

Técnicas basadas en kernel para el análisis de texturas en imagen biomédica

Autor: Carlos Fernández Lozano

Tesis doctoral UDC 2014

Directores: Julián Dorado de la Calle / Marcos Gestal Pose

Programa de doctorado en Tecnologías de la Información y las Comunicaciones



UNIVERSIDADE DA CORUÑA

D. **JULIÁN DORADO DE LA CALLE**, Profesor Titular de Universidad en el área de Ciencias de la Computación e Inteligencia artificial de la Universidade da Coruña y D. **MARCOS GESTAL POSE**, Profesor Contratado Doctor en el área de Ciencias de la Computación e Inteligencia Artificial de la Universidade da Coruña, pertenecientes al Departamento de Tecnologías de la Información y las Comunicaciones como directores de la Tesis Doctoral titulada “Técnicas basadas en kernel para el análisis de texturas en Imagen Biomédica” del doctorando Carlos Fernández Lozano

HACEN CONSTAR QUE:

La tesis titulada “**Técnicas basadas en kernel para el análisis de texturas en imagen biomédica**” realizada por D. **Carlos Fernández Lozano**, bajo nuestra dirección en el Departamento de Tecnologías de la Información y las Comunicaciones cumple los requisitos para optar a grado de Doctor y pretende obtener la mención internacional

Y para que conste, firman la presente a 30 de Enero de 2014,

Julián Dorado de la Calle

Marcos Gestal Pose

*A mi mujer y mis hijos,
por el tiempo robado*

*“Investigar es ver lo que todo el mundo ya ha visto
y pensar lo que nadie ha pensado todavía.”*

Albert Szent-Györgyi. (1893-1986)

Fisiólogo Húngaro

Agradecimientos

En primer lugar, quiero agradecer al Laboratorio RNASA-IMEDIR y a sus responsables por el apoyo dado para la realización de esta Tesis Doctoral y por proporcionarme los medios necesarios para realizarla.

Especialmente a mis directores el Dr. Julián Dorado de la Calle y el Dr. Marcos Gestal Pose por su tiempo y dedicación y por todo lo que he podido aprender trabajando a su lado.

Al profesor Stefano Cagnoni y su grupo de investigación en la *Università degli studi di Parma* por acogerme durante mi estancia en Parma, y al profesor Colin Campbell y Thomas R. Gaunt de la *University of Bristol*, por todo lo que pude aprender de ellos.

A mis padres y mi hermana, por creer en mí y disfrutar este momento tanto como yo.

A José Antonio, que ha sido mi mentor en este mundo de la investigación y sin el cual, estaría perdido en muchas ocasiones.

A mi mujer Patricia y a mis hijos Samuel y Raquel por la paciencia que han tenido y lo bien que han llevado el tiempo que les he robado para lograr terminar esta Tesis Doctoral. Siempre una sonrisa y una palabra de ánimo para continuar con el trabajo, incluso cuando las estancias me tenían fuera de casa.

A mis amigos Merexo, Yoda y Rosmi por vuestro apoyo, ánimos, comprensión, paciencia y eternas sesiones de calma para focalizar el final del camino.

A todos vosotros, muchas gracias.

Resumo

En problemas do mundo real é relevante o estudo da importancia de todas as variables obtidas de maneira que sexa posible a eliminación de ruído, é neste punto onde xorden as técnicas de selección de variables. O obxectivo destas técnicas é pois encontrar o subconxunto de variables que describan do mellor xeito posible a información útil contida nos datos permitindo mellorar o rendemento.

En espazos de alta dimensionalidade son especialmente interesantes as técnicas baseadas en *kernel*, onde demostraron unha alta eficiencia debido á súa capacidade para xeneralizar nos devanditos espazos.

Neste traballo realízase unha nova proposta para a análise de texturas en imaxe biomédica mediante a integración, utilizando técnicas baseadas en *kernel*, de diferentes tipos de datos de textura para a selección das variables máis representativas co obxectivo de mellorar os resultados obtidos en clasificación e en interpretabilidade das variables obtidas.

Para validar esta proposta formalizouse un deseño experimental con catro fases diferenciadas: extracción e preprocesar dos datos, aprendizaxe e selección do mellor modelo asegurando a reproducibilidade dos resultados á vez que unha comparación en condicións de igualdade. Utilizáronse imaxes de xeles de electroforese bidimensional.

Resumen

En problemas del mundo real es relevante el estudio de la importancia de todas las variables obtenidas de manera que sea posible la eliminación de ruido, es en este punto donde surgen las técnicas de selección de variables. El objetivo de estas técnicas es pues encontrar el subconjunto de variables que describan de la mejor manera posible la información útil contenida en los datos permitiendo mejorar el rendimiento.

En espacios de alta dimensionalidad son especialmente interesantes las técnicas basadas en *kernel*, donde han demostrado una alta eficiencia debido a su capacidad para generalizar en dichos espacios.

En este trabajo se realiza una nueva propuesta para el análisis de texturas en imagen biomédica mediante la integración, utilizando técnicas basadas en *kernel*, de diferentes tipos de datos de textura para la selección de las variables más representativas con el objetivo de mejorar los resultados obtenidos en clasificación y en interpretabilidad de las variables obtenidas.

Para validar esta propuesta se ha formalizado un diseño experimental con cuatro fases diferenciadas: extracción y preprocesado de los datos, aprendizaje y selección del mejor modelo asegurando la reproducibilidad de los resultados a la vez que una comparación en condiciones de igualdad.

Abstract

In real-world problems it is of relevance to study the importance of all the variables obtained, so that denoising could be possible, because it is at this point when the variable selection techniques arise. Therefore, these techniques are aimed at finding the subset of variables that describe in the best possible way the useful information contained in the data, allowing improved performance.

In high-dimensional spaces, the kernel-based techniques are of special relevance, as they have demonstrated a high efficiency due to their ability to generalize in these spaces.

In this work, a new approach for texture analysis in biomedical imaging is performed by means of integration. For this procedure, kernel-based techniques were used with different types of texture data for the selection of the most representative variables in order to improve the results obtained in classification and interpretability of the obtained variables.

To validate this proposal, an experimental design has been concluded, consisting of four different phases: 1) Data extraction; 2) Data pre-processing; 3) Learning and 4) Selection of the best model to ensure the reproducibility of results while making a comparison under conditions of equality. In this regard, two-dimensional electrophoresis gel images have been used.

Prefacio

Las observaciones que provienen de problemas del mundo real, a menudo están compuestas de un elevado número de variables. Para cada problema que se intenta resolver, en ocasiones existen múltiples conjuntos de datos que pueden colaborar para resolverlo. Así la cantidad masiva de información existente en la actualidad, hace que cada vez sea más necesario el uso de técnicas de minería de datos o de aprendizaje máquina para extraer información y conocimiento de los datos. En este contexto las técnicas de selección de variables emergen para discriminar cuáles son las variables que aportan más información a la hora de resolver un determinado problema. El objetivo del uso de estas técnicas consiste en encontrar el subconjunto de variables que describan de la mejor manera posible la información útil contenida en las mismas sin empeorar el rendimiento, o incluso mejorándolo. La selección de variables juega un papel vital en el aprendizaje estadístico y los nuevos descubrimientos científicos [1]. No es extraño que diferentes combinaciones de variables sean solución del problema que se está estudiando, por lo que es interesante conocer cuáles son las que pueden aportar la misma información que el conjunto total y poder inferir conocimiento después del proceso de selección.

Éste tipo de técnicas permiten abordar de manera sencilla el problema mediante un estudio explorativo de las posibles soluciones. Este tipo de técnicas suele requerir de un tiempo de desarrollo relativamente pequeño en comparación con el tiempo necesario para, mediante entrenamiento y/o simulación encontrar las mejores soluciones para el problema. Dentro de este tipo de problemas son especialmente interesantes las técnicas basadas en *kernel*, ya que han probado ser excepcionalmente eficientes en problemas de

clasificación debido a su habilidad para generalizar en espacios de alta dimensionalidad [2-7].

Existen varias posibilidades para utilizar dichas técnicas basadas en *kernel* para selección de características, entre las que se van a destacar dos.

- a. En la primera de ellas, se pueden utilizar técnicas de optimización bioinspiradas como los Algoritmos Genéticos y la Optimización por Enjambre de Partículas para que optimicen el proceso de selección de variables. A grandes rasgos estas técnicas asignan un valor de bondad o aptitud a cada una de las soluciones y, mediante un proceso iterativo bioinspirado, exploran y explotan el espacio de soluciones en busca del mejor subconjunto de ellas.
- b. En la segunda opción, se utiliza la capacidad de las técnicas basadas en *kernel* para, o bien mediante procesos iterativos estudiar el peso que cada variable tiene para resolver el problema y eliminar así las menos importantes hasta llegar a la mejor solución, o mediante combinaciones de diferentes *kernels* sencillos estudiar la correlación existente entre ellas y la solución del problema.

Actualmente los problemas no se estudian de manera aislada sino que se buscan interacciones entre conjuntos de datos relacionados. Un ejemplo puede ser el caso de los estudios de asociación genéticos entre genotipos y fenotipos, donde cientos de miles de SNPs son potenciales covariantes de fenotipos (i.e. los niveles de colesterol).

Surge así el concepto de fusión de datos, en el que la idea consiste en incorporar diferentes conjuntos de datos al análisis de manera que pueda producirse un efecto beneficioso en forma de reducción de ruido. Como

consecuencia es posible que se produzca un aumento en la significancia estadística de los resultados obtenidos, además de aprovechar las correlaciones e interacciones entre datos para obtener información más refinada y de más alto nivel [8,9].

Dentro del ámbito biomédico las técnicas de análisis de imágenes son el núcleo de las aplicaciones médicas de ayuda en el diagnóstico y tratamiento de enfermedades; de todas formas, no se debe olvidar que también son ampliamente utilizadas en el dominio educacional. Existen en la actualidad multitud de tipos de imágenes biomédicas, como por ejemplo la resonancia magnética, la tomografía por emisión de positrones, ultrasonidos, tomografía computarizada, geles de electroforesis bidimensional, etc. La primera y principal desventaja que tienen todas las modalidades de imagen médica está en que la interpretación y el modelado de formas están basados en procesos visuales.

El desarrollo de sistemas automáticos de procesado de imagen tiene como objetivo la ayuda en la toma de decisiones diagnósticas y son aplicados a múltiples datos médicos [10]. Las imágenes digitales y las biomédicas están compuestas por píxeles individuales, con valores discretos de brillo o color. Estas aplicaciones se utilizan para el procesado y extracción automáticos de diferentes variables de la imagen para ser utilizadas después en procesos de clasificación o segmentación entre otros. Como ejemplo, las variables que se extraen pueden describir color, morfología o propiedades de textura de una imagen.

El modelo que se va a proponer en esta Tesis Doctoral es general y puede ser aplicado en la fusión de datos en diferentes dominios. En concreto se va a utilizar con datos de textura en imagen de electroforesis bidimensional

debido a que gracias a las particularidades que presentan estas imágenes, tal y cómo se verá en el Capítulo 2, se espera la obtención de buenos resultados.

