, se muestra que los momentos Dual Hahn no alcanzan tan buenos resultados de clasificación como los demás. En particular, para las bases *Celegans*, *HeLa* y *Lymphoma*, los mejores resultados promedio de clasificación se obtienen con los momentos de Krawtchouk. Los mejores resultados para la base *CHO* se obtienen con los momentos de Tchebichef y, para la base *Liver aging*, con los momentos de Shmaliy. Esta situación, en la que cada base de datos responda o se obtengan mejores resultados con cierta familia de momentos, puede ser debido a que estas mismas bases de datos están mayormente correlacionadas con las familias de momentos. Sin embargo, se recalca que los resultados obtenidos con estas tres últimas familias de momentos son muy parecidos entre sí.

La elección del clasificador también tiene relevancia estadística, excepto para la base de datos *Celegans*, es decir, que se obtendrán buenos resultados de clasificación sin importar el clasificador. Ahora, la relevancia significativa del clasificador se debe a que con *Random Forest* se obtienen resultados por debajo de los demás clasificadores. Una de las posibilidades es que el funcionamiento del *Random Forest* esté condicionado por la misma naturaleza del clasificador (Apéndice B.3) y al reducido número de características por la aplicación del LDA, resultando en la generación de árboles correlacionados.

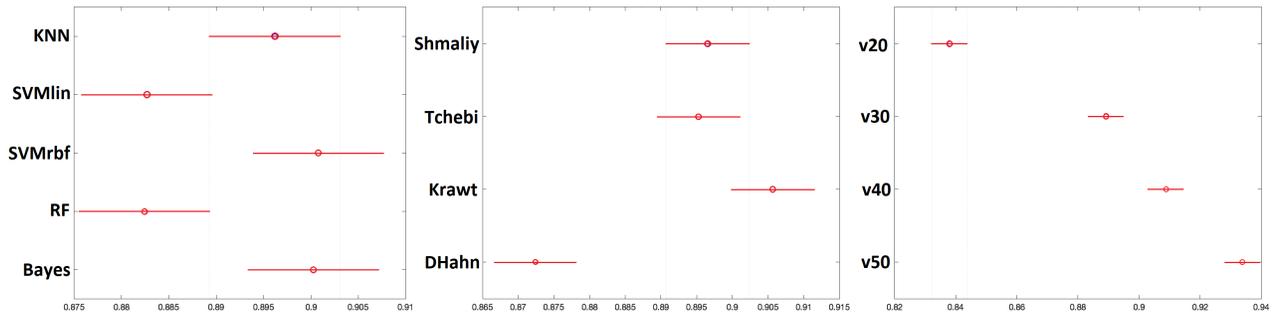
Finalmente, la interacción ventana/momento también es relevante estadísticamente, excepto para la base de datos *HeLa*. De nuevo, esta relevancia estadística no indica que alguna interacción sea la mejor, sino que, muestra que la combinación entre ventanas pequeñas junto con los momentos Dual Hahn obtienen los resultados más bajos de clasificación. Esta conclusión concuerda con lo dicho anteriormente para el tamaño de ventana y la elección del momento discreto.

Los mejores resultados de este experimento son comparados con otros trabajos que han utilizado las mismas bases de datos: Shamir *et al.* [117, 118], Siji *et al.* [119] y Meng *et al.* [124].

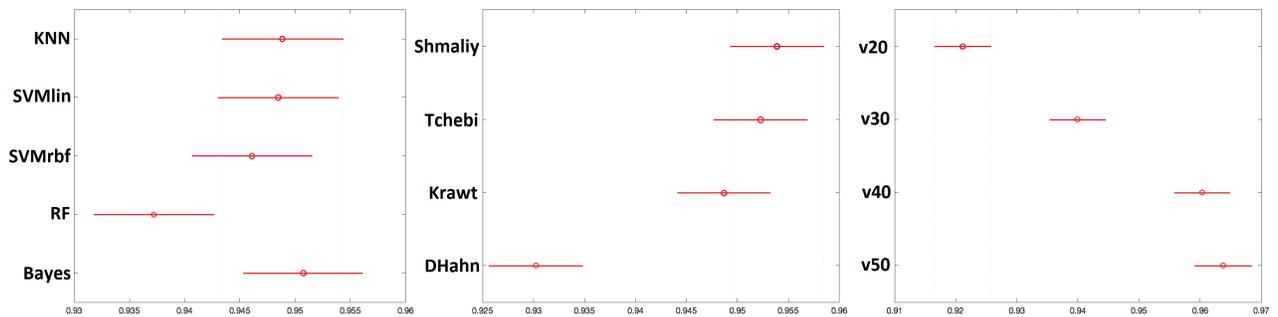
Shamir *et al.* presentaron la base de datos IICBU-2008 en [117]. Además, ellos utilizaron una herramienta muy popular llamada WND-CHRM [125, 126] para extraer y clasificar diferentes conjuntos de características, por ejemplo: transformada de Radon, coeficientes de Gabor, histogramas multiescala,



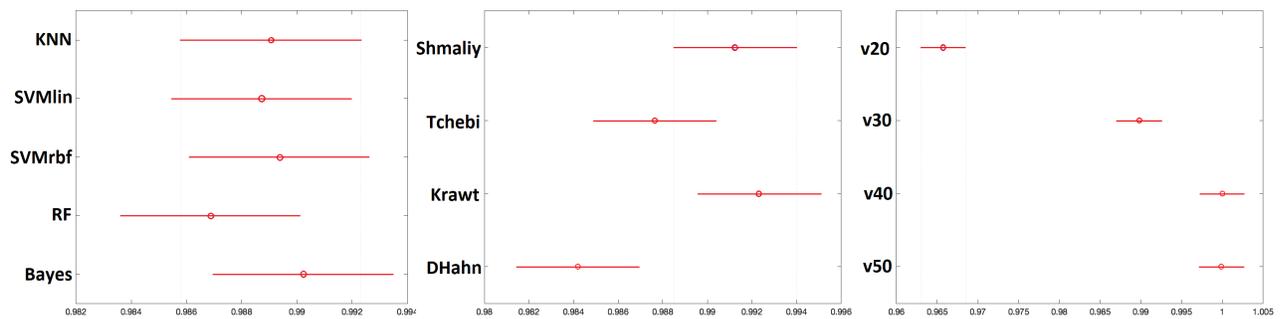
**Figura 5.18:** ANOVA de las variables independientes de CHO. El resultado promedio de los clasificadores se muestran en la imagen de la izquierda, todos los clasificadores tienen resultados parecidos, menos Random Forest. En el centro, están los resultados por momento, Dual Hahn muestra un peor rendimiento que los demás. En la imagen de la izquierda, se observa que, conforme aumenta el tamaño de ventana, los resultados mejoran.



**Figura 5.19:** ANOVA de las variables independientes de HeLa. En la imagen de la izquierda, se observa que los clasificadores SVM lineal y Random Forest tienen los peores resultados promedio. En la imagen del centro, se nota que los momentos Dual Hahn tiene el peor rendimiento. En la imagen de la izquierda, se observa que, conforme aumenta el tamaño de ventana, los resultados mejoran.



**Figura 5.20:** ANOVA de las variables independientes de Liver aging. En la imagen izquierda, los resultados promedio indican que el peor rendimiento pertenece al Random Forest. En la imagen central, Dual Hahn tiene el peor rendimiento, mientras que las demás familias tienen un rendimiento similar entre ellas. En la imagen derecha, se observa que conforme aumenta el tamaño de ventana, el rendimiento mejora.



**Figura 5.21:** ANOVA de las variables independientes de *Lymphoma*. En la imagen izquierda, se muestra que el rendimiento de los clasificadores es similar. En la imagen central, se muestra que los momentos Dual Hahn tienen, ligeramente, un peor resultado comparándolo con los demás momentos. Los resultados de cada ventana se muestran en la imagen de la derecha, si la ventana mide  $40 \times 40$  o más, tendrá los mejores resultados

Base de datos	Shmaliy	Tchebi	Krawt	DHahn	Shamir(1)	Shamir(2)	Siji	Meng
Celegans	100	100	100	100	53	-	75	-
CHO	100	100	100	100	93	94	99	-
HeLa	94.31	94.99	95.49	93.25	84	83	74	-
Liver aging	98.10	98.30	96.97	95.84	51	-	-	95.09
Lymphoma	96.29	96.27	94.67	94.13	85	-	-	84.79

**Tabla 5.3:** Comparación de resultados de exactitud promedio, en porcentaje, entre las firmas estadísticas de textura con momentos ortogonales discretos y los métodos encontrados en la literatura.

características de textura de Tamura, matriz de co-ocurrencia, etc. Posteriormente, Shamir *et al.* usaron la Transformada de Fourier, momentos de Tchebichef y la Transformada *Wavelet* para extraer características de textura y, de nuevo, utilizaron WND-CHRM para la implementación de la clasificación. Ellos reportaron resultados con las bases *CHO* y *HeLa*.

Siji *et al.* propusieron aplicar análisis de componentes principales (PCA) y análisis de Fisher basados en la selección de características para aplicaciones de recuperación de imágenes. Ellos también utilizaron las mismas características calculadas por el WND-CHARM para las bases de datos *Binucleate*, *Celegans*, *CHO* y *HeLa*. También utilizaros *naive* Bayes y SVM como clasificadores.

Meng *et al.* propusieron un modelo de clasificación llamado Modelado de Proyección de Subespacio Representativo Colateral (C-RSPM, por sus siglas en inglés) para las bases *Liver aging* y *Lymphoma*. Las bases utilizadas fueron *Liver aging* y *Lymphoma*.

La Tabla 5.3 resume y compara los resultados de este experimento con los de los autores mencionados anteriormente.

Base de datos	Exactitud	Descriptor	Ventana	Regla	K-Fold
Lymphoma	66.57 ± 6.36	TKS	21	Máximo	10
CHO	87.77 ± 6.26	T	21	Máximo	10
Liver Aging	60.87 ± 9.59	TKS	21	Máximo	10
HeLa	64.39 ± 6.21	TK	16	Máximo	10

Tabla 5.4: Resultados de clasificación con el Diccionario de Textons con Polinomios Ortogonales Discretos.

### 5.3 Experimento con Textons y Firmas de Textura de $k$ -medias

Para este experimento, las metodologías de clasificación propuestas en la Sección 3.3 son implementadas y probadas sobre las bases de datos *Lymphoma*, *CHO*, *Liver aging* y *HeLa* que pertenecen a la IICBU-2008. Además, en este experimento utilizamos los polinomios de Shmaliy, Tchebichef y Krawtchouk. Se ha descartado a los polinomios dual Hahn por el bajo rendimiento que mostraron en experimento de la Sección 5.2.

En primer lugar, se pone a prueba la propuesta de Textons con Polinomios Ortogonales Discretos de la Sección 3.3.2. El primer experimento utiliza como datos de clasificación al Diccionario de Textons con polinomios ortogonales (Figura 3.12) y consta de las siguientes características:

- Utiliza 5 tipos de descriptores o bancos de filtros: Tchebichef, Krawtchouk, Shmaliy, Tchebichef y Krawtchouk, y Tchebichef, Krawtchouk y Shmaliy.
- Los tamaños de cada filtro son de 11, 16 y 21 píxeles.
- Las reglas de reducción del vector de características son *Máximo*, ecuación 3.6, y *Suma*, ecuación 3.7.
- El número de  $X$ -Fold es: 3, 5 y 10.

En experimentos previos bajo este esquema, se observó que con ventanas mayores a 21 píxeles los resultados de clasificación bajaban, por lo tanto, se decidió conservar estos tres tamaños que obtuvieron los resultados más relevantes. Las reglas de reducción del vector de características forman parte de las propuestas de este trabajo de investigación. Finalmente, la variación del número de grupos en la validación cruzada es para observar si hay relación entre el número de imágenes de entrenamiento y los resultados de clasificación. Los mejores resultados de exactitud en la clasificación de este esquema para cada base de datos se muestran en la Tabla 5.4.

Base de datos	Exactitud	Descriptor	Regla
Lymphoma	70.4	S	Máximo
CHO	74.31	TK	Máximo
Liver Aging	96.02	TK	Máximo
HeLa	63.76	T	Máximo

**Tabla 5.5:** Resultados de clasificación con los histogramas de *Textons* con Polinomios Ortogonales Discretos.

El segundo experimento considera el cálculo de los histogramas a partir del Diccionario de *Textons* con polinomios ortogonales y que, a partir de la similitud entre histogramas, se puedan clasificar los datos de validación. La prueba es mucho más tardada, por lo tanto, las bases de datos se dividieron solamente en conjunto de validación y de entrenamiento, es decir, no hubo validación cruzada. El procedimiento se muestra en la Figura 3.11 y tiene las siguientes consideraciones:

- Utiliza 5 tipos de descriptores o bancos de filtros: Tchebichef, Krawtchouk, Shmaliy, Tchebichef y Krawtchouk, y Tchebichef, Krawtchouk y Shmaliy.
- Los tamaños de cada filtro son de 11, 16 y 21 píxeles.
- Las reglas de reducción del vector de características son *Máximo* y *Suma*.

La elección del tamaño de los filtros se realizó bajo el mismo criterio del experimento anterior. Los mejores resultados de clasificación para este procedimiento están en la Tabla 5.5.

Posteriormente, la clasificación es realizada bajo el esquema de las Firmas de textura de *k*-medias, descrito en la Sección 3.3.3. En primer lugar, el experimento utiliza al Diccionario de las Firmas de Textura como conjunto de datos (Figura 3.15) para que el clasificador *Random Forest* genere la regla de discriminación. Las características del experimento son:

- Utiliza 5 tipos de descriptores o bancos de filtros: Tchebichef, Krawtchouk, Shmaliy, Tchebichef y Krawtchouk, y Tchebichef, Krawtchouk y Shmaliy.
- Los tamaños de ventana son de 20, 30, 40 y 50 píxeles.
- El número de *k*-medias es 10, 15 y 20.
- El número de *X-Fold* es 3, 5 y 10.

El tamaño de las ventanas se eligieron para comparar directamente con los resultados del experimento con firmas estadísticas de textura de la Sección 5.2. El número de *k*-medias y el número de grupos

Base de datos	Exactitud	Descriptor	Ventana	K-Medias	K-Fold
Lymphoma	75.4 ± 7.04	S	30	20	5
CHO	97.55 ± 1.06	TK	50	15	3
Liver Aging	86.95 ± 3.75	TKS	30	10	5
HeLa	83.36 ± 4.01	TKS	20	15	10

**Tabla 5.6:** Resultados de clasificación con el diccionario de las Firmas de textura de  $k$ -medias.

de validación, es para saber si hay correlación entre los resultados de clasificación y estos parámetros. Los mejores resultados de clasificación de este esquema para cada base de datos se muestran en la Tabla 5.6.