OBJETIVOS

Algunas de las técnicas más utilizadas hasta el momento para la selección de variables están basadas en modelos matemáticos, mientras que otras lo están en métodos evolutivos o bioinspirados con el fin de explotar las posibilidades de búsqueda en paralelo a lo largo de todo el espacio de soluciones del problema.

Sin embargo, es necesario un estudio en profundidad de las posibilidades que, tal y como se comentó anteriormente, existen con el uso de técnicas basadas en *kernel* para la selección de variables y es el motivo principal por el que se plantea la realización de esta Tesis Doctoral: buscar un mecanismo de selección de variables basado en *kernels* que permita determinar cuáles son las variables que aportan mayor cantidad de información y el mejor ajuste a una solución global del problema sin pérdida de interpretabilidad de las variables solución.

El objetivo específico que se pretende conseguir con esta Tesis Doctoral es la formalización de un nuevo modelo de clasificación y de un diseño experimental para el proceso de selección de variables en problemas de alta dimensionalidad mediante el empleo de técnicas basadas en *kernels*. Más concretamente se particularizará con la integración de información heterogénea de textura en imágenes de geles de electroforesis bidimensional (2D-gel). Para ello, se seguirán los siguientes pasos:

- Estudio de distintas técnicas basadas en *kernel* para mejorar los resultados de clasificación en un análisis de texturas en imagen biomédica.

- Estudio de distintas técnicas de selección de variables y reducción de dimensionalidad basadas en *kernels*. Un análisis de texturas es un problema de elevada dimensionalidad.
- Estudio del uso combinado de técnicas bioinspiradas y basadas en *kernel* para la selección de variables de textura en imagen biomédica.
- Estudio de diferentes métricas para calcular la correlación entre ejemplos y clase de salida.
- Desarrollo de nuevos *kernels* para el análisis de texturas.
- Desarrollo de técnicas basadas en *kernel* para el análisis de textura en imagen biomédica que permita la clasificación y detección de áreas de interés que ayuden en el diagnóstico. Este modelo debe permitir la extracción de conocimiento relacionando las variables texturales con diferentes regiones de interés que pueden tener correspondencia con determinadas patologías en cada tipo de imagen particular.
- Desarrollo de varios prototipos para el estudio de la información textural que permita evaluar la técnica propuesta y probar la hipótesis planteada en el ámbito de aplicación biomédico.
- Propuesta de un diseño experimental que permita realizar una comparación en igualdad de condiciones entre los modelos de referencia y la propuesta presentada en esta Tesis Doctoral.

Como resultado final del trabajo realizado en esta Tesis Doctoral se buscan alcanzar los siguientes aspectos:

- Mejora de los resultados de clasificación: mediante el estudio de las variables que más información aportan para resolver el problema es posible eliminar ruido presente en los datos y mejorar el rendimiento de los clasificadores.
- Mejora del tiempo de ejecución del clasificador: al tener menos datos de entrada, el clasificador mejora en cuanto al tiempo de ejecución para

el procesado de la información y necesitará menos tiempo para devolver un resultado.

- Reducción del tiempo de análisis de imagen: existen una gran cantidad de variables de textura que pueden ser calculadas para una imagen. El proceso de selección de variables permitirá reducir el coste en tiempo y cómputo de las mismas.
- Extracción de conocimiento mejorada: cuanto menor sea el número de variables importantes para el proceso de clasificación, más fácil será sacar inferencias acerca de por qué esa combinación concreta es la solución del problema.
- Selección de técnicas, *kernels* y métricas más adecuadas para la selección y reducción de dimensionalidad de un problema de análisis de texturas.
- Establecimiento de un marco justo de comparación entre diferentes modelos aplicados al mismo conjunto de imágenes.

Por lo tanto, se plantea la obtención de un sistema que aborde la solución de un proceso de selección de variables basado en *kernels*, para lo que se estudiará la combinación de dichas técnicas con técnicas de optimización bioinspiradas y también, diversas técnicas basadas exclusivamente en *kernels* para resolver el problema de selección de variables.

Para esto se aportará además, un diseño experimental claro y conciso para llevar a cabo todo el proceso y se comprobará el correcto funcionamiento del sistema propuesto con un caso real asegurando además que esta comparación se haga en condiciones de igualdad y reproducibilidad. Más concretamente, este caso real está basado en la clasificación de proteínas en imágenes de electroforesis bidimensional mediante un análisis de las texturas presentes en la imagen. En este tipo de imágenes, la forma tanto del ruido

como de las proteínas sigue una curva Gaussiana, por lo que su discriminación mediante otro tipo de técnicas de visión artificial es una tarea muy compleja y se ha de tener en cuenta que las texturas son inmunes al ruido Gaussiano, por lo que son de esperar unos buenos resultados.

CONTENIDOS

ESTRUCTURA DE LA TESIS.....	13
1. FUNDAMENTOS.....	15
1.1 Texturas.....	16
1.1.1 Definición de textura	16
1.1.2 Análisis de Texturas	17
1.1.2.1 Objetivos de un análisis de texturas.....	19
1.1.2.2 Aproximación estadística para el análisis de texturas.....	20
1.1.2.2.1 Operadores de primer orden	21
1.1.2.2.2 Operadores de segundo orden	22
1.1.2.2.3 Operadores de orden superior.....	23
1.2 Metaheurísticas Bioinspiradas.....	25
1.2.1 Algoritmos genéticos	25
1.2.2 Optimización por enjambre de partículas.....	28
1.3 Métodos basados en Kernels.....	30
1.3.1 Máquinas de vectores soporte	30
1.3.1.1 Hiperplano de separación óptimo	32
1.3.1.2 Datos no separables linealmente	35
1.3.1.3 Truco del kernel.....	37
1.3.1.4 Kernels más habituales de propósito general	39
1.3.2 Aprendizaje con múltiples kernels	40
1.4 Selección de variables	42
1.4.1 Métodos híbridos.....	45
1.4.2 SVM-RFE.....	46
1.4.3 MKL con selección de variables	47
1.5 Fusión de datos	48
1.6 Imagen de electroforesis bidimensional	49

2. ESTADO DE LA CUESTIÓN	53
2.1 Análisis de texturas en imagen médica	53
2.2 Selección de variables en análisis de texturas	55
2.3 Consideraciones	57
3. HIPÓTESIS DE TRABAJO	59
4. PROPUESTA PARA EL ANÁLISIS BASADO EN TEXTURAS EN IMAGEN BIOMÉDICA	61
4.1 Formalización de las fases del diseño experimental	63
4.1.1 Extracción de datos	64
4.1.2 Preprocesado	65
4.1.2.1 Preparación de los datos	65
4.1.2.2 Normalización	66
4.1.3 Aprendizaje	67
4.1.3.1 Entrenamiento del modelo	67
4.1.3.2 Reducción de la dimensionalidad	69
4.1.3.3 Estimación del error real del modelo	70
4.1.4 Selección del mejor modelo	71
4.2 Mejoras propuestas en el proceso de análisis basado en texturas	76
4.2.1 Desarrollo de nuevos kernels	77
4.2.2 Métricas para la selección de variables	81
5. PRUEBAS Y RESULTADOS	87
5.1 Obtención de modelos de referencia	90
5.1.1 Extracción de datos	90
5.1.1.1 Variables de textura basadas en el histograma	94
5.1.1.2 Variables de textura basadas en el gradiente	94
5.1.1.3 Variables de textura basadas en la matriz de longitud de carrera	94
5.1.1.4 Variables de textura basadas en la matriz de coocurrencia	95
5.1.1.5 Variables de textura basadas en modelos autoregresivos	96
5.1.1.6 Variables de textura basadas en Wavelet	96
5.1.2 Preprocesado	97

5.1.3	Aprendizaje	97
5.1.3.1	Técnicas basadas en kernel	98
5.1.3.1.1	SVM	98
5.1.3.1.2	MKL	99
5.1.3.2	Reducción de dimensionalidad.....	101
5.1.3.2.1	Algoritmo Genético	101
5.1.3.2.2	Optimización por enjambre de partículas.....	103
5.1.3.2.3	SVM-RFE	104
5.1.3.2.4	FSMKL.....	105
5.1.4	Selección del mejor modelo.....	107
5.2	Resultados obtenidos con las mejoras propuestas.....	116
5.2.1	Aprendizaje	116
5.2.1.1	Nuevos Kernels aplicados en el sistema propuesto.....	117
5.2.1.2	Nuevo kernel suma de Gaussiano y Lineal	117
5.2.1.3	Nuevo kernel producto de Gaussiano y Lineal	119
5.2.2	Kernels propuestos: Selección del mejor modelo.....	120
5.2.3	Aprendizaje	127
5.2.3.1	Nuevas métricas de selección de variables	127
5.2.3.1.1	Distancia de Fisher	127
5.2.3.1.2	Distancia de Bhattacharyya.....	129
5.2.3.1.3	Distancia de Wilcoxon	130
5.2.3.1.4	Distancia de ROC	132
5.2.3.1.5	Distancia Entropía-Relativa	133
5.2.3.1.6	Distancia Entropía-Relativa transformada	134
5.2.4	Nuevas métricas: Selección del mejor modelo	135
5.2.5	Kernel Mahalanobis	144
5.3	Discusión.....	146
6.	CONCLUSIONES	149
7.	CONCLUSIONS.....	151
8.	FUTUROS DESARROLLOS	153
	REFERENCIAS.....	155

ANEXOS	171
ANEXO A: Variables de textura basadas en la matriz de coocurrencia	171
ANEXO B: Análisis con curvas ROC.....	173
ANEXO C: Synthetic Minority Over-Sampling Technique (SMOTE)	177
ÍNDICE GENERAL DE PALABRAS SIGNIFICATIVAS.....	179