El siguiente experimento considera el cálculo de histogramas para usarlos como datos de clasificación. En este caso, el experimento requiere de mucho tiempo, por esa razón, sólo se consideran dos tamaños de ventana, que serían los casos extremos, para observar si el tamaño de ventana tiene relevancia en este experimento. El esquema general del experimento está en la Figura 3.14 y tiene los siguientes parámetros experimentales:

- Utiliza 5 tipos de descriptores o bancos de filtros: Tchebichef, Krawtchouk, Shmaliy, Tchebichef y Krawtchouk, y Tchebichef, Krawtchouk y Shmaliy.
- Los tamaños de ventana son de 20 y 50 píxeles.
- El número de  $k$ -medias es 10, 15 y 20.

El número de  $k$ -medias y el número de grupos de validación, es para saber si hay correlación entre los resultados de clasificación y estos parámetros. Los mejores resultados de clasificación para cada base de datos en este esquema están en la Tabla 5.7.

Observando los resultados de los 4 experimentos, se concluye que la clasificación basada en las Firmas de Textura  $k$ -medias funciona mejor que el de *Textons* con polinomios ortogonales. En particular, el esquema que utiliza al Diccionario de Firmas de textura como datos de clasificación es el que tiene los mejores resultados observados en la Tabla 5.6. Esta conclusión conduce al siguiente experimento: agregar al esquema la Selección de parámetros a través del *Relieff* (Sección 4.1.2) y el Análisis Discriminante Kernel Fisher (Sección 4.2.4). Las características del experimento son:

- Utiliza 5 tipos de descriptores o bancos de filtros: Tchebichef, Krawtchouk, Shmaliy, Tchebichef y Krawtchouk, y Tchebichef, Krawtchouk y Shmaliy.
- Los tamaños de ventana son de 30, 40 y 50 píxeles.

Base de datos	Exactitud	Descriptor	Ventana	K-Medias
Lymphoma	69 ± 3.99	S	50	10
CHO	88.38 ± 3.47	TKS	20	10
Liver Aging	79.59 ± 5.39	T	20	10
HeLa	79.47 ± 1.08	TKS	20	10

**Tabla 5.7:** Resultados de clasificación con los histogramas de las Firmas de textura de *k*-medias.

Base de datos	Exactitud	Descriptor	Ventana	K-Medias	K-Fold	% relieff
Lymphoma	81.81 ± 5.48	TKS	50	20	10	90
CHO	98.15 ± 2.17	TK	50	15	10	90
Liver Aging	87.7 ± 5.01	TKS	50	10	10	70
HeLa	85.62 ± 3.75	TKS	30	15	10	90

**Tabla 5.8:** Resultados de clasificación con el diccionario de las Firmas de textura de *k*-medias con Relieff y Kernel Fisher.

- El número de *k*-medias es 10, 15 y 20.
- El número de X-Fold es 10.
- El tamaño del nuevo vector es con el 40, 50, 60, 70, 80 y 90 por ciento de las características ordenadas conforme al resultados de la selección de parámetros.
- El Kernel utilizado en el Análisis Discriminante es Gaussiano con una desviación estándar de 10.
- El clasificador es *KN* vecinos.

Los mejores resultados para las bases de datos se muestran en la Tabla 5.8.

Finalmente, la Tabla 5.9 muestra los mejores resultados de clasificación para cada experimento, incluyendo los de la Sección 5.2 y la clasificación con WND-CHARM. Se observa que los mejores resultados pertenecen a las firmas estadísticas de textura, el método más rápido de todos que, no obstante, también tiene ciertas condicionantes que limitan su aplicación. Todo esto se discute en la Sección 6.

Base de datos	WNDCHARM	Firmas estadísticas	Textons POD diccionario	Textons POD histogramas	Firmas K-medias diccionario	Firmas K-medias histograma	Firmas K-medias diccionario (KF/Relieff)
Lymphoma	85	96.29	66.57 ± 6.36	70.4	75.4 ± 7.04	69 ± 3.99	81.81 ± 5.48
CHO	93	100	87.77 ± 6.26	74.31	97.55 ± 1.06	88.38 ± 3.47	98.15 ± 2.17
Liver Aging	51	98.30	60.87 ± 9.59	96.02	86.95 ± 3.75	79.59 ± 5.39	87.7 ± 5.01
HeLa	84	95.49	64.39 ± 6.21	63.76	83.36 ± 4.01	79.47 ± 1.08	85.62 ± 3.75

**Tabla 5.9:** Comparación entre todos los resultados de exactitud de clasificación de todos los esquemas propuestos.

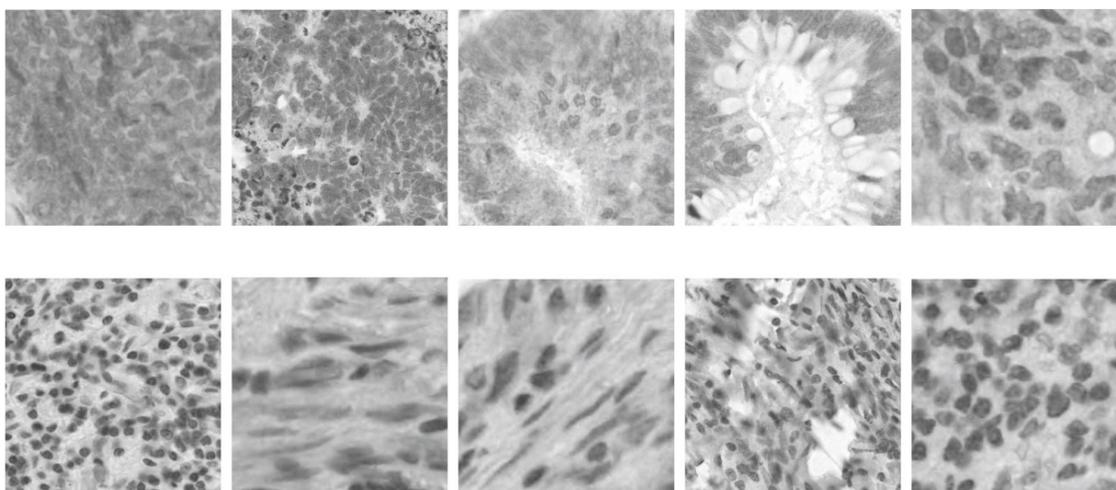
## 5.4 Estudios secundarios

Esta sección es una recapitulación de los experimentos que se realizaron y publicaron durante el tiempo que duró este trabajo de investigación. Estas pruebas están enfocadas en la clasificación de cáncer colorrectal [127] y enfermedades hematológicas [128].

### 5.4.1 EPITELIO Y ESTROMA EN CÁNCER COLORRECTAL

El cáncer colorrectal (CRC) es el tercer tipo más común de cáncer alrededor del mundo con más de 1.4 millones de casos registrados en 2012 [129]. La mortalidad de este tipo de cáncer es del 70 % en adultos mayores a los 65 años. Además, más de la mitad de la población desarrollará al menos un tumor intestinal benigno durante su tiempo de vida. En la mayoría de los datos, el CRC comienza como un pólipo o adenoma, caracterizado por la acumulación de células en la pared epitelial del tracto gastrointestinal. Una cantidad pequeña de esos pólipos se convierten en carcinomas tras la acumulación de alteraciones genéticas. Este proceso se conoce como la secuencia adenoma-carcinoma (ACS).

La metástasis a través de los vasos sanguíneos o linfáticos hacia el hígado o los pulmones es la causa principal de muerte y ocurre en más del 25 % de los pacientes. En contraste con la ACS, la metástasis colorrectal no está fuertemente correlacionada con alguna alteración genética, pero sí con las células sanas que rodean a los tumores. Estas células, llamadas estroma, están compuestas normalmente



**Figura 5.22:** Muestras de cáncer colorrectal en TMAs digitalizados, sólo el canal rojo. La primera fila muestra solamente el epitelio del tumor y la segunda fila muestra el estroma extraída desde un bloque de parafina.

por tejido conectivo y son esenciales para el mantenimiento del tejido epitelial sano y su contraparte enferma. Los cambios oncogénicos en el tejido epitelial modifican el compartimiento del estroma, que es responsable de establecer y propiciar un ambiente adecuado y eventualmente promueve el crecimiento y metástasis. Por lo tanto, el estroma juega un rol fundamental en el desarrollo y progresión de la enfermedad.

Los microarreglos de tejido (TMAs) son el estándar para determinar y monitorear la prevalencia de las alteraciones asociadas con la carcinogénesis colorrectal. Este procedimiento recoge secciones histológicas de tejidos y las coloca en un arreglo para formar un bloque de parafina. Los TMAs típicos pueden contener más de 1000 espacios que son utilizados para su interpretación simultánea. Entonces, la gran cantidad de información es el principal inconveniente para la clasificación manual y, por tanto, la motivación principal para el desarrollo de este estudio. Además, la identificación de las regiones de interés depende de la evaluación visual por parte del patólogo experto, lo cual genera un sesgo.

Utilizando los momentos ortogonales discretos de Tchebichef, se pretende clasificar la base de datos de imágenes de epitelio y estroma de cáncer colorrectal. Esta base de datos ya cuenta con el conjunto de validación y entrenamiento por separado. La base de entrenamiento tiene 400 imágenes de epitelio y 256 de estroma, mientras que la de validación contiene 425 imágenes de epitelio y 295 de estroma. Las imágenes tienen tamaños distintos, son imágenes a color (RGB) y es una base de datos pública proporcionada por Linder *at. al.* [130]. En la Figura 5.22 se muestran ejemplos del epitelio y estroma.

El procedimiento utilizado para la clasificación de tejidos de cáncer colorrectal es el siguiente:

1. Todas las imágenes, para entrenamiento y validación, son reescaladas a la mitad del tamaño original de cada una para acelerar el procesamiento, además, sólo se considera el canal rojo (R) del espacio de color RGB. Estas condiciones fueron propuestas por Linder *at. al.* [130] y las replicamos para comparar nuestra propuesta bajo las mismas condiciones de preprocesamiento.

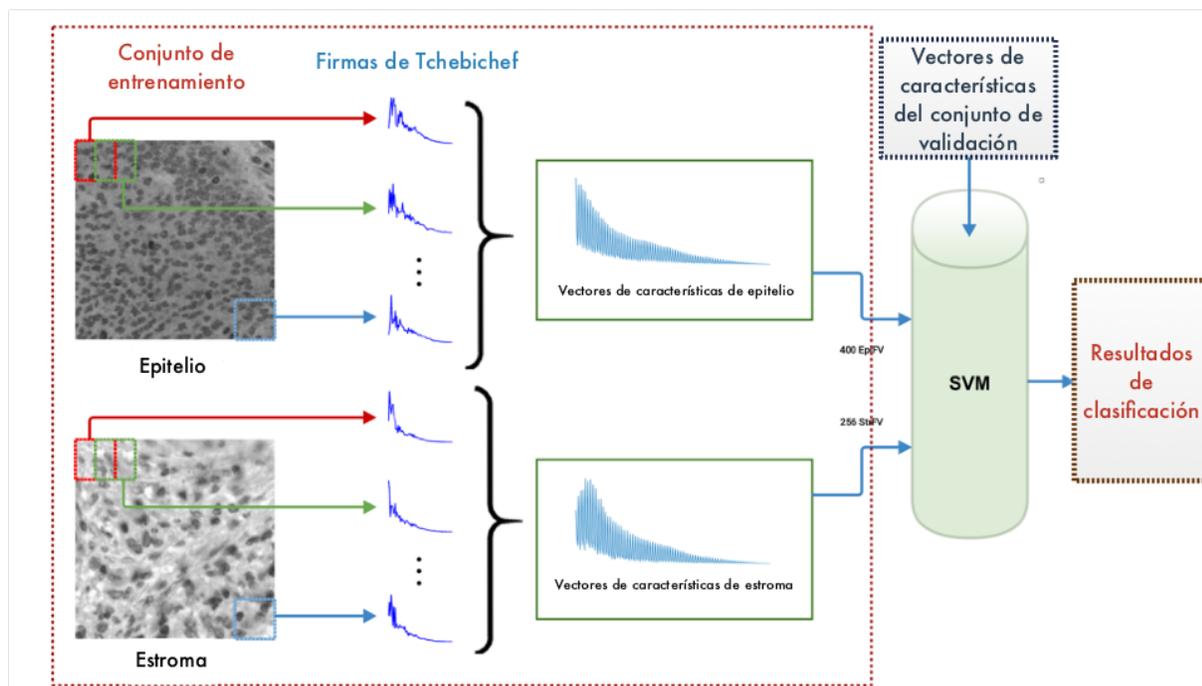


Figura 5.23: Esquema general de la clasificación del epitelio y estroma del cáncer colorrectal.

2. La propuesta evaluada en esta base de datos es la clasificación por Firmas de Textura Estadísticas, descrita en la Sección 3.3.1. El tamaño de la ventana se define entre  $10 \times 10$  y  $80 \times 80$  píxeles, con pasos de  $10 \times 10$  píxeles. Los mejores resultados de clasificación se obtuvieron con la ventana de análisis de  $40 \times 40$  píxeles. Finalmente, el vector característico calculado para cada imagen del conjunto de entrenamiento es dado por la ecuación 3.4.
3. El clasificador elegido es una máquina de soporte vectorial (SVM) por ser un clasificador definido inicialmente para problemas de 2 clases. En la Figura 5.23, observamos el esquema general para la clasificación de epitelio y estroma.

La propuesta clasifica correctamente 700 de 725 imágenes, que significa una exactitud de 96.53 %. También se realizaron experimentos utilizando un clasificador  $k$  vecinos más cercanos, los mejores resultados con este clasificador se obtuvieron con  $k = 11$ . Por motivos de comparación, se implementó el descriptor de Patrones Binarios Locales ( $LBP$ ) [88], mencionados en el Apéndice A. Para cada imagen, en cada una de las ventanas, se construye el vector de características a partir de la concatenación de los histogramas  $LBP_{8,1}$  y del  $LBP_{16,2}$ . Entonces, todos los vectores  $LBP$  correspondientes a una imagen son agrupados y caracterizados por su promedio y desviación estándar. Además, se incluyen los resultados reportados por Linder *et al.* [130], ellos proponen un descriptor que combina  $LBP$  con información de contraste ( $LBP/C$ ) para caracterizar la textura de los tejidos. En este caso, el contraste se define a través del concepto de variación local invariante a la rotación ( $VAR_{P,R}$ ) [88], mencionado en el Apéndice A.1.