TABLAS

TABLA 1. OPERADORES ESTADÍSTICOS DE PRIMER, SEGUNDO Y ORDEN SUPERIOR.....	21
TABLA 2. TESTS ESTADÍSTICOS MÁS UTILIZADOS DEPENDIENDO DEL NÚMERO DE VARIABLES.....	72
TABLA 3. VARIABILIDAD INTER- E INTRA-OPERADOR DEL PROCESO DE MARCADO DE PROTEÍNAS.....	91
TABLA 4. VARIABLES DE TEXTURA EXTRAÍDAS CON MAZDA.....	93
TABLA 5. SVM: DOMINIO DE LOS PARÁMETROS.....	98
TABLA 6. SVM: RESULTADOS DE LOS 10 EXPERIMENTOS.....	98
TABLA 7. MKL: DOMINIO DE LOS PARÁMETROS.....	100
TABLA 8. MKL: RESULTADOS DE LOS 10 EXPERIMENTOS.....	100
TABLA 9. GA: DOMINIO DE LOS PARÁMETROS.....	101
TABLA 10. GA: RESULTADOS DE LOS 10 EXPERIMENTOS.....	102
TABLA 11. PSO: RESULTADOS DE LOS 10 EXPERIMENTOS.....	103
TABLA 12. SVM-RFE: RESULTADOS DE LOS 10 EXPERIMENTOS.....	105
TABLA 13. FSMKL: DOMINIO DE LOS PARÁMETROS.....	106
TABLA 14. FSMKL: RESULTADOS DE LOS 10 EXPERIMENTOS.....	106
TABLA 15. MODELOS DE REFERENCIA: RESULTADOS.....	110
TABLA 16. TEST FRIEDMAN ALINEADO: RANKING DE LOS ALGORITMOS.....	112
TABLA 17. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL.....	112
TABLA 18. FSMKL: VARIABLES FINALES.....	115
TABLA 19. DOMINIO DE LOS PARÁMETROS UTILIZADOS CON FSMKL CON NUEVO KERNEL SUMA	117
TABLA 20. FSMKL-SUMA: RESULTADOS DE LOS 10 EXPERIMENTOS.....	118
TABLA 21. FSMKL-PRODUCTO: RESULTADOS DE LOS 10 EXPERIMENTOS.....	119
TABLA 22. NUEVOS KERNELS PROPUESTOS: RESULTADOS.....	122
TABLA 23. TEST NO PARAMÉTRICO QUADE: RANKING DE LOS ALGORITMOS.....	123
TABLA 24. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL.....	124
TABLA 25. FSMKL-PRODUCTO: VARIABLES FINALES SELECCIONADAS.....	127
TABLA 26. DISTANCIA FISHER: RESULTADOS DE LOS 10 EXPERIMENTOS.....	128
TABLA 27. DISTANCIA BHATTACHARYYA: RESULTADOS DE LOS 10 EXPERIMENTOS.....	129
TABLA 28. DISTANCIA WILCOXON: RESULTADOS DE LOS 10 EXPERIMENTOS.....	130
TABLA 29. DISTANCIA WILCOXON: VARIABLES SELECCIONADAS.....	131
TABLA 30. DISTANCIA ROC: RESULTADOS DE LOS 10 EXPERIMENTOS.....	132
TABLA 31. DISTANCIA ENTROPÍA RELATIVA: RESULTADOS DE LOS 10 EXPERIMENTOS.....	133
TABLA 32. DISTANCIA ENTROPÍA RELATIVA TRANSFORMADA: RESULTADOS DE LOS 10 EXPERIMENTOS.....	134
TABLA 33. NUEVAS DISTANCIAS PROPUESTAS: RESULTADOS.....	138
TABLA 34. TEST NO PARAMÉTRICO DE FRIEDMAN ALINEADO: RANKING DE LOS ALGORITMOS	140

TABLA 35. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL.....	140
TABLA 36. DISTANCIA BHATTACHARYYA: VARIABLES FINALES.....	144
TABLA 37. KERNEL MAHALANOBIS PARSIMONIOSO: RESULTADOS DE LOS 10 EXPERIMENTOS	145
TABLA 38. MATRIZ DE CONFUSIÓN PARA UN PROBLEMA DE CLASIFICACIÓN BINARIO.....	173

ILUSTRACIONES

FIGURA 1. TEXTURAS EN EL MUNDO REAL	16
FIGURA 2. VECINDARIO DE UN PIXEL CON DISTANCIA DE SEPARACIÓN IGUAL A UNO	23
FIGURA 3. ESQUEMA DE FUNCIONAMIENTO GENERAL DE UN GA	27
FIGURA 4. CONCEPTO DE MOVIMIENTO EN UN PSO DE ACUERDO A LAS MEJORES POSICIONES GLOBALES Y DE LA PROPIA PARTÍCULA	29
FIGURA 5. ESQUEMA DE FUNCIONAMIENTO GENERAL DE UN PSO.....	30
FIGURA 6. HIPERPLANO DE ÓPTIMO DE SEPARACIÓN, MÁRGENES MÁXIMOS ENTRE CLASES. PROBLEMA LINEALMENTE SEPARABLE	31
FIGURA 7. LÍNEA DE SEPARACIÓN ENTRE CLASES	33
FIGURA 8. VECTORES SOPORTE	35
FIGURA 9. DATOS NO SEPARABLES LINEALMENTE	35
FIGURA 10. VECTORES SOPORTE Y VALORES DE DESVIACIÓN SOBRE EL MARGEN DE LOS ELEMENTOS MAL CLASIFICADOS. PROBLEMA NO LINEALMENTE SEPARABLE	37
FIGURA 11. EL ESPACIO DE ENTRADA ES MAPEADO A UN ESPACIO DE DIMENSIÓN SUPERIOR DONDE LOS DATOS SON SEPARABLES	39
FIGURA 12. MKL	40
FIGURA 13. MKL PARA FUSIÓN DE DATOS CON DIFERENTES DATASETS HETEROGÉNEOS.....	41
FIGURA 14. TÉCNICAS DE SELECCIÓN DE VARIABLES	42
FIGURA 15. TÉCNICAS DE SELECCIÓN DE VARIABLES <i>FILTER</i>	43
FIGURA 16. TÉCNICAS DE SELECCIÓN DE VARIABLES <i>WRAPPER</i>	44
FIGURA 17. TÉCNICAS HÍBRIDAS BIO-INSPIRADAS CON UN CLASIFICADOR (SVM) PARA SELECCIÓN DE VARIABLES	46
FIGURA 18. SVM-RFE: ELIMINACIÓN DE VARIABLES.....	47
FIGURA 19. IMAGEN DE ELECTROFORESIS BIDIMENSIONAL	51
FIGURA 20. FASES DEL DISEÑO EXPERIMENTAL PROPUESTO.....	64
FIGURA 21. MÉTODO DE VALIDACIÓN CRUZADA	68
FIGURA 22. TEST PARAMÉTRICO: DIAGRAMA DE FLUJO	73
FIGURA 23. DIAGRAMA DE FLUJO DE UN ANÁLISIS DE VARIANZA	74
FIGURA 24. PROPAGACIÓN DE LAS MEJORAS ENTRE TÉCNICAS QUE EMPLEAN SVM	77
FIGURA 25. FASE DEL DISEÑO EXPERIMENTAL EN LA QUE SE PROPONE LA INCLUSIÓN DE NUEVOS KERNELS	77
FIGURA 26. FASE DEL DISEÑO EXPERIMENTAL EN LA QUE SE PROPONE LA INCLUSIÓN DE NUEVAS MÉTRICAS	82
FIGURA 27. MEJORA PROPUESTA EN EL PROCESO DE SELECCIÓN DE VARIABLES	82
FIGURA 28. REPRESENTACIÓN IDEAL DE UNA PROTEÍNA CON FORMA DE CAMPANA DE GAUSS.92	
FIGURA 29. ASPECTO REAL DE UNA PROTEÍNA RESULTADO DE UN PROCESO DE ELECTROFORESIS BIDIMENSIONAL	93
FIGURA 30. SVM: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	99

FIGURA 31. MKL: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	101
FIGURA 32. GA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.	103
FIGURA 33. PSO: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	104
FIGURA 34. SVM-RFE: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	105
FIGURA 35. FSMKL: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	107
FIGURA 36. CURVAS ROC DE LOS MODELOS DE REFERENCIA.....	108
FIGURA 37. MODELOS DE REFERENCIA: RESULTADOS	109
FIGURA 38. GRÁFICA QQ: MODELOS DE REFERENCIA.....	111
FIGURA 39. TEST DE FRIEDMAN ALINEADO: MODELOS DE REFERENCIA.....	113
FIGURA 40. FSMKL: IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA	114
FIGURA 41. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHS KERNEL EN LA SOLUCIÓN FINAL.....	115
FIGURA 42. FSMKL-SUMA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS	118
FIGURA 43. FSMKL-PRODUCTO: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.	120
FIGURA 44. NUEVOS KERNELS: CURVAS ROC	121
FIGURA 45. NUEVOS KERNELS PROPUESTOS: RESULTADOS	122
FIGURA 46. GRÁFICA QQ: NUEVOS KERNELS	123
FIGURA 47. TEST NO PARAMÉTRICO QUADE: NUEVOS KERNELS	124
FIGURA 48. IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA.....	125
FIGURA 49. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHS KERNEL EN LA SOLUCIÓN FINAL.....	126
FIGURA 50. DISTANCIA FISHER: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.	129
FIGURA 51. DISTANCIA BHATTACHARYYA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	130
FIGURA 52. DISTANCIA WILCOXON: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS	131
FIGURA 53. DISTANCIA ROC: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS	133
FIGURA 54. DISTANCIA ENTROPÍA RELATIVA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	134
FIGURA 55. ENTROPÍA RELATIVA TRANSFORMADA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.....	135
FIGURA 56. CURVAS ROC DE LOS MODELOS CON LAS DISTINTAS DISTANCIAS PROPUESTAS	136
FIGURA 57. NUEVAS MEDIDAS DE DISTANCIA: RESULTADOS.....	137
FIGURA 58. GRÁFICO QQ: NUEVAS DISTANCIAS PROPUESTAS	139
FIGURA 59. TEST NO PARAMÉTRICO DE FRIEDMAN ALINEADO: NUEVAS DISTANCIAS PROPUESTAS.....	141

FIGURA 60. IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA.....	142
FIGURA 61. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHS KERNEL EN LA SOLUCIÓN FINAL.....	143
FIGURA 62. IDENTIFICACIÓN DE PUNTOS IMPORTANTES EN UNA CURVA ROC	175
FIGURA 63. PUNTOS INDIVIDUALES CORRESPONDIENTES A CLASIFICADORES DISCRETOS	176
FIGURA 64. SMOTE: GENERACIÓN DE NUEVOS EJEMPLOS SINTÉTICOS SEGÚN LOS VECINOS PRÓXIMOS	178

LISTA DE ABREVIATURAS

2D-gel	Two Dimensional Gel Electrophoresis
AUROC	Area Under the ROC curve
FSMKL	Feature Selection Multiple Kernel Learning
GA	Genetic Algorithm
GLCM	Gray-Level Co-occurrence Matrices
i. e.	Id est (lat.: 'esto es')
MAS	Momento Angular Segundo
MDI	Momento Diferencial Inverso
MKL	Multiple Kernel Learning
post-hoc	(lat.: 'después de')
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
ROC	Receiver Operating Characteristics
SMOTE	Synthetic Minority Over-Sampling Technique
SVM	Support Vector Machines
SVM-RFE	Support Vector Machines - Recursive Feature Elimination

ESTRUCTURA DE LA TESIS

Esta Tesis Doctoral está estructurada en forma de capítulos, de forma que partiendo de una introducción de las técnicas y métodos que existen en la actualidad para realizar un proceso de selección de variables, sea posible entender el proceso seguido para el desarrollo de la misma

El presente capítulo servirá como presentación de la estructura que seguirá la presente Tesis Doctoral.

En el capítulo 1 de Fundamentos se presentará, partiendo del significado de textura y de los operadores estadísticos para el análisis de las mismas, las diferentes metodologías que existen para la selección de variables. Entre ellas se destacarán las metaheurísticas bioinspiradas y las técnicas basadas en *kernels*. Dentro de este capítulo se presentará una breve introducción a las técnicas basadas en *kernel*, y en concreto a las diferentes posibilidades existentes en la actualidad para realizar un proceso de selección de variables, por ser el eje central del desarrollo de esta Tesis Doctoral.