Método	Exactitud	Sensibilidad	Especificidad
DTMs/SVM	<b>96.47</b>	97.62	<b>95</b>
DTMs/KNN	94.12	94.79	91.61
LBP/SVM	90.35	89.1	85.81
LBP/KNN	91.53	83.48	85.83
LBP/C	95.53	<b>99.02</b>	93.87

**Tabla 5.10:** Comparación de los resultados de clasificación para epitelio y estroma.

Finalmente, el vector de características de textura utilizado por Linder *et al.* es:

$$LBP/C = LBP_{8,1} + VAR_{8,1} + LBP_{16,2} + VAR_{16,2}. \quad (5.3)$$

Además de la exactitud, definida en la ecuación 5.2, se utilizan otras dos métricas para la evaluación de los resultados: la sensibilidad y la especificidad. La sensibilidad es la efectividad promedio del clasificador por cada clase para identificar las etiquetas de clase verdaderas y se define como:

$$Sensibilidad = \frac{1}{C} \sum_{i=1}^C \frac{VP_i}{VP_i + FN_i}, \quad (5.4)$$

donde  $VP_i$  es el número de aciertos para la clase  $i$  y  $FN_i$  es el número de errores resultado de asignar la clase  $i$  cuando en verdad correspondía otra clase.

Mientras que, la especificidad es la efectividad promedio del clasificador por cada clase para identificar las etiquetas de clase falsas y tiene la siguiente definición:

$$Especificidad = \frac{1}{C} \sum_{i=1}^C \frac{VN_i}{VN_i + FP_i}, \quad (5.5)$$

donde  $VN_i$  es el número de aciertos para las clases que son diferentes a  $i$  y  $FP_i$  es el número de errores resultado de asignar cualquier otra clase diferente de  $i$ , cuando en verdad correspondía la clase  $i$ .

Todos los resultados de clasificación son mostrados en la Tabla 5.10 y hacen evidente que la propuesta aquí desarrollada es la mejor. Los mejores resultados por métrica están resaltados en negritas.

#### 5.4.2 CLASIFICACIÓN DE ENFERMEDADES HEMATOLÓGICAS

A través del tiempo, el conjunto de bases de datos IICBU-2008 ha sido utilizado para realizar pruebas de algoritmos de descripción, clasificación y validación; y no necesariamente sobre todo el

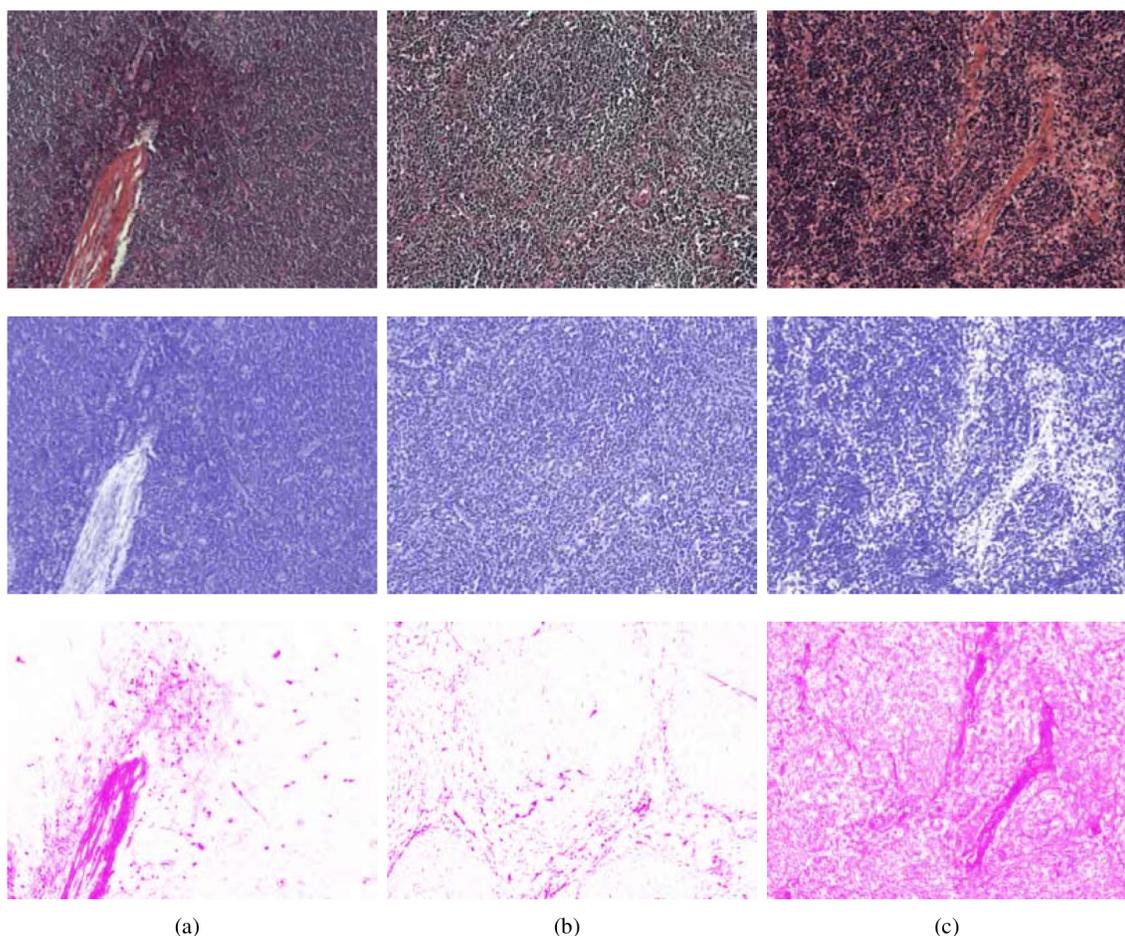
conjunto de bases, si no que se han reportado trabajos en los cuales sólo ocupan una de las bases de datos [120–122, 131, 132].

Las enfermedades hematológicas son un conjunto de enfermedades que comprenden diversos tipos de leucemias y linfomas. A pesar de que estas enfermedades son de poca incidencia, han sido objeto de estudio en años recientes. Por tal razón, se ha considerado que la base de datos de *Lymphoma* del IICBU-2008 tiene una repercusión de mayor interés. Esta base contiene tres tipos de enfermedades hematológicas teñidas con hematoxilina/eosina (H+E). La base de datos está compuesta por 133 muestras de **Leucemia Linfocítica Crónica (CLL)**, 139 de **Linfoma Folicular (FL)** y 122 de **Linfoma de Células de Manto (MCL)**. Las imágenes están en el espacio de color RGB y no contienen información alguna sobre los pacientes. Las enfermedades tienen las siguientes características:

- **Leucemia Linfocítica Crónica:** es el tipo de leucemia más común en el mundo occidental. Cerca del 80 % de los casos son diagnosticados en fase terminal. La CLL es un tipo de leucemia de lento desarrollo que ataca a los linfocitos B, que son responsables de la producción de inmunoglobulinas. Los linfocitos B dañados tienen una vida prolongada y reaccionan a la proteína CD38, de tal manera que, si las células contienen gran cantidad de esta proteína, la CLL tiende a empeorar rápidamente.
- **Linfoma Folicular:** el linfoma no Hodgkin (NHL) es el tipo de cáncer más común del sistema linfático porque representa alrededor del 90 % de los casos diagnosticados. Entonces, el FL es un tipo común de linfoma porque representa el 25 % de los casos de NHL. Es asintomático y afecta principalmente a los nodos linfáticos. El FL progresa lentamente desde los linfocitos B y eventualmente afecta a la médula ósea. Su nombre proviene del parecido de la distribución de éstas células anormales con respecto a la distribución folicular.
- **Linfoma de Células de Manto:** es una enfermedad hematológica que representa el 6 % de los NHL y usualmente afecta a hombres mayores de 50 años. Se desarrolla en la zona del manto, que es la parte externa de los nodos linfáticos. El MCL se propaga a la médula espinal y a veces hacia el tracto gastrointestinal. En comparación con CLL y FL, el MCL es una enfermedad agresiva que reacciona ante la proteína ciclina-D1, que ayuda en el control del crecimiento celular.

Ya que los fenotipos citopatológicos de las enfermedades hematológicas pueden diferir significativamente, es muy importante detectar patrones específicos para proponer terapias acordes al perfil de cada paciente. La biopsia es la práctica estándar para la adquisición, detección y evaluación de la enfermedad. Sin embargo, los patólogos deben manipular un gran número de casos al día mediante una evaluación visual, que representa una dificultad importante y mucho tiempo invertido en esta tarea.

Para evitar los problemas que estas situaciones suponen, se han propuesto diversos métodos automáticos de clasificación. Por ejemplo, Orlov *et al.* [132] clasifica imágenes de *Lymphoma* usando un conjunto de características conformadas por descriptores polinomiales, de textura, espectrales y espaciales. Sertel *et al.* [133] modifican la matriz de coocurrencia de nivel de gris de las imágenes de citopatología usando una cuantización de color no lineal. Además, Tuzel *et al.* [134] usan *Textons* y



**Figura 5.24:** Base de datos *Lymphoma*. Las imágenes fueron transformadas al espacio de color Hematoxilina y Eosina (H+E). La primera fila muestra las imágenes RGB originales, la segunda fila corresponde al canal Hematoxilina y la tercera fila al canal Eosina. **(a)** CLL, **(b)** FL y **(c)** MCL.

máquinas de soporte vectorial para clasificar cinco clases de enfermedades hematológicas. En [117], los autores usan WND-CHARM para clasificar esta base de datos.

El método propuesto de clasificación incluye una etapa de preprocesamiento de las imágenes más exhaustiva y su procedimiento es el siguiente:

1. Transformar el espacio de color de las imágenes de RGB al H+E, para tener canales que representen el efecto del tinte aplicado a través de una deconvolución de color [135]. Entonces, las imágenes estarán en un espacio acorde al problema de interés. En la Figura 5.24 se muestran ejemplos de cada clase y el resultado después de aplicar la deconvolución.
2. Calcular los momentos de las imágenes ya transformadas al espacio H+E, la ventana de traslape será de  $50 \times 50$  píxeles porque con ese tamaño se obtuvo el mejor resultado con las firmas estadísticas de textura.
3. El descriptor estará conformado por los promedios de cada familia de momentos (Tchebichef, Krawtchouk y Dual Hahn), es decir, no habrá desviaciones estándar ni curtosis como en la propuesta de la Sección 3.3.1.

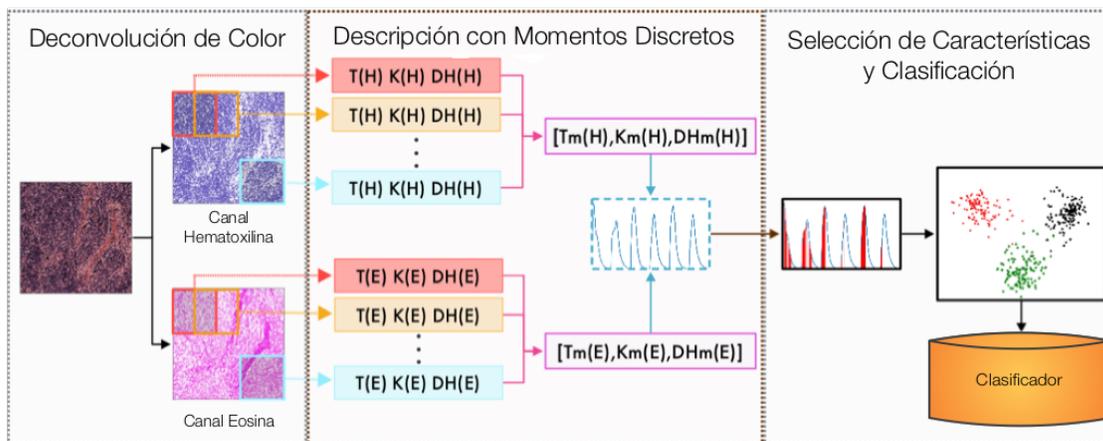


Figura 5.25: Diagrama general para la clasificación de la base de datos Lymphoma.

Método	Exactitud	Sensibilidad	Especificidad
Propuesta	<b>94 ± 4</b>	<b>94.31 ± 2.13</b>	<b>97 ± 1.21</b>
Tchebichef	82.06 ± 6.06	82.90 ± 6.26	91.38 ± 2.94
Krawtchouk	75.97 ± 5.32	76.63 ± 5.52	88.30 ± 2.63
Dual Hahn	70.88 ± 8.86	71.24 ± 8.52	85.70 ± 4.21
LBP	76.77 ± 6.07	76.01 ± 5.99	88.35 ± 2.89

Tabla 5.11: Resultados de clasificación de la base Lymphoma con diferentes descriptores.

- El vector de características es muy largo, por lo tanto, debe ser optimizado para obtener el mejor resultado posible con el vector de características más corto. Para esto, se utiliza la función *Relieff* (Sección 4.1.2) que calcula el peso y la relevancia de cada característica al hacer una pre-clasificación con  $KN$  vecinos.
- Utilizar el Kernel Fisher (Sección 4.2.4) en los datos de entrada para proyectarlos en un espacio en el que se encuentren mejor separados entre sí.
- Finalmente, se realiza una validación cruzada con 10 grupos con los clasificadores  $K$  vecinos más cercanos y *Random Forest*.

El mejor resultado de exactitud es  $94 \pm 4\%$  con un vector de 115 características ordenadas por importancia utilizando  $KN$  vecinos. Finalmente, un esquema del procedimiento se muestra en la Figura 5.25. Además, la base de datos fue caracterizada con los (LBPs) de radio  $R = 1$  y  $P = 8$  vecinos usando la misma metodología descrita anteriormente. Los resultados son mejores que los presentados por Sertel *et al.* [133] quienes reportan una exactitud de 90.3%. Los resultados son mostrados en la Tabla 5.11.

# 6

## Conclusiones y Trabajo Futuro

### 6.1 Conclusiones

En el desarrollo de este trabajo de investigación hemos presentado nuevos métodos para la descripción de texturas basados en Momentos Ortogonales Discretos con el fin de obtener la clasificación de diferentes bases de datos de imágenes de interés médico. En la mayoría de las pruebas se observa que estos métodos son lo suficientemente generales para aplicarse a distintos tipos de datos porque el preprocesamiento aplicado a estas bases de datos es muy básico, ya que prácticamente consiste en transformar las imágenes a color en imágenes en escala de grises y, en algunos casos, reducir el tamaño de las imágenes para acelerar el proceso de descripción y clasificación. Esto nos lleva a pensar que los métodos propuestos son robustos y generalizables, muy probablemente, a cualquier otro tipo de bases de datos, no necesariamente pertenecientes al ámbito biomédico.

Otro punto importante de la propuesta es la descripción a través de las ventanas traslapadas. Las ventanas traslapadas nos resultan útiles por dos situaciones: la inestabilidad de la implementación de los polinomios de Shmaliy, Krawtchouk y Dual Hahn, y la diferencia de tamaño entre las imágenes dentro de una misma base de datos. La inestabilidad en las oscilaciones aparece cuando se calculan polinomios de orden alto porque involucra funciones factoriales que crecen muy rápidamente, culminando en un desbordamiento en el cálculo numérico. De manera que, analizar imágenes en ventanas traslapadas de menor tamaño, se vuelve necesario si se desea mantener la estabilidad del cálculo de los momentos. Por otra parte, si se calculasen momentos sobre las imágenes completas y si estas imágenes, dentro de la misma base de datos, tuviesen un tamaño distinto entre sí, entonces, tendríamos vectores de características de distintos tamaños y tendríamos que definir otro criterio de tal manera que nos permita

utilizar estos vectores en particular.

Desde su reciente presentación, los momentos Shmaliy no han sido tratados o probados en el análisis de texturas de imágenes. En este trabajo de investigación, mostramos por primera vez la capacidad de los momentos Shmaliy para calcular las características de la textura y compararlas con las familias más populares de momentos ortogonales discretos: los momentos de Tchebichef, Krawtchouk y Dual Hahn. Con base en los resultados de clasificación, la descripción de textura basada en los momentos y polinomios discretos de Shmaliy tiene un rendimiento similar al que muestran las familias de Tchebichef y Krawtchouk, y es mejor que el de la familia dual Hahn.

Por otro lado, la comparación entre momentos discretos ortogonales y otras técnicas encontradas en la literatura muestra que los resultados obtenidos son mejores. Además, los momentos ortogonales discretos mostraron la misma capacidad entre sí para la descripción de texturas, por lo tanto, su consideración es pertinente para el análisis de textura y trabajo futuro.

### 6.2 Trabajo Futuro

En todos los experimentos se han utilizado diferentes tamaños de ventanas para la descripción de las bases de datos. Sin embargo, no se tiene certeza sobre el tamaño óptimo de ventana. Suponemos que este tamaño dependerá directamente de las texturas de los tejidos a analizar. Una primera idea es encontrar el elemento fundamental de textura de un tejido en particular, y partiendo de este elemento, establecer el tamaño de la ventana de análisis. Sin embargo, esta situación también generaría ciertas complicaciones, ya que, si estos elementos fundamentales varían de tamaño entre clases dentro de una base de datos, entonces habría que definir un criterio para homogeneizar el tamaño, en primera instancia, para que los vectores de características por clase no sean de diferentes tamaños.

En algunas de las propuestas se combinan las características producidas por cada base polinomial al agruparlas en un vector de características, generando así, descripciones sobrecompletas con información sobre las variaciones locales de los tejidos analizados. Además, la selección de características toma ventaja de esta sobre descripción para preservar las características más relevantes, optimizando así la caracterización de los tejidos o las clases. También, existen métodos para medir la redundancia de los datos, por lo que sería interesante saber qué tan redundante es la descripción a través de la combinación de los coeficientes de los momentos ortogonales.

La mayoría de los esquemas propuestos propuestos están diseñados para clasificar imágenes que en su totalidad comprenden a una sola clase, por lo tanto, para tareas de clasificación de diversas clases dentro de una imagen (segmentación), estos esquemas no son óptimos. Para atacar este problema, en este trabajo se propuso un esquema que combina polinomios ortogonales discretos con *Textons*. Por ahora, los resultados no están al nivel de los experimentos basados en firmas de textura, sin embargo, representan los primeros pasos para poder tener un método de descripción y segmentación con momentos ortogonales discretos.



## Patrones Binarios Locales (LBPs)

El operador básico *LBP*, introducido por Ojala *et al.* [11], se basaba en la suposición de que la textura tiene dos aspectos complementarios a nivel local: un patrón y su fuerza. En ese trabajo, el *LBP* fue propuesto como una versión de dos niveles de la unidad de textura [136] para describir los patrones de textura locales.

La versión original del operador patrón *LBP* trabaja en un bloque de  $3 \times 3$  píxeles de una imagen. Los píxeles en este bloque son *umbralizados* con respecto al valor de pixel central, multiplicados por potencias de dos y luego son sumados para obtener una etiqueta correspondiente al pixel central. Como la vecindad se compone de 8 píxeles, un total de  $2^8 = 256$  etiquetas diferentes se pueden obtener en función de los valores relativos entre el pixel central y sus vecinos. En la Figura A.1 se muestra el operador básico *LBP*.

Varios años después de la publicación original, el operador *LBP* es presentado en una forma más genérica revisada por Ojala *et al.* en [88]. En contraste con el *LBP* básico que usa 8 píxeles en un

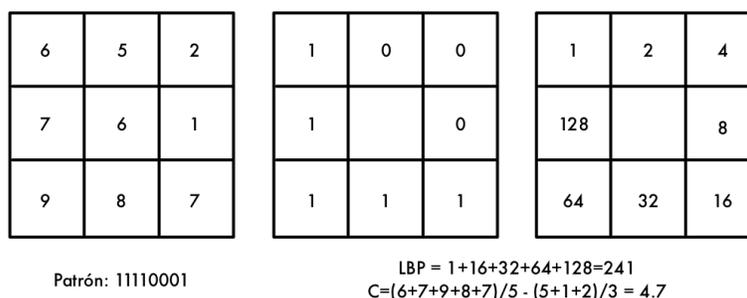
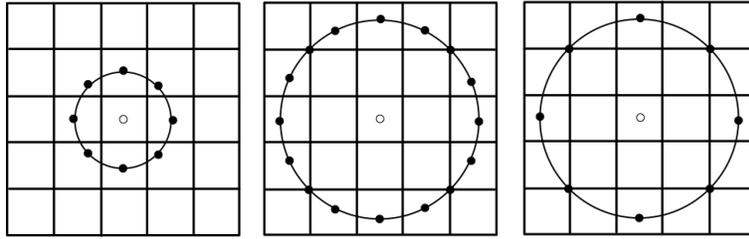


Figura A.1: El *LBP* Original.

bloque de  $3 \times 3$ , la formulación genérica no pone ninguna limitación para el tamaño de la vecindad o para el número de puntos de muestreo.



**Figura A.2:** Las vecindades de  $(8,1)$ ,  $(16,2)$  y  $(8,2)$ . Los valores de pixel son interpolados bilinealmente cuando el punto de muestreo no está justo en el centro del pixel.

Considerando una imagen en escala de grises  $I(x, y)$  y  $g_c$  es el nivel de gris de un pixel en cualquier posición  $((x, y)$ , i.e.  $g_c = I(x, y)$ ).

Por otra parte,  $g_p$  denota el valor de gris de un punto en una zona circular uniformemente espaciada de puntos de muestreo  $P$  y radio  $R$  alrededor del punto  $(x, y)$ :

$$g_p = I(x_p, y_p), \quad (\text{A.1})$$

donde  $p = 0, \dots, P - 1$  y:

$$x_p = x + R \cos(2\pi p/P), \quad (\text{A.2})$$

$$y_p = y - R \sin(2\pi p/P). \quad (\text{A.3})$$

En la Figura A.2 se muestran ejemplos de las vecindades circulares.

Suponiendo que la textura local de la imagen  $I(x, y)$  se caracteriza por la distribución conjunta de los valores de gris de los pixeles  $P + 1$  ( $P > 0$ ):

$$T = t(g_c, g_0, g_1, \dots, g_{P-1}). \quad (\text{A.4})$$

Sin pérdida de información, el valor del pixel central se puede restar con respecto a la vecindad:

$$T = t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c). \quad (\text{A.5})$$

En el siguiente paso, la distribución conjunta se aproxima suponiendo que el pixel central es estadísticamente independiente de las diferencias, lo que permite la factorización de la distribución:

$$T \approx t(g_c)l(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c). \quad (\text{A.6})$$

Ahora el primer factor  $t(g_c)$  es la distribución de intensidad sobre  $I(x, y)$ . Desde el punto de vista del análisis de patrones de textura locales, no contiene información útil. En cambio, la distribución

conjunta de las diferencias

$$t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (\text{A.7})$$

puede ser utilizada para modelar la textura local. Sin embargo, la estimación fiable de esta distribución multidimensional a partir de los datos de la imagen puede ser difícil. Una solución a este problema, propuesta por Ojala *et al.* en [137], es aplicar cuantificación vectorial y se creó un libro de códigos de 384 palabras para reducir la dimensionalidad del espacio de características. Sin embargo, este enfoque todavía tenía ciertas debilidades: en primer lugar, las diferencias  $g_p - g_c$  son invariantes a los cambios del valor promedio de intensidad de la imagen, pero no a otro tipo de cambios en los niveles de intensidad y, en segundo lugar, con el fin de utilizarlo para la clasificación de la textura, requiere de un entrenamiento. Para superar estos problemas, sólo se consideran los signos de las diferencias:

$$t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)), \quad (\text{A.8})$$

donde  $s(z)$  es la función escalón definida por:

$$s(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (\text{A.9})$$

El operador genérico del *LBP* se deriva de esta distribución conjunta. Al igual que en el caso del *LBP* básico, este operador se obtiene mediante la suma de las diferencias umbralizadas por la ecuación (A.9) ponderadas por potencias de dos. El operador  $LBP_{P,R}$  se define como

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p. \quad (\text{A.10})$$

En la práctica, la ecuación (A.10) significa que los signos de las diferencias en una vecindad son interpretados como un número binario de  $P$  bits, lo que resulta en  $2^P$  valores distintos para el código *LBP*.

El *LBP* original (Figura A.1) es muy similar a  $LBP_{8,1}$ , con dos diferencias: en primer lugar, la vecindad en la definición general se indexa circularmente, por lo que es más fácil obtener descriptores de textura invariantes a la rotación y, en segundo lugar, los píxeles ubicados en las diagonales en la ventana de  $3 \times 3$  se interpolan en  $LBP_{8,1}$ .

## A.1 Contraste invariante a la rotación

El *LBP* es una métrica excelente para medir patrones espaciales, pero, por definición, descarta el contraste. Si la invarianza a la escala de grises no es requerida y se quiere incorporar información local de contraste, entonces, Ojala *et al.* [88] propusieron el concepto de variación local invariante a la

rotación ( $VAR_{P,R}$ ):

$$VAR_{R,P} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad (\text{A.11})$$

donde:

$$\mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p. \quad (\text{A.12})$$

$VAR_{R,P}$  es invariante ante los desplazamientos en escala de gris. Ya que el  $LBP_{P,R}$  y  $VAR_{P,R}$  son complementarios, su concatenación se convierte en una métrica invariante a la rotación de la textura local de la imagen.

# B

## Clasificadores

En este apéndice se detallan algunos aspectos que concierne a los clasificadores utilizados en los experimentos de este trabajo de tesis. Los algoritmos son:  $K$  vecinos más cercanos ( $k$ NN), Máquinas de Soporte Vectorial (SVM) y *Random Forest* (RF).

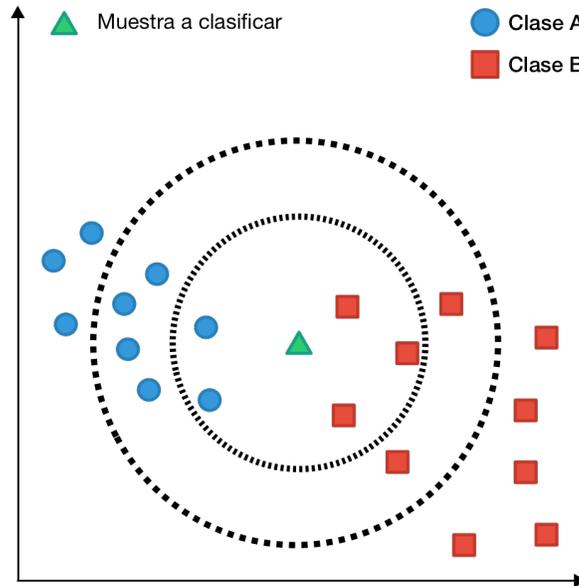
### B.1 $K$ vecinos más cercanos

La idea principal sobre la que se fundamenta este método de clasificación es que un nuevo caso se clasificará en la clase más frecuente a la que pertenecen sus  $k$  vecinos más cercanos.

$K$  vecinos más cercanos ( $k$ NN, por sus siglas en inglés) es un método de clasificación supervisada, es decir, se necesita un conocimiento *a priori* obtenido a través de un conjunto de entrenamiento con el fin de clasificar un conjunto nuevo (de validación) a través de las características obtenidas en el entrenamiento. Este es un método de clasificación no paramétrico que estima el valor de la función de probabilidad  $F(x, C_j)$  de que un elemento  $x$  pertenezca a la clase  $C_j$  a partir de la información obtenida en el conjunto de entrenamiento.

#### B.1.1 DESARROLLO

Los vectores de entrenamiento se definen en un espacio característico multidimensional  $X$ , y cada uno está descrito en términos de  $p$  características considerando  $q$  clases para la clasificación. Los valores



**Figura B.1:** Ejemplo del algoritmo de  $K$  vecinos. El triángulo verde es la muestra que se debe clasificar. Para  $k = 5$ , círculo interno, el triángulo será clasificado como cuadrado (Clase B). Mientras que, si  $k = 11$ , círculo externo, la muestra será clasificada como círculo o clase A

de las características del  $i$ -ésimo vector de entrenamiento  $x_i$  se define como:

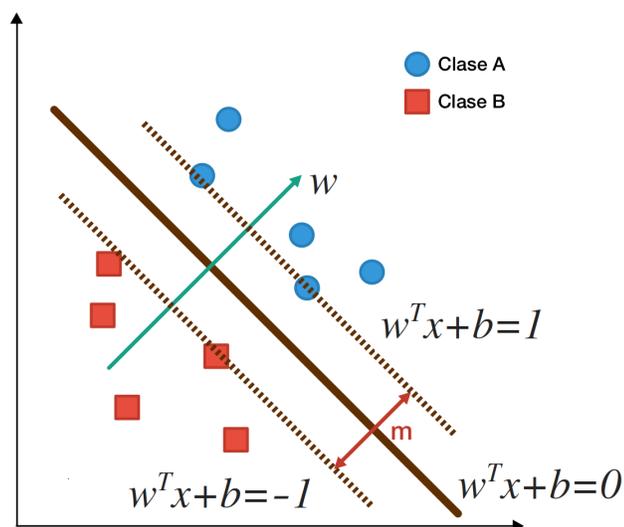
$$x_i = (x_{1i}, x_{2i}, \dots, x_{pi}) \in X, \tag{B.1}$$

donde  $1 \leq i \leq n$  y  $n$  es el número total de vectores de características del conjunto de entrenamiento.