En el capítulo 2 de Estado de la cuestión, se repasará el estado actual en el análisis de texturas en imagen biomédica, así como las técnicas de selección de variables basadas en *kernel* para el análisis de texturas. Se razonará acerca de las características más importantes detectadas y además, se indicará el camino a seguir por esta Tesis Doctoral para no incurrir en los posibles errores o carencias de las aproximaciones previas. En base a lo expuesto en este capítulo, se presentará en el capítulo 3 la hipótesis de trabajo sobre la que se sustentará todo el trabajo desarrollado.

En el capítulo 4 se presenta la propuesta para el análisis de texturas en imagen biomédica. En este capítulo, se desgranará el diseño experimental seguido en esta Tesis Doctoral, así como el significado de cada una de las fases a desarrollar de forma pormenorizada y se concretarán las mejoras que se propondrán. Se verán en el capítulo 5 las pruebas desarrolladas y los resultados obtenidos siguiendo el diseño experimental y añadiendo las mejoras propuestas en el capítulo anterior.

En el capítulo 6 de Conclusiones se recogen, en base al trabajo realizado, las conclusiones que se pueden extraer, tanto de la aplicación concreta de la metodología propuesta como de los resultados obtenidos para los diferentes métodos de selección de variables basados en kernel utilizados. Asimismo se valorarán las mejoras propuestas sobre los métodos existentes.

Para terminar, en el capítulo 8 de Futuros Desarrollos, se indicarán las futuras líneas de trabajo que se seguirán a partir del desarrollo de esta Tesis Doctoral. Posteriormente se incluye la sección de Referencias donde se recogen las diferentes referencias bibliográficas empleadas en el desarrollo de la presente Tesis Doctoral. Se finaliza con una sección de Anexos en la que presentan las fórmulas matemáticas de las variables calculadas para la matriz de coocurrencia de los niveles de gris dada su importancia en la solución obtenida y para su consulta, un anexo sobre análisis de resultados con curvas ROC y de una técnica para balanceado de datos como SMOTE (*Synthetic Minority Over-Sampling Technique*).

1. FUNDAMENTOS

Se hará una pequeña introducción y se desarrollará a continuación durante el capítulo con más detenimiento las problemáticas relacionadas con un análisis de textura, un proceso de selección de variables, el uso de técnicas bioinspiradas y/o basadas en *kernel*, la fusión de datos y las imágenes de 2D-gel

Para esta Tesis Doctoral se definirá una textura como *la variación espacial de los valores de intensidad de gris a nivel de píxel*, aunque como se verá a continuación es un concepto complejo de definir, debido a que su uso está extendido en múltiples ámbitos, y cada uno de ellos proporciona su propia definición.

Las metaheurísticas bioinspiradas son técnicas de optimización basadas en comportamientos observados en la naturaleza que buscan encontrar el mínimo local de una función cuyo análisis resulte complicado de realizar por métodos convencionales. Para ello disponen habitualmente de una población de individuos que evalúan de forma paralela múltiples posibles soluciones y que durante el transcurso de sucesivas iteraciones buscan solucionar el problema.

Un *kernel* [11,12] básicamente, es una función que mapea un espacio de entrada en otro de una dimensionalidad mucho más alta, dónde se espera que los datos puedan ser linealmente separables. Sin embargo, la inclusión de estos *kernels* requiere de un nuevo nivel de parametrización, donde las funciones de *kernel* y sus parámetros deben ser seleccionados de forma mucho más cuidadosa.

Se puede definir la extracción de características de una imagen como la transformación o cambio de representación que se hace de ella. Este proceso suele hacerse de manera habitual pasando de los conocidos como dominios espacial y de frecuencia al dominio de las características, representando la información en diferentes vectores.

Esta transformación busca que las técnicas de visión artificial tengan la misma información que un ser humano al ver la misma escena, ya que éste último es capaz de interpretar y comprender escenas a pesar de la gran cantidad de información que puedan llegar a contener.

1.1 TEXTURAS

1.1.1 DEFINICIÓN DE TEXTURA

La textura está presente en todas las imágenes, ya sea en mayor o menor grado de importancia y es uno de los conceptos centrales en el área de la visión artificial. La textura, al contrario que el color o el brillo, es una propiedad espacial, por lo que un píxel aislado no tiene textura por sí mismo. A pesar de que es muy fácil encontrar ejemplos de textura en el mundo real como la que se puede ver en la Figura 1, es muy complicado dar una definición sencilla y precisa de la misma, ya que es algo sorprendentemente difícil de definir. Imagen tomada de *Datasets for Computer Vision Research*. URL: http://www-cvr.ai.uiuc.edu/ponce_grp/data/

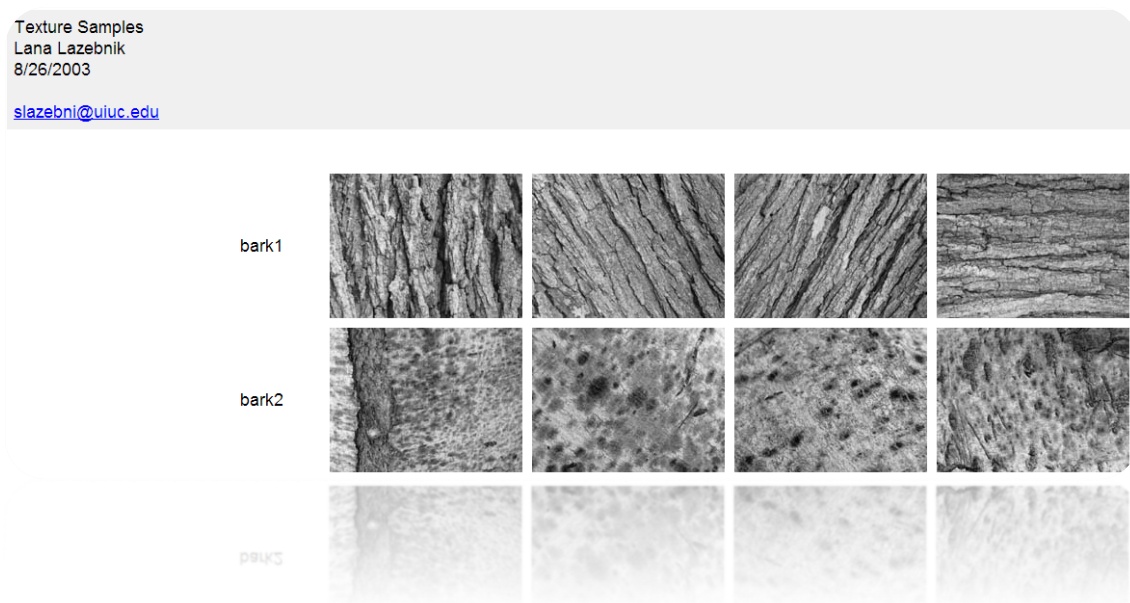


FIGURA 1. TEXTURAS EN EL MUNDO REAL.

En la comunidad científica no hay una única definición de textura aceptada por todos, aunque sea posible identificarla fácilmente al verla, la definición varía dependiendo del ámbito de aplicación [13]. Coggins [14] preparó, desde trabajos

previos publicados en la literatura, un catálogo completo de definiciones de textura. Es una de las características más importantes usadas para identificar objetos o regiones de interés en una imagen y está relacionada con la distribución espacial (estadística) de los niveles de gris que hay en la imagen [15]. Se puede consultar la colección de ejemplos de Brodatz [16] para ver múltiples ejemplos de texturas. Se incluirán a continuación alguna de las definiciones más aceptadas para el término:

- “Textura es un atributo que presenta la distribución espacial de los niveles de gris de una determinada región”. (Estándar IEEE 610.4)
- “Se entiende por textura un conjunto de características de una escena como son su rugosidad, homogeneidad, contraste, regularidad, direccionalidad o granulosidad, entre otras, las cuales se conocen como propiedades texturales”. [17]
- “Las texturas son patrones visuales complejos compuestos de entidades o subpatrones, que tienen similares características de brillo, color, forma, tamaño, etc.”. [18]
- “El término textura se refiere, en general, a la repetición de elementos básicos llamados texels. Cada texel contiene varios píxeles, cuya ubicación puede ser periódica, quasi-periódica o aleatoria. Las texturas naturales son generalmente aleatorias, mientras que las artificiales son a menudo deterministas o periódicas. Una textura puede ser gruesa, fina, suave, granulada, rugosa, regular, irregular, lineal, etc.”. [19]

1.1.2 ANÁLISIS DE TEXTURAS

El análisis de texturas no es nuevo en la literatura, de hecho, es un problema que ha sido investigado de manera intensiva y está relacionado con un gran conjunto de aplicaciones como el procesado de imagen biomédica.

La extracción de las variables es el primer paso de un análisis de texturas. Después de que esas variables sean extraídas, los resultados son usados para la segmentación, discriminación o determinación de formas dentro de la imagen. La percepción visual de las texturas fue estudiada por primera vez por Julesz a finales de los años setenta [20] evaluando la habilidad de los seres humanos para discriminar

determinados tipos de texturas. La conclusión final de su trabajo estableció a los operadores estadísticos de segundo orden como las variables de textura más importantes para objetivos de discriminación. De tal manera que, cuanto más diferentes fuesen los operadores de segundo orden entre dos imágenes, más fácil sería discriminarlas utilizando información textural. De acuerdo con Julesz, los ojos humanos no son capaces de utilizar los operadores de tercer orden o superior para discriminar texturas. Es por esto que, dependiendo del área de trabajo, una mala interpretación de la información presente en una imagen es crítica y se deberían usar aproximaciones basadas en diferentes estrategias para resolverlas. Un ejemplo de esta situación se presenta en el campo de la imagen médica donde se usa para reducir la variabilidad intra- e inter- operador en la identificación y segmentación de tumores para poder maximizar la zona de radiación en terapias radioterápicas (reduciendo la irradiación de zonas de tejido sano no afectado) [21].

Los aproximaciones de análisis de textura son normalmente caracterizadas en estructurales, estadísticas, basadas en modelos y en métodos de transformación [18].

- Las aproximaciones estructurales representan las texturas mediante el uso de primitivas y una jerarquía de relaciones espaciales entre esas primitivas [15,22] desde la micro- hasta la macrotextura. Esta aproximación es más útil para tareas de síntesis porque describe muy bien de manera simbólica la imagen. La textura se considera un patrón repetitivo que puede ser estudiado mediante un análisis de Fourier.
- Las aproximaciones estadísticas, al contrario, hacen uso de las relaciones y de la distribución de los niveles de gris entre píxeles en las imágenes, evitando las relaciones jerárquicas. La textura es considerada como la distribución de los niveles de gris de una imagen.
- Las aproximaciones basadas en modelos hacen uso de modelos fractales o estocásticos para interpretar las texturas y obtienen muy buenos resultados en el modelado de texturas naturales.
- Los métodos de transformación modifican el sistema de coordenadas de la imagen para encontrar uno en el que la representación de la imagen esté

relacionada con características de textura, como por ejemplo el dominio de frecuencia. Se hace utilizando las transformadas de Fourier, Gabor y Wavelet [23-28]. El problema de la transformada de Fourier está en la falta de localización espacial, mientras que la transformada Gabor tiene mejor localización espacial pero no hay una única resolución en el filtro para las texturas naturales donde se puedan encontrar estructuras espaciales, así que su uso es muy limitado. Comparado con los filtros Gabor, la transformada Wavelet mejora la resolución espacial y presenta muchas ventajas.

Weszka [29] comparó los métodos de transformación, estructurales y estadísticos de segundo orden, y llegó a la conclusión de que los estadísticos que se calculan para la relaciones entre pares de píxeles obtienen mejores resultados (segundo orden). Según Niemann [30] los estadísticos de orden superior a dos podrían ser los más importantes debido a que las texturas que difieren en los momentos de segundo orden son más fáciles de discriminar que los de orden superior. Existe pues una controversia en la que los resultados de esta Tesis aportarán su opinión al respecto. La matriz de coocurrencia de Haralick es el conjunto de variables estadísticas de segundo orden más popular [31], y de acuerdo con lo publicado en la literatura, (ver capítulo 2), ha demostrado un gran potencial discriminatorio cuando es usado en imagen biomédica.

1.1.2.1 OBJETIVOS DE UN ANÁLISIS DE TEXTURAS

Como se ha mencionado en el apartado anterior de esta Tesis, el análisis de textura ocupa un lugar muy importante en el área de la visión artificial ya que es una propiedad que se puede encontrar en toda imagen real y por lo tanto es un componente necesario para analizar.

Los problemas básicos [32] a los que aplicar un análisis de texturas son los siguientes: síntesis, clasificación, segmentación y discriminación de formas desde la textura.

- La síntesis de texturas es el problema que supone producir una nueva imagen de textura que se parezca a una textura dada, pero que no es exactamente igual.

En gráficos por ordenador, se puede dar el caso de que se quiera reproducir una textura sobre una superficie dada y darle un aspecto realista. Establece la forma en que se registran y son vistas las texturas por un observador, de forma que pueda diferenciarlas a partir de elementos estructurales, parámetros y distribuciones comunes y diferentes de las mismas.

- La clasificación de texturas es un problema importante en el análisis de texturas, donde el objetivo consiste en asignar una imagen a un conjunto de texturas previamente conocido. La clasificación de texturas ha sido usada en múltiples campos que van desde la detección de defectos en maderas hasta aplicaciones médicas para diagnóstico y clasificación de enfermedades. Una textura dada se clasifica como perteneciente a una clase determinada a partir del conocimiento a priori de un conjunto previo de texturas.
- La segmentación de texturas consiste en la partición de una imagen en regiones que poseen diferentes texturas, sin conocimiento de la diferenciación de las texturas y sin suponer ninguna hipótesis de su ubicación.
- La discriminación de formas desde la textura realiza el cálculo de características de una imagen digital que permite describir numéricamente sus propiedades de textura.

1.1.2.2 APROXIMACIÓN ESTADÍSTICA PARA EL ANÁLISIS DE TEXTURAS

Un análisis de texturas está basado en grandes conceptos matemáticos. La textura está caracterizada por las variaciones de las intensidades del nivel de gris en una imagen. Como se ha mencionado con anterioridad, no hay información de textura en un píxel aislado, de modo que la textura es una propiedad relacionada con el contexto de los niveles de gris del vecindario de un píxel. En la definición de textura juegan un papel muy importante propiedades comunes como: densidad, aspereza, uniformidad, rugosidad, regularidad, linealidad, direccionalidad, dirección, frecuencia y fase como propuso Laws [33]. Dichas propiedades están a menudo relacionadas [34]. En esta aproximación, la textura es descrita por una colección de operadores estadísticos que normalmente son clasificados entre operadores de primer, segundo y orden superior.

El histograma de la imagen es usado en las aproximaciones que utilizan operadores de primer orden para extraer información global de las texturas presentes en la imagen. Al tratarse de una medida global, no tiene en cuenta las posibles correlaciones entre píxeles. Por el contrario, en las aproximaciones que utilizan operadores estadísticos de segundo orden se utiliza la información que se obtiene de las relaciones entre píxeles en la imagen. No existe un consenso entre todos los investigadores que utilizan estos operadores debido a las diferentes áreas dónde son aplicados pero, con los más aceptados para cada grupo, se puede construir la Tabla 1.

TABLA 1. OPERADORES ESTADÍSTICOS DE PRIMER, SEGUNDO Y ORDEN SUPERIOR

Orden	Grupo	Variabes
Primero	Histograma	Media, varianza, skewness, kurtosis
Segundo	Matriz de coocurrencia	Segundo momento angular, contraste, correlación, suma de cuadrados, momento diferencial inverso, suma promedio, suma varianzas, suma entropías, entropía, diferencia varianzas, diferencia entropías
Superior	Matriz de longitud de carreras	Longitud de carrera no-uniforme, no-uniformidad de niveles de gris, énfasis de carrera larga, énfasis de carrera corta, fracción de imagen en carreras

Tsatsanis y Giannakis [35] demostraron que estos operadores son inmunes al ruido Gaussiano e insensibles a la translación de objetos con un detector de energía en el dominio acumulativo y así resolvieron la clasificación invariante de texturas y el modelado de problemas. Tsatsanis y Giannakis han obtenido resultados prometedores para resolver problemas de clasificación de texturas utilizando operadores de orden superior, sin embargo esta afirmación está en contradicción con lo dicho por Julesz [20] acerca de la poca información textural útil proporcionada por estos operadores.

1.1.2.2.1 OPERADORES DE PRIMER ORDEN

En este tipo de operadores, la textura se calcula realizando diferentes medidas sobre el histograma de la imagen (probabilidad de ocurrencia relativa a nivel de píxel del nivel de gris). No tiene en cuenta las posibles relaciones entre los píxeles de la imagen. Los

operadores conocidos como momento central se encuentran en este grupo: media, varianza, skewness y kurtosis [22,36,37], aunque también podrían haberse incluido otros como dispersión, valor cuadrático medio, energía, rugosidad o entropía.

La media calcula el promedio mientras que la varianza calcula la variación de intensidades alrededor de la media. *Skewness* es una medida de la simetría del histograma alrededor de la media (cero si es simétrico). La *kurtosis* es una medida de la forma del histograma que evalúa lo plano que es mientras que la entropía mide la uniformidad de la imagen.

El histograma de una imagen es sencillo de calcular e incluso aporta mucha de la información estadística presente en la imagen. Su principal inconveniente es que al estar basado en los valores de gris individuales de la imagen, no proporciona información acerca de la posición o de las relaciones de los píxeles. La ventaja clara de utilizar los operadores de primer orden es su simplicidad.

1.1.2.2.2 OPERADORES DE SEGUNDO ORDEN

Haralick fue el primero que propuso un *framework* para calcular los operadores de segundo orden y los valores de coocurrencia a nivel de píxel para obtener información textural y aplicarlos a las imagen de fotografía aérea [15]. En ese trabajo, propuso como calcular las matrices de coocurrencia y se convirtieron en las variables de textura estado-del-arte; como Julesz argumenta en [20], no se pueden discriminar texturas que sean iguales en sus operadores estadísticos de segundo orden.

Lo comentado en el párrafo anterior explica por qué la medida de distribución de probabilidad de niveles de gris para pares de píxeles es el operador más utilizado para realizar análisis de texturas. Las matrices de coocurrencia de los niveles de gris (GLCM) se construyen con las medidas de distribución de probabilidad de niveles de gris para pares de píxeles y miden la frecuencia con la que cada nivel de gris aparece en la imagen, en una posición fija con respecto a otras posiciones, como una función de los niveles de gris presentes en la imagen. El tamaño de la matriz se calcula de acuerdo al número umbral de niveles que se quieran calcular. Para cada posición relativa distinta entre dos píxeles que se quiera estudiar, se crea una matriz de coocurrencia

distinta. El análisis comienza con una distancia mínima de uno entre píxeles vecinos. En su trabajo [31], Haralick sugirió considerar como máximo solo dos píxeles de distancia y cuatro ángulos $\theta = 0^\circ, 45^\circ, 90^\circ$ y 135° debido a razones computacionales, ver Figura 2, hoy en día tanto por los equipos informáticos que procesan esta información como por la calidad de las imágenes que se analizan, esto se ve ampliamente superado.

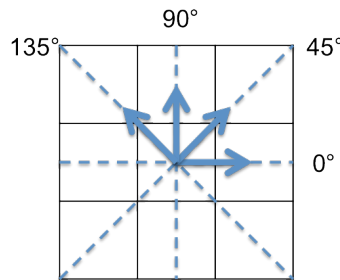


FIGURA 2. VECINDARIO DE UN PIXEL CON DISTANCIA DE SEPARACIÓN IGUAL A UNO

Cada píxel tiene un vecindario compuesto por 8 píxeles conectados a él. Extraer información de las matrices GLCM (son simétricas), permite obtener información acerca de la estructura de la textura, si se hace utilizando variables de textura. Haralick [15] propuso un total de 14 variables locales: segundo momento angular, contraste, correlación, suma de cuadrados, momento diferencial inverso, suma de medias, suma de varianzas, suma de entropías, entropía, diferencia de varianzas, diferencia de entropías y dos medidas con información de los coeficientes de correlación y de máxima correlación. Estas variables iniciales fueron aumentándose con el tiempo [38] aunque siguen predominando los operadores que utilizan las variables originales de Haralick.

1.1.2.2.3 OPERADORES DE ORDEN SUPERIOR

Como propuso Galloway [39], y después aplicaron Chu et al. [40], haciendo uso de los operadores de orden superior se calcula la matriz de longitud de carrera o *run-length matrix*. El método obtiene información del recorrido de un nivel de gris tomando una dirección en particular. De esta manera, el número de píxeles encontrado en el recorrido es conocido como longitud de carrera o *run-length*. Esto quiere decir que una textura rugosa, por ejemplo, estará dominada por recorridos largos y, por el contrario, una textura mucho más fina estará compuesta por un menor número de recorridos.

De manera similar a lo que ocurría con la GLCM, se utilizan cuatro matrices para describir los recorridos en los ángulos θ (0° , 45° , 90° y 135°) en píxeles adyacentes. Un conjunto de once medidas numéricas de textura son calculadas a partir de estos operadores: énfasis de carrera corta, énfasis de carrera larga, no-uniformidad de los niveles de gris, longitud de carrera para no uniformidad, porcentaje de carrera, énfasis de carrera de niveles bajos de gris, énfasis de carrera de niveles altos de gris, énfasis de longitud de carrera corta de niveles bajos de gris, énfasis de longitud de carrera corta de niveles altos de gris, énfasis de longitud de carrera larga de niveles de gris altos, énfasis de longitud de carrera larga de niveles altos de gris y énfasis de longitud de carrera larga de niveles de gris altos.

Los operadores de orden superior han demostrado ser capaces de resolver problemas no lineales y no Gaussianos gracias a su insensibilidad al ruido Gaussiano aditivo y a que son capaces de caracterizar mejor las señales no Gaussianas [41]. De acuerdo con estas propiedades, el uso de operadores de orden superior en un problema de clasificación de una señal no Gaussiana integrada en ruido Gaussiano está justificada por la posibilidad de obtener mejores resultados de clasificación. La razón principal fue apuntada en [42,43] desarrollando el Test de Gaussianidad de Hinich y basándose en el hecho de que muchas texturas naturales pueden ser modeladas como procesos no Gaussianos.