Para cada vector de características del conjunto de validación se calculan las distancias, generalmente euclidiana, respecto a los vectores de entrenamiento y se eligen los  $k$  vectores más cercanos. Entonces, el vector del conjunto de validación es asignado a la clase que más se repite de los  $k$  vectores más cercanos.

$k$ NN es un esquema de clasificación atípico si se compara con el resto de paradigmas clasificatorios, ya que en la mayoría de estos paradigmas se necesita la *deducción* del modelo de clasificación y su *aplicación* en el conjunto de evaluación, mientras que en  $k$ NN, al no existir un modelo explícito, la *deducción* y la *aplicación* se realizan al mismo tiempo y se le denomina *transducción*.

Otro tema que se debe mencionar es la elección de  $k$  porque depende fundamentalmente de los datos. Generalmente, valores grandes de  $k$  reducen el efecto de ruido en la clasificación, pero crean límites entre clases parecidas. Aprovechando que el algoritmo es computacionalmente rápido, la elección de  $k$  se puede realizar empíricamente a través de su variación.



**Figura B.2:** El margen máximo  $m$  es la máxima distancia que hay entre los vectores de soporte más cercanos entre clases.

## B.2 Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial (SVM, por sus siglas en inglés *Support Vector Machines*) están basadas en la teoría del aprendizaje estadístico y fueron introducidas por Cortes y Vapnik [138]. Aunque las SVM fueron ideadas originalmente para resolver problemas de clasificación binaria, actualmente se usan para resolver problemas de clasificación multi clase, regresión y agrupamiento (*clustering*). La teoría de la SVM esta basada en la idea de minimización de riesgo estructural [139]. En muchas aplicaciones, las SVM han mostrado tener gran desempeño, más que las máquinas de aprendizaje tradicional como las redes neuronales [140] y han sido introducidas como herramientas poderosas para resolver problemas de clasificación.

Una SVM primero mapea los puntos de entrada a un espacio de características de una dimensión mayor ( por ejemplo: si los puntos de entrada están en  $\mathbb{R}^2$  entonces son mapeados por la SVM a  $\mathbb{R}^3$  ) y encuentra un hiperplano que los separe y maximice el margen  $m$  entre las clases en este espacio extra, como se muestra en la Figura B.2.

Maximizar el margen  $m$  es un problema de programación cuadrática (QP) y puede ser resuelto a través de multiplicadores de Lagrange. Sin ningún conocimiento sobre el mapeo de las muestras, la SVM encuentra el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características que son llamadas *kernels*. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada que son llamados vectores de soporte.

B.2.1 MÁQUINA DE SOPORTE VECTORIAL PARA MUESTRAS LINEALMENTE SEPARABLES

Dado el conjunto de muestras  $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ , donde  $x_i \in \mathbb{R}^d$  y  $y_i \in \{+1, -1\}$ , para  $i = 1, \dots, n$ , con la distribución que se muestra en la Figura 2. En la mayoría de los casos, la búsqueda de un hiperplano adecuado en un espacio de entrada es demasiado restrictivo para ser de uso práctico. Una solución a esta situación es mapear el espacio de entrada en un espacio de características de una dimensión mayor y buscar el hiperplano óptimo en ese espacio. Sea  $z = \phi(x)$  el vector de características en el espacio de características con un mapeo  $\phi$  de  $\mathfrak{X}^N$  a un espacio de características  $Z$ . Entonces, se pretende encontrar el hiperplano:

$$w \cdot z + b = 0. \tag{B.2}$$

Las variables  $w$  y  $b$  son definidas de tal manera que el punto  $x_i$  sea clasificado de acuerdo a la función:

$$f(x_i) = \text{sign}(w \cdot z_i + b) = \begin{cases} 1 & y_i = 0 \\ -1 & y_i = -1, \end{cases} \tag{B.3}$$

donde  $w \in Z$  y  $b \in \mathfrak{R}$ . Entonces, el conjunto  $S$  se dice que es linealmente separable si existen  $w$  y  $b$  de tal manera que sean válidas las inecuaciones para todo el conjunto  $S$ :

$$\begin{cases} (w \cdot z_i + b) \geq 1, & y_i = 1 \\ (w \cdot z_i + b) \leq -1, & y_i = -1. \end{cases} \tag{B.4}$$

Tal y como se observa en la Figura B.3(a), el hiperplano de separación no es único, es decir, existen infinitos hiperplanos separables que cumplen las condiciones de restricción dadas en la ecuación B.4. Entonces, para el caso linealmente separable de  $S$ , podemos encontrar un único hiperplano óptimo, para el cual, el margen entre las proyecciones de los puntos de entrenamiento de dos diferentes clases es maximizado.

B.2.2 MÁQUINA DE SOPORTE VECTORIAL PARA MUESTRAS NO LINEALMENTE SEPARABLES

Para tratar con datos que no son linealmente separables, como se observa en la Figura B.3(b), el análisis previo puede ser generalizado introduciendo algunas variables no-negativas  $\xi \geq 0$ , de tal modo que la ecuación B.4 se reescribe de la siguiente manera:

$$y_i(w \cdot z_i + b) \geq 1 - \xi_i. \tag{B.5}$$

Los valores  $\xi \neq 0$  en la ecuación B.5 son aquellos para los cuales el punto  $x_i$  no satisface a la ecuación B.4. Entonces el término  $\sum_{i=1}^n \xi_i$  se toma como una medida de error en la clasificación.

El problema del hiperplano óptimo es entonces redefinido como la solución al problema de minimi-

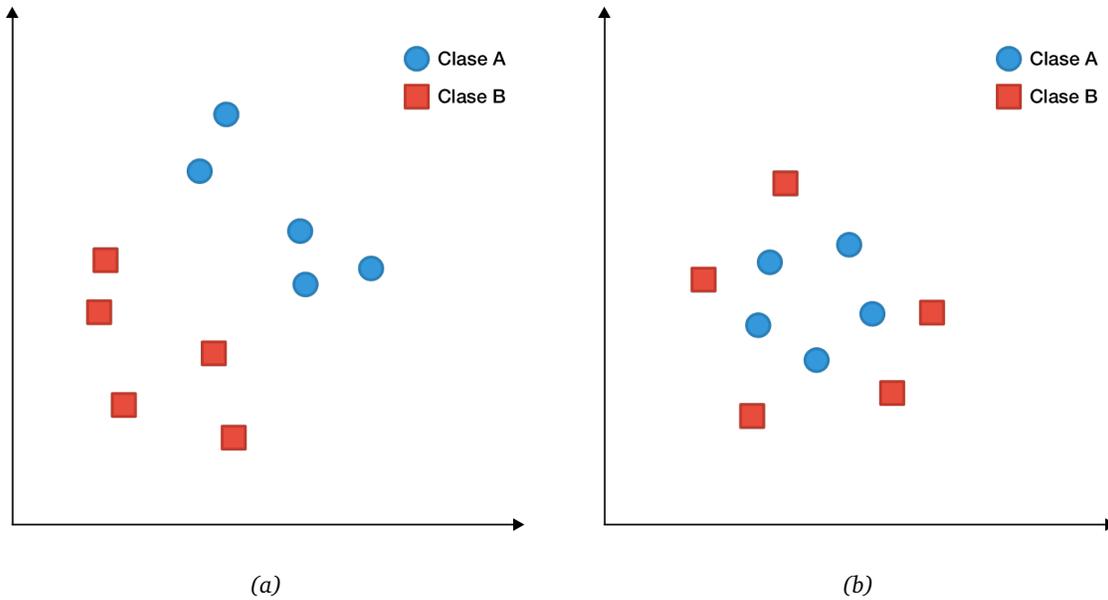


Figura B.3: (a) Muestras linealmente separables y (b) Muestras no linealmente separables.

zación:

$$\begin{aligned} \min & \left\{ \frac{1}{2} w \cdot w + C \sum_{i=1}^n \xi_i \right\} & (B.6) \\ \text{s.a.} & y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, n \\ & \xi_i \geq 0, \quad i = 1, \dots, n, \end{aligned}$$

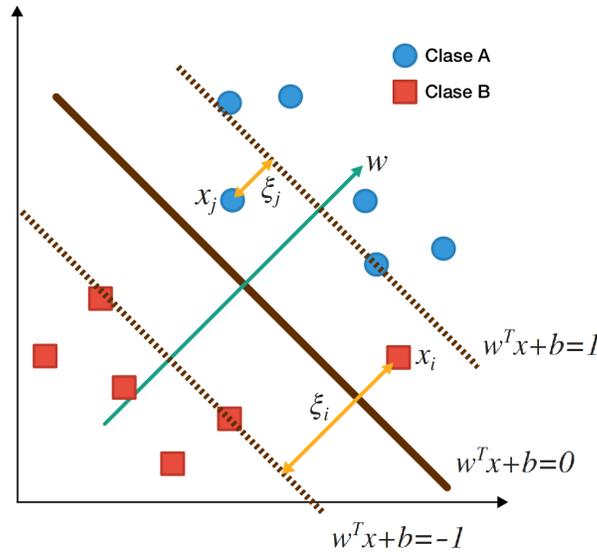
donde  $C$  es una constante. El parámetro  $C$  puede ser definido como un parámetro de regularización. Este es el único parámetro libre de ser ajustado en la formulación de la SVM. El ajuste de éste parámetro puede hacer un balance entre la maximización del margen y el error en la clasificación.

Buscando el hiperplano óptimo en la ecuación B.2.2 es un problema QP, que puede ser resuelto construyendo un Lagrangiano y transformándolo en el dual

$$\begin{aligned} \max & W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (z_i \cdot z_j) & (B.7) \\ \text{s.a.} & \sum_{i=1}^n y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1, \dots, n, \end{aligned}$$

donde  $\alpha = (\alpha_1, \dots, \alpha_n)$  es un vector de multiplicadores de Lagrange positivos asociados con las constantes en la ecuación B.5.

El teorema de Khun-Tucker juega un papel importante en la teoría de las SVM. De acuerdo a este



**Figura B.4:** Los puntos  $x_i$  y  $x_j$  son errores o outliers y se encuentran a una distancia  $\xi_i$  y  $\xi_j$ , respectivamente, que se interpreta como la variable de error o de holgura.

teorema, la solución  $\bar{\alpha}_i$  del problema, ecuación B.2.2, satisface:

$$\bar{\alpha}_i(y_i(\bar{w} \cdot z_i + b) - 1 + \bar{\xi}_i) = 0, \quad i = 1, \dots, n \quad (\text{B.8})$$

$$(C - \bar{\alpha}_i)\xi_i = 0, \quad i = 1, \dots, n. \quad (\text{B.9})$$

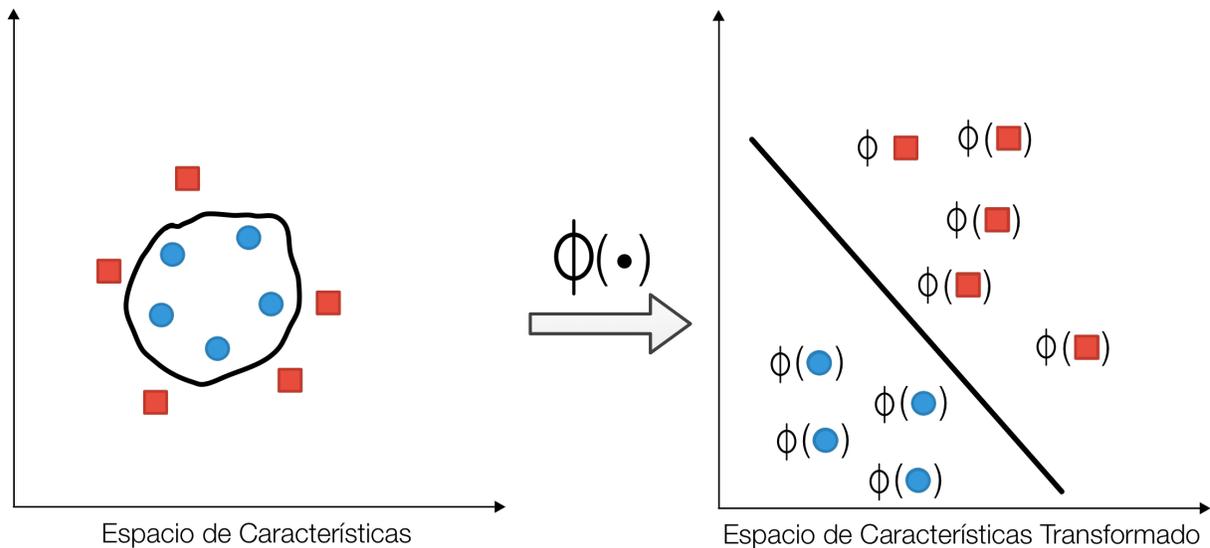
De esta igualdad se deduce que los únicos valores  $\bar{\alpha}_i \neq 0$  son aquellos que para las constantes en la ecuación B.5 son satisfechas con el signo de igualdad. El punto  $x_i$  correspondiente con  $\bar{\alpha}_i > 0$  es llamado *vector de soporte*. Pero hay dos tipos de vectores de soporte en un caso no separable. En el caso  $0 < \bar{\alpha}_i < C$ , el vector de soporte  $x_i$  satisface las igualdades  $y_i(w \cdot z_i + b) = 1$  y  $\xi_i = 0$ . En el caso  $\bar{\alpha}_i = C$ , el correspondiente  $\xi_i$  es diferente de cero y el correspondiente vector de soporte  $x_i$  no satisface la ecuación B.4, por lo tanto, estos vectores de soporte son clasificados erróneamente. El punto  $x_i$  con  $\bar{\alpha}_i = 0$  es clasificado correctamente y esta claramente alejado del margen de decisión. Estas condiciones se muestran en la Figura B.4.