Es por esto que se considera que los operadores de orden superior pueden caracterizar un proceso aleatorio mejor que un operador de segundo orden, ya que permiten obtener información de fase y de la desviación de Gaussianidad. Para entender el aumento de importancia y relevancia de estos operadores [44-46] no se debe olvidar su insensibilidad e inmunidad al ruido Gaussiano.

No obstante, la controversia mencionada con estos operadores acerca de su utilidad en relación con los operadores estadísticos de primer y segundo orden ya empezó con la comparación que hacían los primeros trabajos que los consideraban menos eficientes [29,47] y termina con trabajos que los consideran igual al menos a los otros [48,49].

1.2 METAHEURÍSTICAS BIOINSPIRADAS

La creciente importancia de heurísticas basadas en la naturaleza o bioinspiradas ha generado grandes expectativas en el desarrollo de nuevas metaheurísticas [50] entre las que destacan los Algoritmos Genéticos y las técnicas de Optimización por Enjambre de Partículas entre otras.

Un proceso de optimización, o de minimización sin perder generalidad, trata de encontrar el mínimo global de una función analíticamente muy compleja de resolver, por ejemplo por tratarse de una función no convexa con múltiples mínimos locales y cuyo gradiente no pueda ser calculado fácilmente [51]. Los métodos de optimización habituales tratan de encontrar mediante procesos iterativos una solución y van mejorando hasta que encuentran un mínimo local. Es el caso de *hill climbing* o *simulated annealing*. Por el contrario, las heurísticas inspiradas en la naturaleza (habitualmente técnicas con una población de individuos) mediante comportamientos *inteligentes* evalúan de manera paralela múltiples soluciones en cada iteración, aumentando las posibilidades de no caer en un mínimo local.

En las secciones siguientes se ofrece una pequeña introducción a los Algoritmos Genéticos y la Optimización por Enjambre de Partículas.

1.2.1 ALGORITMOS GENÉTICOS

Los algoritmos genéticos, *genetic algorithms* (GA), son métodos adaptativos, usados de manera recurrente en la solución de problemas de búsqueda y en la optimización de parámetros. Están basados en los principios de la selección natural y la supervivencia del individuo más apto propuestos por Darwin [52] y sus bases computacionales fueron establecidos por [53]. Se encuentran bien descritos en varios textos [54-56].

Los GA usan una analogía con un comportamiento natural que se describe como la competición que se produce en la naturaleza entre individuos de una población por la búsqueda de recursos que les permitan sobrevivir. Incluso se podría incluir la competición entre individuos de una misma especie para encontrar un compañero. La probabilidad de generar un gran número de descendientes es, por tanto, directamente proporcional al éxito en sobrevivir y a la capacidad de procurarse

compañeros. Es fácilmente entendible que los genes de los individuos que más capacidad tienen de tener descendientes más se propagarán en sucesivas generaciones. De esta combinación de “buenos” genes puede a veces surgir un “superindividuo”, cuya adaptación es mucho mayor que la adaptación que haya podido tener cualquiera de sus ancestros. De esta manera, la evolución de la especie se encamina cada vez hacia una mejor adaptación del entorno en el que viven.

En base a esta analogía, los GA trabajan con una población de individuos, donde cada uno representa una posible solución al problema. La población inicial es generada de manera aleatoria. Cada posible solución puede ser representada mediante parámetros a los que se les dan valores. Este conjunto de parámetros es conocido como genes y se codifica en una cadena de valores que se denomina cromosoma o genotipo. Por tanto, el genotipo es el conjunto de parámetros representado por un cromosoma en particular y contiene toda la información necesaria para resolver el problema, que se conoce como fenotipo. Hablando en términos biológicos, el genotipo sería la información contenida en el ADN mientras que la expresión de ese ADN, correspondiente al mismo individuo, sería el fenotipo.

El alfabeto utilizado para representar los individuos puede ser binario, real, de números enteros o en punto flotante [57]. En los orígenes de los GA, se utilizó la codificación binaria valorándose el paralelismo implícito a este tipo de representación, pero ha de tenerse en cuenta que el resto de representaciones comentadas permiten especificar de modo más adecuado operadores genéticos necesarios en determinados campos de aplicación.

Para conocer el grado de satisfacción o bondad de la solución se le asigna a cada individuo una puntuación, esta medida recibe el nombre de ajuste o *fitness*. Se pueden diferenciar cuatro tipos de ajustes [58]: puro, estandarizado, ajustado y normalizado. Por lo general se usa el ajuste puro o estandarizado. Cuanto mejor sea este valor, mejor será la adaptación de este individuo para resolver el problema y mayor por tanto la probabilidad de que resulte seleccionado para reproducirse y poder así combinar su material genético con otro individuo (seleccionado del mismo modo) mediante los operadores de cruce. De esta manera y con el paso de las generaciones, se va

reemplazando la generación anterior y se verifica que la nueva generación de individuos, contiene una mayor proporción de buenas características que la anterior. Además, se realizarán mutaciones con cierta probabilidad en los cromosomas con el objetivo de introducir nueva información y evitar una convergencia prematura. En caso de haber sido correctamente diseñado, con el paso de las generaciones el algoritmo genético convergerá hacia una solución óptima del problema. Esto evidentemente no quiere decir que el GA sea capaz de encontrar la solución óptima del problema, aunque sí una solución aceptable en un tiempo razonable.

Pueden aplicarse diferentes operadores genéticos para la obtención de la nueva generación a partir de la predecesora, básicamente son el operador de cruce si se genera un nuevo descendiente a partir del mismo número de individuos de la generación anterior y operador de copia, si se permite el paso de un determinado número de individuos que no han sufrido ninguna variación. Una vez se ha generado esta nueva población, es el momento de que sea aplicado el operador de mutación, en el que con una probabilidad que suele estar comprendida en el rango 0.5-2 % se altera el entorno o vecindad de los individuos. Existen trabajos que tratan de determinar el valor de mutación más adecuado para cada problema [59,60], otros trabajos lo calculan en función de la longitud de la cadena y el número de individuos [61,62] e incluso existen trabajos en los que el valor de mutación varía con el tiempo [63]. Se puede observar el esquema general de funcionamiento en la Figura 3.

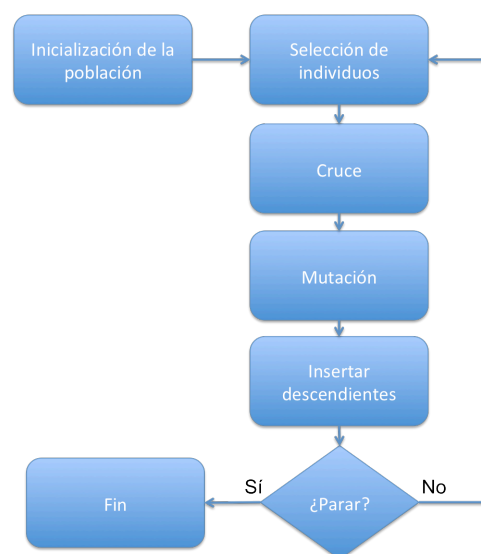


FIGURA 3. ESQUEMA DE FUNCIONAMIENTO GENERAL DE UN GA

De lo comentado anteriormente se puede inferir que el ámbito de trabajo de los GAs se centra en problemas para los cuáles no existen técnicas especializadas para resolverlos. Sin embargo, esto no quiere decir que no sean capaces de resolver problemas para los que existen técnicas específicas, pero sí es cierto que con una alta probabilidad lo harán de una manera más lenta y menos eficaz. Se puede hacer uso de aproximaciones híbridas en las que los GAs optimicen una de esas técnicas específicas.

Se admite que con el operador de cruce se realiza la búsqueda a lo largo del espacio de soluciones mientras que con el operador de mutación se consigue un componente de aleatoriedad que adquiere relevancia a medida que la población va convergiendo [64], dado que la población tiende a homogeneizarse.

1.2.2 OPTIMIZACIÓN POR ENJAMBRE DE PARTÍCULAS

El algoritmo de optimización por enjambre de partículas, o *Particle Swarm Optimization* (PSO) es un algoritmo de optimización bioinspirado basado en la simulación del comportamiento social de una bandada de pájaros. En los últimos 15 años PSO ha sido aplicado a gran variedad de problemas distintos [65] y se han presentado en la literatura científica numerosas variaciones del algoritmo [66].

Durante la ejecución de PSO, un conjunto de partículas se mueve dentro del dominio de la función buscando el óptimo de la función (mejor valor posible de bondad o *fitness*). El movimiento de cada partícula se consigue con las mejores posiciones visitadas por la propia partícula y por todo el enjambre (*gbest* PSO) o por un vecindario predefinido de la partícula (*pbest* PSO). En [67] utilizaron el concepto de *pbest* y *gbest* para separar entre aprendizaje cognitivo y social en el aprendizaje de cada partícula.

El concepto de movimiento puede verse en la Figura 4, imagen extraída de [68]. En base a esto se puede afirmar que cada partícula depende de una *inteligencia* individual y otra de enjambre y el movimiento. Ésta se puede describir con dos ecuaciones simples que regulan su posición y velocidad. Cada partícula representa una posible solución (posición en el espacio). Se representa la posición de una partícula i en la iteración t como $x_i^t = \{x_{i1}^t, x_{i2}^t, \dots, x_{in}^t\}$, siendo n el número de dimensiones [65,68]. De

manera análoga, se representa la velocidad de una partícula como $v_i^t = \{v_{i1}^t, v_{i2}^t, \dots, v_{in}^t\}$. Si se representa la mejor solución de una partícula i hasta la iteración t como $p_i^t = \{p_{i1}^t, p_{i2}^t, \dots, p_{in}^t\}$ y la mejor solución obtenida por p_i^t en la población durante la iteración t como $p_g^t = \{x_{g1}^t, x_{g2}^t, \dots, x_{gn}^t\}$ se puede calcular el cambio de la velocidad de la partícula como $V_{id}^t = V_{id}^{t-1} + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (p_{gd}^t - x_{id}^t)$, siendo $d = 1, 2, \dots, n$ y c_1 indica el factor de aprendizaje cognitivo; c_2 el factor de aprendizaje social; r_1 y r_2 son números aleatorios uniformemente distribuidos en el intervalo $U(0,1)$. Por lo tanto el movimiento de cada partícula queda definido según $x_{id}^{t+1} = x_{id}^t + V_{id}^t$.

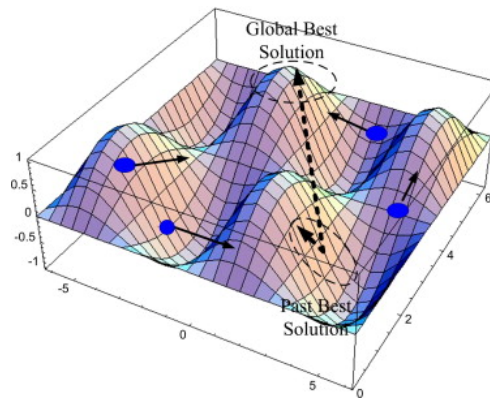


FIGURA 4. CONCEPTO DE MOVIMIENTO EN UN PSO DE ACUERDO A LAS MEJORES POSICIONES GLOBALES Y DE LA PROPIA PARTÍCULA

El esquema de funcionamiento general de un PSO puede verse en la Figura 5. Esta técnica tiene muchas similitudes con los GA en cuanto a que el sistema se inicializa con una población de individuos de manera aleatoria, a la postre posibles soluciones del problema, y busca a lo largo de distintas generaciones el óptimo global. La gran diferencia reside en que PSO, al contrario que GA, no tiene operadores de cruce ni de mutación.