Para construir el hiperplano óptimo  $\bar{w} \cdot z + \bar{b}$ , se utiliza:

$$\bar{w} = \sum_{i=1}^n \bar{\alpha}_i y_i z_i, \quad (\text{B.10})$$

y el escalar  $b$  puede ser determinado de las condiciones de Kuhn-Tucker, mostrados en al ecuación B.9. La función de decisión generalizada de las ecuaciones B.3 y B.10 es tal que:

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i (z_i \cdot z) + b\right) \quad (\text{B.11})$$



**Figura B.5:** Las muestras son proyectadas al espacio del kernel para facilitar su separabilidad y posterior clasificación con una SVM.

### B.2.3 PROYECCIÓN KERNEL PARA EL CASO NO LINEALMENTE SEPARABLE

Como no tenemos ningún conocimiento de  $\phi$ , el cálculo del problema mostrado en las ecuaciones B.2.2 y B.11 es imposible. Sin embargo, las SVM tienen la propiedad de que no es necesario un conocimiento previo sobre  $\phi$ , sino que se necesita una función  $K(\cdot, \cdot)$  llamada *kernel*, que permite calcular el producto punto de las muestras en el espacio de características  $Z$  de la siguiente manera:

$$z_i \cdot z_j = \phi(x_i) \cdot \phi(x_j) = K(x_i, x_j). \quad (\text{B.12})$$

Las funciones que satisfacen el teorema de Mercer pueden ser usadas como productos punto, y por ende pueden ser usadas como *kernels*. Algunas de las funciones *kernel* más utilizadas son:

- Kernel lineal:

$$K(x_i, x_j) = \langle x_i, x_j \rangle.$$

- Kernel polinomial:

$$K_p(x_i, x_j) = [\gamma \langle x_i, x_j \rangle + \tau]^p.$$

- Kernel Gaussiano:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0.$$

- Kernel sigmoidal:

$$K(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + \tau).$$

En síntesis, los puntos que se encuentran en el espacio de características original son proyectados al espacio del *kernel* seleccionado, el resultado es un nuevo espacio de características transformado

en el cual, los nuevos datos transformados son ahora linealmente separables, idealmente. Es evidente que, para lograr una separación óptima entre los nuevos puntos, es importante ajustar los parámetros del *kernel*:  $\gamma$ ,  $\tau$  y  $p$ . Un ejemplo del resultado de la proyección de los datos en el espacio del *kernel* se muestra en la Figura B.5. Finalmente, el hiperplano no lineal de separación puede ser encontrado como la solución de:

$$\begin{aligned} \max \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.a.} \quad \sum_{i=1}^n y_i \alpha_i &= 0, \quad 0 \leq \alpha_i \leq C, i = 1, \dots, n, \end{aligned} \quad (\text{B.13})$$

y la función de decisión es:

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign} \left( \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \right). \quad (\text{B.14})$$

### B.3 *Random Forest*

*Random Forest* (bosque aleatorio, en español) es un método de aprendizaje de ensamblado para clasificación, regresión y otras tareas, que consiste en la combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Para tareas de regresión, el valor estimado es el promedio de los valores resultantes de los árboles predictores. Mientras que, para clasificación, la etiqueta es la moda estadística generada a partir de las respuestas de los árboles.

El primer algoritmo de *Random Forest* fue presentado por Ho [141], quien utilizó el método de subespacio aleatorio propuesto por Barandiaran [142], diferenciándose en que la formulación de Ho, es una forma de implementar el enfoque de *discriminación estocástica* propuesto por Kleinberg [143].

Una extensión de esta propuesta fue desarrollada por Leo Breiman [144] y fue llamada finalmente *Random Forest*, que por cierto es marca registrada<sup>1</sup>. El método combina la idea de *bagging* de Breiman y la selección aleatoria de atributos, introducida independientemente por Ho [141], Amit y Geman [145], para construir una colección de árboles de decisión con variación controlada.

Los árboles de decisión son un método popular para diversas tareas de aprendizaje de máquina porque son invariables a la escala y varias otras transformaciones de las características, es robusto a la inclusión de características irrelevantes y produce modelos inspeccionables. Sin embargo, rara vez son precisos [146].

En particular, los árboles que crecen muy profundamente tienden a aprender patrones muy irregulares porque se sobre ajustan con respecto a sus conjuntos de entrenamiento, es decir, tienen un sesgo bajo y una varianza alta. En cambio, *Random Forest* es una forma de promediar múltiples árboles de decisión profunda, entrenados con diferentes elementos del conjunto de entrenamiento, con el

<sup>1</sup>Número de marca registrada en los EE.UU. 3185828, registrado en 2006/12/19

objetivo de reducir la varianza[146]. Por otra parte, la reducción de la varianza viene acompañada con el aumento del sesgo y cierta pérdida en la interpretabilidad del método, pero, en general, aumenta en gran medida el rendimiento final del modelo.

### B.3.1 ESQUEMA GENERAL DE RANDOM FOREST

El algoritmo de entrenamiento de *Random Forest* aplica la técnica general conocida como *bagging*, o embolsado, a los árboles de aprendizaje. Dado el conjunto de entrenamiento  $X = x_1, \dots, x_n$  con sus respectivas respuestas o etiquetas de clase  $Y = y_1, \dots, y_n$ , se embolsa repetidamente ( $B$  veces) un conjunto aleatorio con reemplazo perteneciente al conjunto de entrenamiento original, normalmente  $2/3$  del conjunto original, y los árboles son entrenados con estas muestras:

Para  $b = 1, \dots, B$ :

1. Obtener el conjunto aleatorio de muestras con reemplazo  $X_b$  y  $Y_b$  a partir del conjunto de entrenamiento  $X$  y  $Y$ .
2. Entrenar el árbol de clasificación  $f_b$  con  $X_b$  y  $Y_b$ .

Después de entrenar, las predicciones de clase para las muestras del conjunto de entrenamiento  $x'$  se obtienen a partir de la moda estadística realizada sobre los resultados obtenidos de todos los árboles entrenados. Sin embargo, es importante mencionar que la implementación de *Random Forest* incluida en la paquetería *sci-kit learn* para *Python*, utilizada en este trabajo de investigación, no usa la moda estadística, en su lugar, calcula las probabilidades de que cierta clase  $c_i$  corresponda a la muestra  $x'_j$ :

$$P(c_i) = \frac{1}{B} \sum_j P(c_i|x'_j) \quad (\text{B.15})$$

Este procedimiento conduce a un mejor rendimiento del modelo porque disminuye la varianza del modelo, sin aumentar el sesgo. Esto significa que, aunque las predicciones de un solo árbol son altamente sensibles al ruido en su conjunto de entrenamiento, el promedio de muchos árboles no lo es, siempre que los árboles no estén correlacionados. Simplemente entrenar muchos árboles en un solo conjunto de entrenamiento daría árboles fuertemente correlacionados (o incluso el mismo árbol muchas veces, si el algoritmo de entrenamiento es determinista). El muestreo del conjunto inicial de entrenamiento (*bagging*) es una forma de descorrelacionar los árboles al inicializarlos con diferentes conjuntos de entrenamiento.

El procedimiento mencionado anteriormente describe el algoritmo original de *Tree Bagging*. *Random Forest* contiene un paso adicional: se utiliza un algoritmo modificado de árboles de aprendizaje que no solamente selecciona aleatoriamente muestras del conjunto de entrenamiento, sino también selecciona un subconjunto aleatorio de características. Este procedimiento también es conocido como *feature bagging*. La razón para hacer esto es que existe correlación entre los árboles en una muestra de *bagging*: si una o algunas características resultan ser predictores muy fuertes para la variable de respuesta

(regresión) o etiqueta objetivo (clasificación), estas características serán seleccionadas en muchos de los árboles  $B$ , provocando correlación entre ellos. Un análisis de cómo el *bagging* y la selección un subconjunto aleatorio de características contribuyen a los resultados de precisión en diferentes condiciones fue reportado por Ho [147]. En la Figura B.6 se muestra el procedimiento de entrenamiento del *Random Forest*.

Normalmente, para un problema de clasificación cuyas muestras tienen una longitud de  $p$  características, un número de  $\sqrt{p}$  características (con redondeo hacia abajo) para cada subconjunto de muestras aleatorias. Mientras que, para regresión, se sugiere utilizar  $p/3$  características [146].

## B.4 Naive Bayes

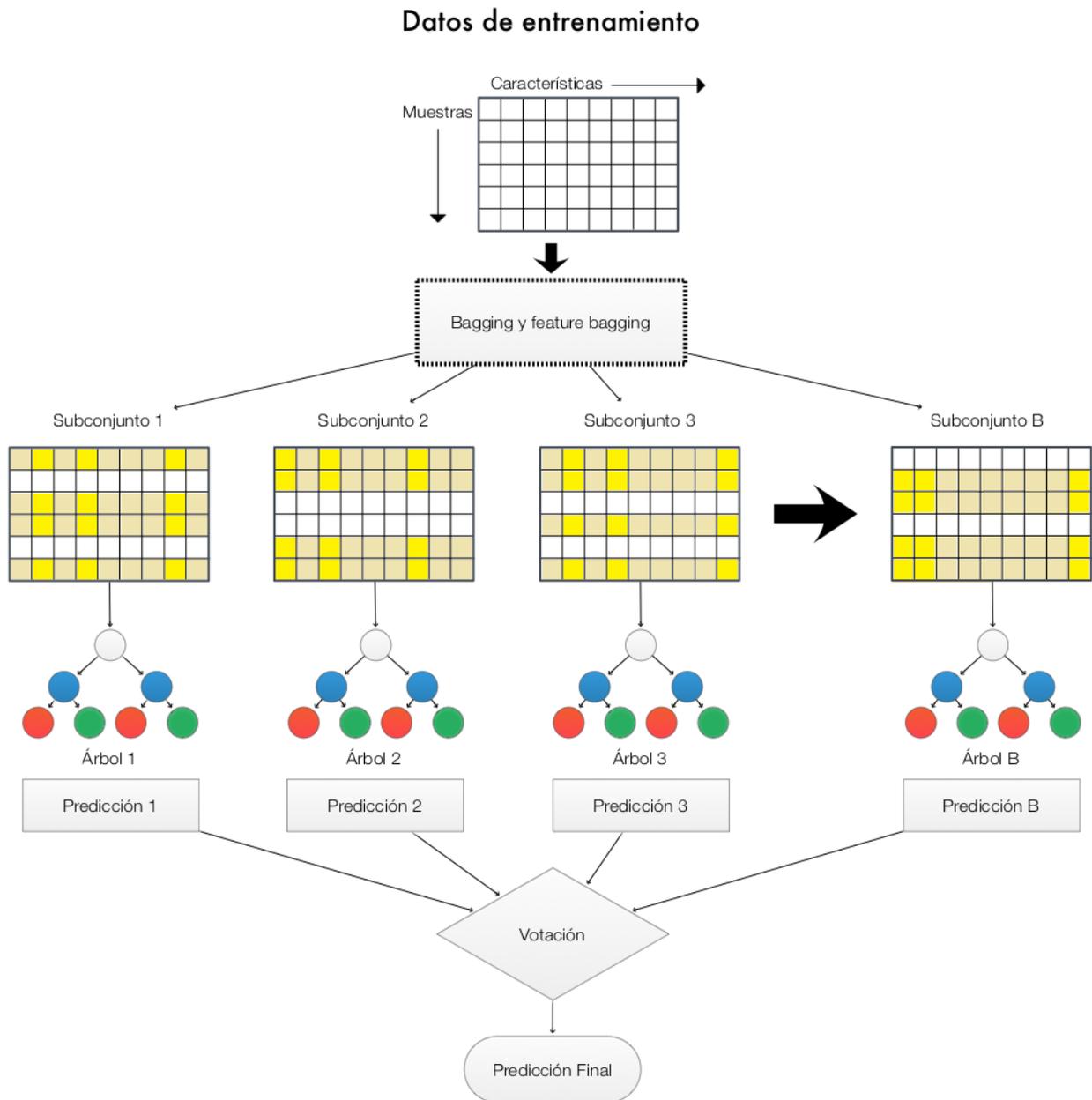
Los clasificadores *naive Bayes*, Bayes ingenuo en español, pertenecen a la familia de clasificadores probabilísticos, su funcionamiento se basa en la aplicación del teorema de Bayes la suposición de independencia fuerte (ingenua) entre las características.

*Naive Bayes* es una técnica que se introdujo para categorización de texto [148] que, hasta nuestros días, sigue siendo popular en este tipo de aplicaciones, por ejemplo: separar correo *spam*, discernir entre si un texto pertenece al ámbito político o deportivo, etcétera. Con un pre procesamiento adecuado y en este tipo de aplicaciones, se muestra competitivo frente a otros métodos de clasificación más complejos como las máquinas de soporte vectorial [149]. También se pueden encontrar aplicaciones en diagnóstico médico [150].

*Naive Bayes* es una técnica simple para construir clasificadores, es decir, no es un algoritmo único para entrenar tales clasificadores, sino una familia de algoritmos basados en un principio común: todos los clasificadores *naive Bayes* suponen que el valor de una característica particular es independiente del valor de cualquier otra característica, dada la variable de clase.

Para algunos tipos de modelos de probabilidad, los clasificadores *naive Bayes* pueden ser entrenados de manera muy eficiente en un entorno de aprendizaje supervisado. En muchas aplicaciones prácticas, la estimación de parámetros para modelos *naive Bayes* utiliza el método de máxima verosimilitud; en otras palabras, uno puede trabajar con el modelo *naive Bayes* sin aceptar la probabilidad bayesiana o usar cualquier método bayesiano.

A pesar de su diseño simplista, los clasificadores *naive Bayes* han funcionado bastante bien en muchas situaciones complejas del mundo real. En 2004, un análisis del problema de clasificación bayesiano mostró que hay razones teóricas sólidas para la eficacia de los clasificadores *naive Bayes* [151]. Aún así, una comparación exhaustiva con otros algoritmos de clasificación en 2006 mostró que la clasificación de Bayes está superada por otros enfoques, como *Random Forest* [152].



**Figura B.6:** Esquema del algoritmo del Random Forest. Los subconjuntos se forman aleatoriamente a partir de los datos de entrenamiento aplicando bagging y feature bagging. Las muestras seleccionadas a partir del bagging están de color amarillo pálido las características elegidas a partir del feature bagging están en color amarillo intenso.

B.4.1 MODELO PROBABILÍSTICO

*Naive Bayes* es un modelo de probabilidad condicional. Se pretende clasificar al vector  $x = (x_1, \dots, x_n)$ , que contiene  $n$  características (variables independientes) y se le asigna la probabilidad:

$$p(C_k|x_1, \dots, x_n), \tag{B.16}$$

para cada posible clase o etiqueta  $k$ . El problema con esta formulación es que si el número de características  $n$  es grande o si una característica puede asumir una gran cantidad de valores, entonces no es factible basar dicho modelo en tablas de probabilidad. Por lo tanto, el modelo es reformulado usando el teorema de Bayes, entonces, la probabilidad condicional se puede definir como:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}. \tag{B.17}$$

En la práctica, sólo es de interés el numerador de la ecuación B.17, porque  $p(x)$  no depende de las clases  $C$  y los valores  $x_i$  ya son conocidos. El denominador es equivalente al modelo de probabilidad conjunta:

$$p(C_k, x_1, \dots, x_n), \tag{B.18}$$

que puede ser reescrito usando la regla de la cadena, aplicando repetidas veces la definición de probabilidad condicional:

$$p(C_k, x_1, \dots, x_n) = p(x_1, \dots, x_n, C_k) \tag{B.19}$$

$$\begin{aligned} &= p(x_1|x_2, \dots, x_n, C_k)p(x_2, \dots, x_n, C_k) \\ &= p(x_1|x_2, \dots, x_n, C_k)p(x_2|x_3, \dots, x_n, C_k)p(x_3, \dots, x_n, C_k) \\ &= \dots \\ &= p(x_1|x_2, \dots, x_n, C_k)p(x_2|x_3, \dots, x_n, C_k) \dots p(x_{n-1}|x_n, C_k)p(x_n|C_k)p(C_k) \end{aligned} \tag{B.20}$$

Ahora vienen las suposiciones ingenuas (*naive*) de independencia condicional, se asume que cada característica  $x_i$  es condicionalmente independiente de todas las demás características  $x_j$  para  $i \neq j$ , dada la clase  $C_k$ . Esto significa que:

$$p(x_i|x_{i+1}, \dots, x_n, C_k) = p(x_i|C_k). \tag{B.21}$$

Por lo tanto, el modelo conjunto se puede expresar como:

$$p(C_k|x_1, \dots, x_n) \propto p(C_k, x_1, \dots, x_n) = \tag{B.22}$$

$$= p(C_k)p(x_1|C_k)p(x_2|C_k)p(x_3|C_k) \dots \tag{B.23}$$

$$= p(C_k) \prod_{i=1}^n p(x_i|C_k) \tag{B.24}$$

Esto significa que bajo las condiciones de independencia, la distribución condicional sobre la clase  $C$  es:

$$p(C_k|x_1, \dots, x_n) = \frac{1}{p(x)} p(C_k) \prod_{i=1}^n p(x_i|C_k), \quad (\text{B.25})$$

donde  $p(x) = \sum_k p(C_k)p(x|C_k)$  es un factor de escala que depende sólo de  $x_1, \dots, x_n$ , que son constantes si los valores de las características son conocidos.

El clasificador *naive* Bayes combina este modelo de características independientes con una regla de decisión. La regla más es la de decisión máxima *a posteriori*. El clasificador correspondiente, un clasificador de Bayes, es la función que asigna una etiqueta de clase  $\hat{y} = C_k$  para alguna  $k$ :

$$\hat{y} = \arg \max_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i|C_k). \quad (\text{B.26})$$

#### B.4.2 ESTIMACIÓN DE PARÁMETROS Y MODELOS DE EVENTOS

Todos los parámetros del modelo (por ejemplo, clases *a priori* y características de las distribuciones de probabilidad) se puede aproximar con frecuencias relativas del conjunto de entrenamiento. Estas son las estimaciones de máxima verosimilitud de las probabilidades. Una clase *a priori* se puede calcular asumiendo clases equiprobables. Para la estimación de los parámetros de la distribución de una característica, se debe asumir una distribución o generar modelos de estadística no paramétrica de las características del conjunto de entrenamiento.

Las hipótesis sobre las distribuciones de características son llamadas el modelo de eventos del Clasificador Bayesiano Ingenuo. La distribución multinomial y la distribución de Bernoulli son populares para aplicaciones como la clasificación de documentos. Cuando se trata con variables continuas, una de las hipótesis más utilizada es que los valores están asociados con cada clase según una distribución Gaussiana o normal.

Suponiendo que los vectores de características  $x$  son continuos, se separan los vectores por clase para calcular su media y varianza. Si  $\mu_k$  es la media del vector asociada a la clase  $k$  y  $\sigma_k^2$  es la varianza del vector asociado a la clase  $k$ . Suponiendo que se adquiere una observación  $v$ , entonces, la densidad de probabilidad de  $v$  dada una clase  $C_k$  es  $P(x = v|C_k)$ , y puede ser calculada agregando  $v$  en la ecuación para una distribución Gaussiana parametrizada por  $\mu_k$  y  $\sigma_k^2$ , es decir:

$$P(x = v|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}. \quad (\text{B.27})$$

Otra técnica común para la manipulación de valores continuos es usar *binning* para discretizar los valores de las características, obteniendo un nuevo conjunto de características de la distribución de Bernoulli. En general, el método de distribución es una mejor opción si hay pocos datos de entrenamiento, o si se conoce la distribución precisa de los datos. El método de discretización tiende a

ser mejor si hay una gran cantidad de datos de entrenamiento, ya que va a aprender para adaptarse a la distribución de los datos. *Naive Bayes* se utiliza normalmente cuando está disponible una gran cantidad de datos (los modelos más caros computacionalmente pueden lograr una mayor precisión), se prefiere generalmente el método de discretización que el método de distribución [153].

# Bibliografía

- [1] Shapiro, L., Stockman, G.C.: Computer vision. Prentice Hall (2001)
- [2] Xie, X., Mirmehdi, M. In: A Galaxy of Texture Features. Imperial College Press 375–406
- [3] Haralick, R.M.: Statistical and structural approaches to texture. *Proceedings of the IEEE* **67**(5) (May 1979) 786–804
- [4] Wechsler, H.: Texture analysis, a survey. *Signal Processing* **2**(3) (1980) 271 – 282
- [5] Gool, L.V., Dewaele, P., Oosterlinck, A.: Texture analysis anno 1983. *Computer Vision, Graphics, and Image Processing* **29**(3) (1985) 336 – 357
- [6] Price, K.E., Vilnrotter, F.M., Nevatia, R.: Structural analysis of natural textures. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **8** (01 1986) 76–89
- [7] Reed, T., Dubuf, J.: A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding* **57**(3) (1993) 359 – 372
- [8] Tuceryan, M., Jain, A. In: Texture Analysis. World Scientific Publishing 235–276
- [9] Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**(6) (Nov 1973) 610–621
- [10] Giannakis, G., Tsatsanis, M.: Object and texture classification using higher order statistics. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **14** (07 1992) 733–750
- [11] Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* **29**(1) (1996) 51 – 59
- [12] Mandelbrot, B.B., Pignoni, R.: The fractal geometry of nature. Volume 173. WH freeman New York (1983)
- [13] Mao, J., Jain, A.K.: Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition* **25**(2) (1992) 173 – 188
- [14] Li, S.Z.: Markov Random Field Modeling in Image Analysis. Springer-Verlag, Berlin, Heidelberg (2001)
- [15] Zucker, S.W.: Toward a model of texture. *Computer Graphics and Image Processing* **5**(2) (1976) 190 – 202
- [16] Beck, J., Sutter, A., Ivry, R.: Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing* **37**(2) (1987) 299 – 325
- [17] Ade, F.: Characterization of textures by eigenfilters. *Signal Processing* **5**(5) (1983) 451 – 457

- [18] Jolliffe, I. In: *Principal Component Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg (2011) 1094–1096
- [19] Coifman, R.R., Meyer, Y., Wickerhauser, M.V., et al.: Size properties of wavelet packets. *Wavelets and their applications* (1992) 453–470
- [20] Turner, M.R.: Texture discrimination by Gabor functions. *Biological Cybernetics* **55**(2) (Nov 1986) 71–82
- [21] Clark, M., Bovik, A.C., Geisler, W.S.: Texture segmentation using Gabor modulation-demodulation. *Pattern Recognition Letters* **6**(4) (1987) 261 – 267
- [22] Estudillo-Romero, A., Escalante-Ramirez, B.: Rotation-invariant texture features from the steered hermite transform. *Pattern Recognition Letters* **32**(16) (2011) 2150 – 2162 *Advances in Theory and Applications of Pattern Recognition, Image Processing and Computer Vision*.
- [23] Belkasim, S., Shridhar, M., Ahmadi, M.: Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition* **24**(12) (1991) 1117 – 1138
- [24] Flusser, J., Suk, T.: Pattern recognition by affine moment invariants. *Pattern Recognition* **26**(1) (1993) 167 – 174
- [25] Heywood, M., Noakes, P.: Fractional central moment method for movement-invariant object classification. *IEE Proceedings-Vision, Image and Signal Processing* **142**(4) (1995) 213–219
- [26] Goshtasby, A.: Template matching in rotated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-7**(3) (May 1985) 338–344
- [27] Ghosal, S., Mehrotra, R.: Orthogonal moment operators for subpixel edge detection. *Pattern Recognition* **26**(2) (1993) 295 – 306
- [28] Mukundan, R., Ramakrishnan, K.: A quaternion solution to the pose determination problem for rendezvous and docking simulations. *Mathematics and Computers in Simulation* **39**(1) (1995) 143 – 153
- [29] Markandey, V., deFigueiredo, R.J.P.: Robot sensing techniques based on high-dimensional moment invariants and tensors. *IEEE Transactions on Robotics and Automation* **8**(2) (April 1992) 186–195
- [30] Hua-Sheng Hsu, W.H.T.: Moment-preserving edge detection and its application to image data compression. *Optical Engineering* **32** (1993) 32 – 32 – 13
- [31] Bigun, J., du Buf, J.M.H.: N-folded symmetries by complex moments in Gabor space and their application to unsupervised texture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(1) (Jan 1994) 80–87
- [32] Wang, M., Knoesen, A.: Rotation- and scale-invariant texture features based on spectral moment invariants. *Journal of the Optical Society of America A* **24**(9) (Sep 2007) 2550–2557
- [33] Mukundan, R., Ong, S.H., Lee, P.A.: Image analysis by Tchebichef moments. *IEEE Transactions on Image Processing* **10**(9) (Sept 2001) 1357–1364
- [34] Teague, M.R.: Image analysis via the general theory of moments. *Journal of the Optical Society of America* **70**(8) (Aug 1980) 920–930

- [35] Khotanzad, A., Hong, Y.H.: Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(5) (May 1990) 489–497
- [36] Mukundan, R., Ramakrishnan, K.: Fast computation of Legendre and Zernike moments. *Pattern Recognition* **28**(9) (1995) 1433 – 1442
- [37] Wallin, A., Kubler, O.: Complete sets of complex Zernike moment invariants and the role of the pseudoinvariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**(11) (Nov 1995) 1106–1110
- [38] Mukundan, R., Ramakrishnan, K.: Moment functions in image analysis, theory and applications. World Scientific (1998)
- [39] Temme, N.M.: Special functions: An introduction to the classical functions of mathematical physics. John Wiley & Sons (2011)
- [40] Yap, P., Paramesran, R., Ong, S.H.: Image analysis by Krawtchouk moments. *IEEE Transactions on Image Processing* **12**(11) (Nov 2003) 1367–1377
- [41] Flusser, J., Zitova, B., Suk, T.: Moments and moment invariants in pattern recognition. John Wiley & Sons, Chichester (2009)
- [42] Bayraktar, B., Bernas, T., Robinson, J.P., Rajwa, B.: A numerical recipe for accurate image reconstruction from discrete orthogonal moments. *Pattern Recognition* **40**(2) (2007) 659 – 669
- [43] Mukundan, R.: Some computational aspects of discrete orthonormal moments. *IEEE Transactions on Image Processing* **13**(8) (Aug 2004) 1055–1059
- [44] Yap, P., Paramesran, R., Ong, S.: Image analysis using Hahn moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(11) (Nov 2007) 2057–2062
- [45] Zhu, H., Shu, H., Zhou, J., Luo, L., Coatrieux, J.: Image analysis by discrete orthogonal dual Hahn moments. *Pattern Recognition Letters* **28**(13) (2007) 1688 – 1704
- [46] Zhu, H., Shu, H., Liang, J., Luo, L., Coatrieux, J.L.: Image analysis by discrete orthogonal Racah moments. *Signal Processing* **87**(4) (2007) 687 – 708
- [47] Asli, B.H.S., Flusser, J.: New discrete orthogonal moments for signal analysis. *Signal Processing* **141** (2017) 57 – 73
- [48] Marcos, J.V., Cristóbal, G.: Texture classification using discrete Tchebichef moments. *Journal of the Optical Society of America A* **30**(8) (Aug 2013) 1580–1591
- [49] Hu, M.K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* **8**(2) (February 1962) 179–187
- [50] Dudani, S.A., Breeding, K.J., McGhee, R.B.: Aircraft identification by moment invariants. *IEEE Transactions on Computers* **C-26**(1) (Jan 1977) 39–46
- [51] Casasent, D., Cheatham, R.: Image segmentation and real-image tests for an optical moment-based feature extractor. *Optics Communications* **51**(4) (1984) 227 – 230
- [52] Talenti, G.: Recovering a function from a finite number of moments. *Inverse Problems* **3**(3) (1987) 501
- [53] Nikiforov, A.F., Uvarov, V.B.: Special functions of mathematical physics. Springer (1988)