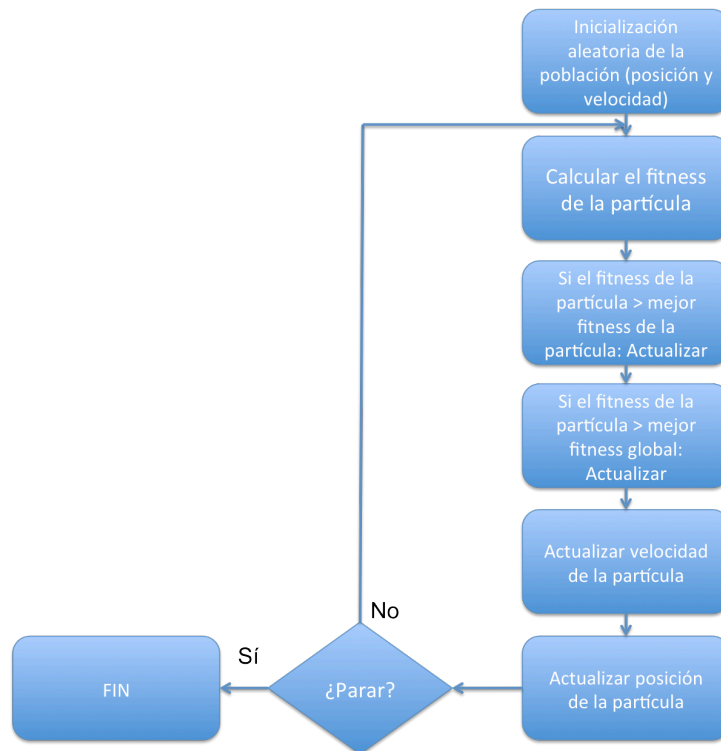


FIGURA 5. ESQUEMA DE FUNCIONAMIENTO GENERAL DE UN PSO

1.3 MÉTODOS BASADOS EN KERNELS

A continuación se dará una breve introducción a los métodos basados en *kernel* de las Máquinas de Vectores Soporte y de Aprendizaje por Múltiples Kernels. En caso de querer profundizar más lo aquí comentado se recomiendan las siguientes lecturas [12,69-74]. Sin embargo, es necesario introducir un mínimo de conceptos para disponer de la notación matemática suficiente como para que la explicación de los métodos utilizados en esta Tesis Doctoral tenga sentido para un lector no especializado en la materia.

1.3.1 MÁQUINAS DE VECTORES SOPORTE

Vapnik introdujo las Máquinas de Vectores Soporte, *Support Vector Machines* (SVM) a finales de los años setenta [75]. Son una de las técnicas más utilizadas por su buen funcionamiento y su capacidad de generalización en dominios de alta dimensionalidad, sobre todo en bioinformática [76]. En aprendizaje máquina se utilizan conjuntos de puntos en un determinado espacio para *aprender* una manera de tratar nuevas observaciones. Los métodos basados en *kernel* utilizan esos puntos para

saber como son de similares las nuevas observaciones y tomar una decisión. Los kernel codifican y miden la semejanza entre objetos [6,73,74].

La implementación trabaja con problemas de dos clases en la que los datos están separados por un *hiperplano*. Siendo n la dimensión de los datos, un hiperplano es un subespacio afín de dimensión $n-1$ que divide el espacio en dos mitades que se corresponde con las entradas de las dos clases [74]. En la tarea de clasificación, el objetivo del SVM es encontrar el hiperplano que separe los ejemplos positivos de los negativos. Este hiperplano separe los ejemplos positivos de los negativos orientado de tal manera que la distancia entre la frontera y el dato más cercano de cada clase sea máxima; los puntos más cercanos son utilizados para definir los márgenes, conocidos como vectores soporte [71]. Se puede ver el concepto de hiperplano óptimo (H_1) de separación en la Figura 6.

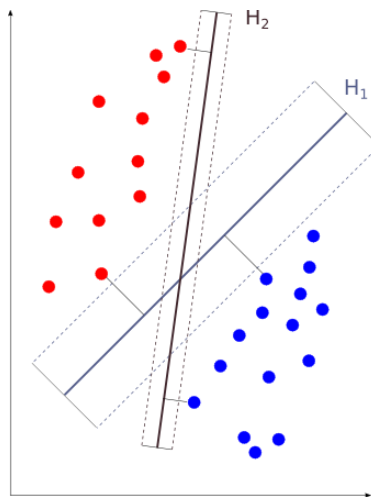


FIGURA 6. HIPERPLANO DE ÓPTIMO DE SEPARACIÓN, MÁRGENES MÁXIMOS ENTRE CLASES.
PROBLEMA LINEALMENTE SEPARABLE

Las técnicas de aprendizaje máquina, los métodos basados en *kernel* y los SVM más específicamente, han demostrado ser excepcionalmente eficientes en problemas de clasificación de alta dimensionalidad [2,77] debido a su habilidad para generalizar en dichos espacios, como el caso de un análisis de texturas.

La mayor parte de los conjuntos de datos complejos no son linealmente separables, por lo que los SVM introducen el concepto de *kernel*. Una función *kernel* es una función que mapea el espacio de entrada a una dimensión superior, donde los

datos puedan ser linealmente separables. Sin embargo, la inclusión de estos *kernels* requiere de un nuevo nivel de parametrización, donde la función del *kernel* y sus parámetros deben ser cuidadosamente seleccionados. Es entonces cuando surgen las técnicas de Aprendizaje de Múltiples Kernels, *Multiple Kernel Learning* (MKL), que se verán en la apartado 1.3.2, para seleccionar automáticamente estos parámetros.

1.3.1.1 HIPERPLANO DE SEPARACIÓN ÓPTIMO

Sea un problema de clasificación binario, donde se utilizan las etiquetas -1/+1 para diferenciar las dos clases. Siendo el par $\{x^t, r^t\}$ el conjunto ordenado de entradas y salidas deseadas donde $r^t = +1$ si $x^t \in C_1$ y $r^t = -1$ si $x^t \in C_2$, se pretende encontrar w y w_0 tal como se muestra en la (Eq. 1)

$$\text{(EQ. 1)} \quad \begin{aligned} w^T x^t + w_0 &\geq +1 \text{ para } r^t = +1 \\ w^T x^t + w_0 &\leq -1 \text{ para } r^t = -1 \end{aligned}$$

lo que puede ser reescrito como $r^t(w^T x^t + w_0) \geq +1$, siendo w el vector de pesos y w_0 el umbral.

No es necesario solamente que los casos se puedan situar adecuadamente a cada lado del hiperplano dependiendo de su clase de pertenencia, si no que han de hacerlo con la mayor distancia posible. Esta distancia es conocida como margen y el objetivo es maximizarlo, ya que cuanto mayor sea mejor será la capacidad de generalización del modelo. En la Figura 7 se puede ver como el hiperplano separa el espacio de entrada en dos regiones, una para cada una de las clases, de manera que si $g(x) > 0$ los puntos se sitúan a la derecha del hiperplano y viceversa.

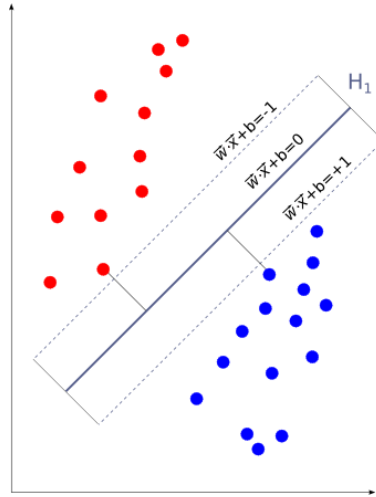


FIGURA 7. LÍNEA DE SEPARACIÓN ENTRE CLASES

Tomando dos puntos sobre el hiperplano de decisión $g(x_1)=g(x_2)=0$, entonces se puede escribir que $w^T x_1 + w_0 = w^T x_2 + w_0$, por lo que $w^T (x_1 - x_2) = 0$. Esto quiere decir que w es normal a cualquier vector que caiga sobre el hiperplano. Tomando x_p como la proyección normal de x en el hiperplano, siendo r la distancia de x al hiperplano, se puede reescribir x en (Eq. 2) como [78],

$$(EQ. 2) \quad x = x_p + r \frac{w}{\|w\|}$$

Se puede ver entonces que para el caso en que $g(x_p)=0$, la distancia al origen queda definida por la localización del hiperplano con respecto del origen w_0 y por la orientación w de la forma siguiente $r_0 = \frac{w_0}{\|w\|}$. Por lo tanto, maximizar el margen como se puede ver en (Eq. 3) es equivalente [79] a

$$(EQ. 3) \quad \text{minimizar } \frac{1}{2} \|w\|^2 \text{ sujeta a la restricción } r^t (w^t x^t + w_0) \geq +1, \text{ para } \forall t.$$

Se sabe que esto es un problema de optimización cuadrático que depende de la dimensionalidad del problema. Ya se ha dicho con anterioridad que la potencia de estos métodos en problemas que no son linealmente separables, reside en la posibilidad de mapear el espacio de entrada en un espacio de características de dimensión superior donde los datos puedan ser linealmente separables. Es por esto que no interesa que la complejidad dependa de la dimensionalidad y para ello, el objetivo es conseguir que la complejidad dependa del número de elementos en el

conjunto de entrenamiento. Además, esto permitirá reescribir las funciones en términos de funciones *kernel*.