- [54] Armenta, J.: Polinomios ortogonales y procesos estocásticos. Tesis de licenciatura, Facultad de Ciencias, Universidad Nacional Autónoma de México (2017)
- [55] Sim, D.G., Kim, H.K., Park, R.H.: Invariant texture retrieval using modified Zernike moments. *Image and Vision Computing* **22**(4) (2004) 331 – 342
- [56] Zernike, F.v.: Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica* **1**(7) (1934) 689 – 704
- [57] Bhatia, A.B., Wolf, E.: On the circle polynomials of Zernike and related orthogonal sets. *Mathematical Proceedings of the Cambridge Philosophical Society* **50**(1) (1954)
- [58] Sheng, Y., Shen, L.: Orthogonal Fourier-Mellin moments for invariant pattern recognition. *Journal of the Optical Society of America A* **11**(6) (Jun 1994) 1748–1757
- [59] Jackson, D.: Fourier series and orthogonal polynomials. Courier Corporation (2012)
- [60] Teh, C., Chin, R.T.: On image analysis by the methods of moments. In: Proceedings CVPR 88: The Computer Society Conference on Computer Vision and Pattern Recognition. (June 1988) 556–561
- [61] Kan, C., Srinath, M.D.: Invariant character recognition with Zernike and orthogonal Fourier–Mellin moments. *Pattern Recognition* **35**(1) (2002) 143 – 154 Shape representation and similarity for image databases.
- [62] Wang, L., Healey, G.: Using Zernike moments for the illumination and geometry invariant classification of multispectral texture. *IEEE Transactions on Image Processing* **7**(2) (Feb 1998) 196–203
- [63] Zhang, D., Lu, G.: Evaluation of mpeg-7 shape descriptors against other shape descriptors. *Multimedia Systems* **9**(1) (Jul 2003) 15–30
- [64] Kim, W.Y., Kim, Y.S.: Robust rotation angle estimator. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(8) (Aug 1999) 768–773
- [65] Bin, Y., Jia-Xiong, P.: Invariance analysis of improved Zernike moments. *Journal of Optics A: Pure and Applied Optics* **4**(6) (2002) 606
- [66] Teh, C.H., Chin, R.T.: On digital approximation of moment invariants. *Computer Vision, Graphics, and Image Processing* **33**(3) (1986) 318–326
- [67] Kamgar-Parsi, B., Kamgar-Parsi, B.: Evaluation of quantization error in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(9) (Sept 1989) 929–940
- [68] De Vega, H.J. In: Yang-Baxter Algebras, Integrable Theories and Quantum Groups. Springer US, Boston, MA (1990) 387–434
- [69] Koepf, W., Schmersau, D.: On a structure formula for classical q-orthogonal polynomials (2001)
- [70] Álvarez-Nodarse, R., Arvesú, J., Yáñez, R.J.: On the connection and linearization problem for discrete hypergeometric q-polynomials. *Journal of Mathematical Analysis and Applications* **257**(1) (2001) 52 – 78
- [71] Álvarez-Nodarse, R., Smirnov, Y.F.: The dual Hahn q-polynomials in the lattice and the q-algebras. *Journal of Physics A: Mathematical and General* **29**(7) (1996) 1435

- [72] Temme, N., López, J.: The Askey scheme for hypergeometric orthogonal polynomials viewed from asymptotic analysis. *Journal of Computational and Applied Mathematics* **133**(1) (2001) 623 – 633 5th Int. Symp. on Orthogonal Polynomials, Special Functions and their Applications.
- [73] Krawtchouk, M.: On interpolation by means of orthogonal polynomials. *Memoirs Agricultural Inst. Kiev* **4** (1929) 21–28
- [74] Koekoek, R., Swarttouw, R.: The Askey–scheme of hypergeometric orthogonal polynomials and its q-analogue (February 1996)
- [75] Shmaliy, Y.S.: An unbiased FIR filter for TIE model of a local clock in applications to GPS-based timekeeping. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **53**(5) (May 2006) 862–870
- [76] Shmaliy, Y.S., Morales-Mendoza, L.J.: FIR smoothing of discrete-time polynomial signals in state space. *IEEE Transactions on Signal Processing* **58**(5) (May 2010) 2544–2555
- [77] Heinonen, P., Neuvo, Y.: FIR-median hybrid filters with predictive FIR substructures. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**(6) (June 1988) 892–899
- [78] Morales-Mendoza, L.J., Gamboa-Rosales, H., Shmaliy, Y.S.: A new class of discrete orthogonal polynomials for blind fitting of finite data. *Signal Processing* **93**(7) (2013) 1785 – 1793
- [79] Papakostas, G., Koulouriotis, D., Karakasis, E.: Computation strategies of orthogonal image moments: A comparative study. *Applied Mathematics and Computation* **216**(1) (2010) 1 – 17
- [80] Papakostas, G., Boutalis, Y., Papaodysseus, C., Fragoulis, D.: Numerical error analysis in Zernike moments computation. *Image and Vision Computing* **24**(9) (2006) 960 – 969
- [81] Li, B., Meng, M.Q.H.: Computer-aided detection of bleeding regions for capsule endoscopy images. *IEEE Transactions on Biomedical Engineering* **56**(4) (2009) 1032–1039
- [82] Wu, K., Garnier, C., Coatrieux, J., Shu, H.: A preliminary study of moment-based texture analysis for medical images. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. (Aug 2010) 5581–5584
- [83] Mukundan, R.: Local Tchebichef moments for texture analysis. In: Engineering: Chapters and Books. Science Gate Publishing (2014) 127–142
- [84] Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. *International journal of computer vision* **62**(1) (2005) 61–81
- [85] Wee, C.Y., Paramesran, R., Mukundan, R., Jiang, X.: Image quality assessment by discrete orthogonal moments. *Pattern Recognition* **43**(12) (2010) 4055–4068
- [86] Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A* **2**(7) (Jul 1985) 1160–1169
- [87] Daugman, J.G.: Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research* **20**(10) (1980) 847–856
- [88] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7) (July 2002) 971–987

- [89] Zhu, S.C., Guo, C.E., Wang, Y., Xu, Z.: What are Textons? *International Journal of Computer Vision* **62**(1) (2005) 121–143
- [90] Hunt, E.B., Marin, J., Stone, P.J.: Experiments in induction. Academic Press (1966)
- [91] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and regression trees. CRC press (1984)
- [92] De Mántaras, R.L.: A distance-based attribute selection measure for decision tree induction. *Machine Learning* **6**(1) (Jan 1991) 81–92
- [93] Smyth, P., Goodman, R.M., Higgins, C.M.: A hybrid rule-based/Bayesian classifier. In: Proceedings of the 9th European Conference on Artificial Intelligence. ECAI'90, London, UK, UK, Pitman Publishing (1990) 610–615
- [94] Kononenko, I.: On biases in estimating multi-valued attributes. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2. IJCAI'95, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1995) 1034–1040
- [95] Sokal, R.R., Rohlf, F.J.: Biometry wh freeman and company. San Francisco **859** (1981)
- [96] John, G.H., Kohavi, R., Pfleger, K.: Irrelevant features and the subset selection problem. In: Machine Learning Proceedings 1994. Elsevier (1994) 121–129
- [97] Svetnik, V., Liaw, A., Tong, C., Wang, T.: Application of breimans random forest to modeling structure-activity relationships of pharmaceutical molecules. In: International Workshop on Multiple Classifier Systems, Springer (2004) 334–343
- [98] Mitchell, M.: An introduction to genetic algorithms. MIT press (1998)
- [99] Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**(4598) (1983) 671–680
- [100] Rutenbar, R.A.: Simulated annealing algorithms: An overview. *IEEE Circuits and Devices Magazine* **5**(1) (1989) 19–26
- [101] Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**(19) (2007) 2507–2517
- [102] Ambroise, C., McLachlan, G.J.: Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences* **99**(10) (2002) 6562–6566
- [103] Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: European conference on machine learning, Springer (1994) 171–182
- [104] Kira, K., Rendell, L.A.: A practical approach to feature selection. In: Proceedings of the ninth international workshop on Machine learning. (1992) 249–256
- [105] Kira, K., Rendell, L.A.: The feature selection problem: Traditional methods and a new algorithm. In: Proceedings of the Tenth National Conference on Artificial Intelligence. AAAI'92, AAAI Press (1992) 129–134
- [106] Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**(2) 179–188

- [107] McLachlan, G.: Discriminant analysis and statistical pattern recognition. Volume 544. John Wiley & Sons (2004)
- [108] Wetcher-Hendricks, D.: Analyzing quantitative data: An introduction for social researchers. John Wiley & Sons (2011)
- [109] Martínez, A.M., Kak, A.C.: PCA versus LDA. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2) (2001) 228–233
- [110] Lebart, L.: Correspondence analysis. In: Data Science, Classification, and Related Methods: Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96), Kobe, Japan, March 27–30, 1996, Springer Science & Business Media (2013) 423
- [111] Büyüköztürk, Ş., Çokluk-Bökeoğlu, Ö.: Discriminant function analysis: Concept and application. *Eurasian Journal of Educational Research* (33) (2008) 73–92
- [112] Lachenbruch, P.A., Goldstein, M.: Discriminant analysis. *Biometrics* **35**(1) (1979) 69–85
- [113] Klecka, W.R.: Discriminant analysis. Volume 19. SAGE Publishing (1980)
- [114] Rao, C.R.: The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B* **10**(2) (1948) 159–203
- [115] Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468). (Aug 1999) 41–48
- [116] González, G., Nava, R., Escalante-Ramírez, B.: A comparative study on discrete Shmaliy moments and their texture-based applications. *Mathematical Problems in Engineering* **2018** (2018)
- [117] Shamir, L., Orlov, N., Eckley, D.M., Macura, T.J., Goldberg, I.G.: IICBU 2008: a proposed benchmark suite for biological image analysis. *Medical & biological engineering & computing* **46**(9) (2008) 943–947
- [118] Shamir, L., Orlov, N., Goldberg, I.G.: Evaluation of the informativeness of multi-order image transforms. In: Image Processing and Computer Vision (IPCV). (2009) 37–42
- [119] Siji, K.K., Mathew, B.S., Chandran, R., Shajeemohan, B.S., Shanthini, K.S.: Feature selection, optimization and performance analysis of classifiers for biological images. In: 2014 IEEE National Conference on Communication, Signal Processing and Networking (NCCSN). (Oct 2014) 1–5
- [120] Di Ruberto, C., Fodde, G., Putzu, L.: Comparison of statistical features for medical colour image classification. In: International Conference on Computer Vision Systems, Springer (2015) 3–13
- [121] Zhang, B., Lu, W.: Classification of malignant lymphomas by classifier ensemble with multiple texture features. In: Life System Modeling and Intelligent Computing. Springer (2010) 155–164
- [122] Coelho, L.P., Kangas, J.D., Naik, A.W., Osuna-Highley, E., Glory-Afshar, E., Fuhrman, M., Simha, R., Berget, P.B., Jarvik, J.W., Murphy, R.F.: Determining the subcellular location of new proteins from microscope images using local features. *Bioinformatics* **29**(18) (2013) 2343–2349
- [123] Dunn, O.J., Clark, V.A.: Applied Statistics: Analysis of Variance and Regression. John Wiley & Sons, Inc., New York, NY, USA (1986)

- [124] Meng, T., Lin, L., Shyu, M., Chen, S.: Histology image classification using supervised classification and multimodal fusion. In: 2010 IEEE International Symposium on Multimedia. (Dec 2010) 145–152
- [125] Orlov, N., Shamir, L., Macura, T., Johnston, J., Eckley, D.M., Goldberg, I.G.: WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recognition Letters* **29**(11) (2008) 1684 – 1693
- [126] Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Johnston, J., Goldberg, I.G.: WNDCHRM – an open source utility for biological image analysis. *Source Code for Biology and Medicine* **3**(1) (Jul 2008) 13
- [127] Nava, R., González, G., Kybic, J., Escalante-Ramírez, B.: Classification of tumor epithelium and stroma in colorectal cancer based on discrete Tchebichef moments. In: Workshop on Clinical Image-Based Procedures, Springer (2015) 79–87
- [128] Nava, R., González, G., Kybic, J., Escalante-Ramírez, B.: Characterization of hematologic malignancies based on discrete orthogonal moments. In: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). (Dec 2016) 1–6
- [129] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F.: Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International Journal of Cancer* **136**(5) E359–E386
- [130] Linder, N., Konsti, J., Turkki, R., Rahtu, E., Lundin, M., Nordling, S., Haglund, C., Ahonen, T., Pietikäinen, M., Lundin, J.: Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagnostic Pathology* **7**(1) (Mar 2012) 22
- [131] Huang, H.L., Hsu, M.H., Lee, H.C., Charoenkwan, P., Ho, S.J., Ho, S.Y.: Prediction of mouse senescence from HE-Stain liver images using an ensemble SVM classifier. In: Asian Conference on Intelligent Information and Database Systems, Springer (2013) 325–334
- [132] Orlov, N.V., Chen, W.W., Eckley, D.M., Macura, T.J., Shamir, L., Jaffe, E.S., Goldberg, I.G.: Automatic classification of lymphoma images with transform-based global features. *IEEE Transactions on Information Technology in Biomedicine* **14**(4) (2010) 1003–1013
- [133] Sertel, O., Kong, J., Lozanski, G., Shana’ah, A., Catalyurek, U., Saltz, J., Gurcan, M.: Texture classification using nonlinear color quantization: Application to histopathological image analysis. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. (March 2008) 597–600
- [134] Tuzel, O., Yang, L., Meer, P., Foran, D.J.: Classification of hematologic malignancies using Texton signatures. *Pattern Analysis and Applications* **10**(4) (2007) 277–290
- [135] Ruifrok, A., Johnston, D.: Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology* **23**(4) (August 2001) 291–299
- [136] Wang, L., He, D.C.: Texture classification using texture spectrum. *Pattern Recognition* **23**(8) (1990) 905–910
- [137] Ojala, T., Valkealahti, K., Oja, E., Pietikäinen, M.: Texture discrimination with multidimensional distributions of signed gray-level differences. *Pattern Recognition* **34**(3) (2001) 727–739
- [138] Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* **20**(3) (Sep 1995) 273–297

- [139] Vapnik, V.: The nature of statistical learning theory. Springer science & business media (2013)
- [140] Burges, C.J.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**(2) (Jun 1998) 121–167
- [141] Ho, T.K.: Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. Volume 1. (Aug 1995) 278–282 vol.1
- [142] Ho, T.K.: The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(8) (Aug 1998) 832–844
- [143] Kleinberg, E.M.: An overtraining-resistant stochastic modeling method for pattern recognition. *The Annals of Statistics* **24**(6) (1996) 2319–2349
- [144] Breiman, L.: Random forests. *Machine Learning* **45**(1) (Oct 2001) 5–32
- [145] Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. *Neural Computation* **9**(7) (October 1997) 1545–1588
- [146] Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning. Springer series in statistics New York, NY, USA: (2001)
- [147] Ho, T.K.: A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis & Applications* **5**(2) (2002) 102–112
- [148] Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited, (2016)
- [149] Rennie, J., Shih, L., Teevan, J., Karger, D.: Tackling the poor assumptions of naive Bayes classifiers. In: Proceedings of the 20th international conference on machine learning (ICML-03), International Conference on Machine Learning (2003) 616–623
- [150] Rish, I., et al.: An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, IBM New York (2001) 41–46
- [151] Zhang, H.: The optimality of naive Bayes. *Association for the Advancement of Artificial Intelligence* **1**(2) (2004) 3–16
- [152] Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd international conference on Machine learning, ACM (2006) 161–168
- [153] Hand, D.J., Yu, K.: Idiots Bayes, not so stupid after all?. *International Statistical Review* **69**(3) 385–398