Mediante los multiplicadores de Lagrange α^t , es posible reescribir el problema, ya sin restricciones de la forma indicada por la (Eq. 4).

$$(EQ. 4) \quad L(w, w_0) = \frac{1}{2} \|w\|^2 - \sum_{t=1}^N \alpha^t [r^t (w^T x^t + w_0) - 1].$$

Dado que se trata de un problema de optimización cuadrático convexo (tanto el término como las restricciones lineales son convexas), se pueden utilizar las condiciones de Karush-Kuhn-Tucker para resolver el problema dual equivalente. La dualidad consiste en maximizar $L(w, w_0)$ con respecto de α^t , con las restricciones de que el gradiente de $L(w, w_0)$ con respecto de w y de w_0 son cero y además $\alpha^t \geq 0$. Tomando las derivadas con respecto de w (Eq. 5) y de w_0 (Eq. 6) e igualándolas a 0, quedando

$$(EQ. 5) \quad \frac{\partial L(w, w_0)}{\partial w} = w - \sum_t \alpha^t r^t x^t = 0 \Rightarrow w = \sum_t \alpha^t r^t x^t$$

$$(EQ. 6) \quad \frac{\partial L(w, w_0)}{\partial w_0} = 0 \Rightarrow \sum_t \alpha^t r^t = 0$$

Si ahora se sustituye la (Eq. 5) y la (Eq. 6) en la (Eq. 4) obteniendo la (Eq. 7), que maximizando con respecto a α^t con la restricción $\sum_t \alpha^t r^t = 0$ y con $\alpha^t \geq 0$ para todo t , puede ser resuelto usando métodos de optimización cuadráticos, pero la complejidad ya depende del número de elementos (N) y no de la dimensionalidad del problema, estando la complejidad temporal entre $\mathcal{O}(N^2)$ y $\mathcal{O}(N^3)$.

$$(EQ. 7) \quad L(w, w_0) = -\frac{1}{2} \sum_t \sum_S \alpha^t \alpha^S r^t r^S (x^t)^T x^S + \sum_t \alpha^t$$

Una vez resuelta esta ecuación, al conjunto de x^t cuyos $\alpha^t \geq 0$ se le conoce como vectores soporte y cumplen que $r^t (w^T x^t + w_0) = 1$ y además están en el margen de separación. Esto puede servir para calcular w_0 como se ve en la (Eq. 8).

$$(EQ. 8) \quad w_0 = r^t - w^T x^t$$

Esto se hace con todos los vectores soporte y se obtiene una media, el discriminante que resulta es conocido como máquina de vectores soporte. Se pueden ver marcados los vectores soporte con círculos verdes en la Figura 8, donde se ilustra un ejemplo sencillo para dos clases en la que todos los elementos son linealmente separables. Si esto no fuese así, y no existiese un hiperplano que separase los casos, sería necesario suavizar el margen de separación y buscar el hiperplano que cometiese el menor error.

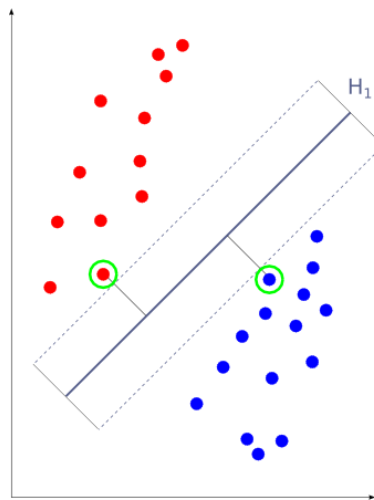


FIGURA 8. VECTORES SOPORTE

1.3.1.2 DATOS NO SEPARABLES LINEALMENTE

En la apartado anterior se presentó un clasificador que funciona para el caso de que los datos sean linealmente separables. En caso de que esto no sea así, como se muestra en la Figura 9, es necesario buscar un hiperplano que pueda separar los casos con el menor número de errores posibles. Para ello, es necesario considerar una nueva variable débil que se utilizará para calcular la desviación sobre el margen $\xi^t \geq 0$.

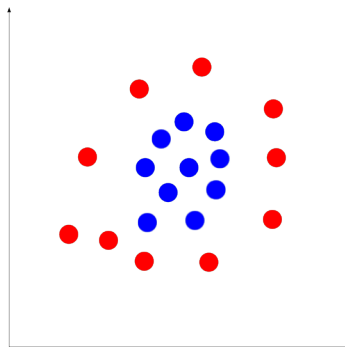


FIGURA 9. DATOS NO SEPARABLES LINEALMENTE

Obviamente se han de considerar dos posibles opciones para computar el error. Por un lado, se puede clasificar mal un ejemplo teniéndolo en el lado contrario del hiperplano, o se puede haber clasificado bien pero no estar lo suficientemente lejos de él. Se debe relajar la ecuación $r^t(w^T x^t + w_0) \geq +1$ añadiendo la variable débil ξ de manera que quedaría $r^t(w^T x^t + w_0) \geq +1 - \xi^t$, por tanto el número de errores de clasificación está determinado por el número $\xi^t > 1$.

Si se añade esta nueva variable en la (Eq. 4) quedaría

$$(EQ. 9) \quad L(w, w_0) = \frac{1}{2} \|w\|^2 + C \sum_t \xi^t - \sum_{t=1}^N \alpha^t [r^t(w^T x^t + w_0) - 1 + \xi^t] - \sum_t \mu^t \xi^t$$

donde C es el factor de penalización por los errores debidos a una mala clasificación y los ejemplos que se sitúan en el margen, para tratar de conseguir una mejor generalización y μ_t son los nuevos multiplicadores de Lagrange que garantizan que ξ^t tenga valor positivo. Tomando de nuevo las derivadas e igualándolas a 0, se puede escribir la (Eq. 10) de la forma

$$(EQ. 10) \quad \frac{\partial L(w, w_0)}{\partial \xi^t} = C - \alpha^t - \mu^t = 0.$$

Como además se sabe que $\mu^t \geq 0$ esto implica que $0 \leq \alpha^t \leq C$. Despejando en (Eq. 9) se obtiene la (Eq. 7) pero ahora sujeta a las nuevas restricciones $\sum_t \alpha^t r^t = 0$ y $0 \leq \alpha^t \leq C, \forall t$, lo que como ya se vio en el caso de los datos pudiesen ser linealmente separables, los casos que se encuentran correctamente clasificados tienen un $\alpha^t = 0$. Los vectores soporte con valor $\alpha^t > 0$ y $\alpha^t < C$ están en el margen, tienen $\xi^t = 0$ y cumplen $r^t(w^T x^t + w_0) = 1$. En la Figura 10 se muestran los vectores soporte y los valores de error por clasificación incorrecta en un problema no linealmente separable.

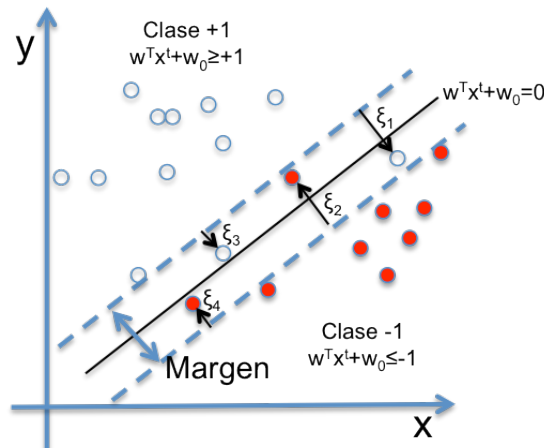


FIGURA 10. VECTORES SOPORTE Y VALORES DE DESVIACIÓN SOBRE EL MARGEN DE LOS ELEMENTOS MAL CLASIFICADOS. PROBLEMA NO LINEALMENTE SEPARABLE

1.3.1.3 TRUCO DEL KERNEL

En el caso en el que los datos no sean linealmente separables, una de las técnicas empleadas es el truco del kernel. La idea es muy sencilla y parte de lo mencionado en las dos secciones anteriores. Las máquinas de vectores soporte buscan el hiperplano que mejor separe los datos maximizando la capacidad de generalización del modelo. Si los datos no son linealmente separables, para intentar que lo sean, el espacio de entrada inicial se puede mapear en un espacio de mayor dimensionalidad [80] (recibe el nombre de espacio de características). En este nuevo espacio debe estar definido el producto escalar, espacio de Hilbert, tal y como se muestra en la Figura 11.

Todo producto vectorial en el espacio de entrada tiene su equivalente en el espacio de características según $K(x^t x^s) = \langle \phi(x^s), \phi(x^t) \rangle_{\mathcal{H}}$ y $d(x, y) = \|\phi(x) - \phi(y)\|_{\mathcal{H}}$, definidos en el espacio de Hilbert, siendo d la distancia y cumpliendo determinadas condiciones: el kernel debe ser semidefinido positivo, simétrico y cumplir con las condiciones de Mercer [81]. Si se cumplen, el kernel representa un producto escalar en el espacio de características o de Hilbert [12,82,83].

El modelo lineal que se genera en el nuevo espacio de características corresponde al modelo no lineal del espacio de entrada original, donde mapear un espacio x de n dimensiones en un espacio z de k dimensiones es lo mismo que

$$(EQ. 11) \quad L(w, w_0) = \frac{1}{2} \|w\|^2 + C \sum_t \xi^t$$

pero con las restricciones $r^t w^T \phi(x^t) \geq 1 - \xi^t$ de manera que ahora la (Eq. 9) pasa ser de la forma

$$(EQ. 12) \quad L(w, w_0) = \frac{1}{2} \|w\|^2 + C \sum_t \xi^t - \sum_t \alpha^t [r^t w^T \phi(x^t) - 1 + \xi^t] - \sum_t \mu^t \xi^t$$

Tomando ahora las derivadas con respecto a w vistas en (Eq. 5), (Eq. 6) y (Eq. 10), el problema dual se convierte en resolver

$$(EQ. 13) \quad L(w, w_0) = -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s \phi(x^t)^T \phi(x^s) + \sum_t \alpha^t$$

sujeto a la restricción $\sum_t \alpha^t r^t = 0$ con $0 \leq \alpha^t \leq C, \forall t$.

La idea es utilizar una función kernel, $K(x^t, x^s)$, para reemplazar el producto escalar $\phi(x^t)^T \phi(x^s)$ de tal manera que se mapea aplicando directamente la función de kernel en el espacio de entrada inicial, dejando (Eq. 13) de la forma

$$(EQ. 14) \quad L(w, w_0) = -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s K(x^t, x^s) + \sum_t \alpha^t.$$

Esto quiere decir que si una función de kernel está correctamente definida, es directo el uso de $K(x^t, x^s)$. Se obtiene así la matriz de Gram (semidefinida positiva y simétrica) con los valores de kernel.

Si el kernel es semidefinido positivo y simétrico tiene un espacio de variables asociado en el espacio Hilbert que permite encontrar el hiperplano en ese espacio [69,84,85].

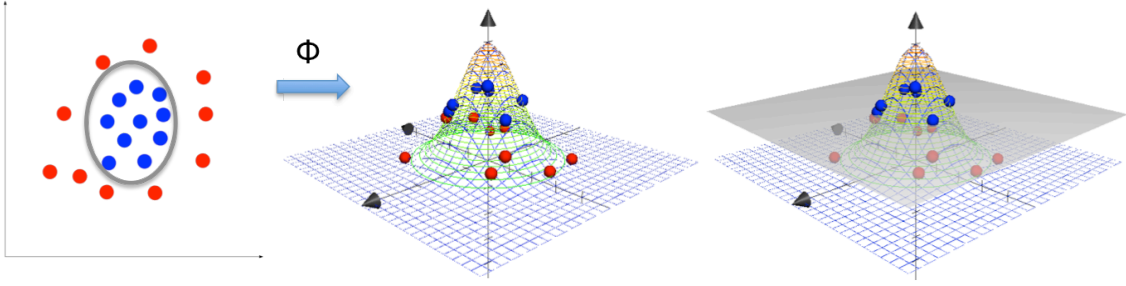


FIGURA 11. EL ESPACIO DE ENTRADA ES MAPEADO A UN ESPACIO DE DIMENSIÓN SUPERIOR DONDE LOS DATOS SON SEPARABLES

1.3.1.4 KERNELS MÁS HABITUALES DE PROPÓSITO GENERAL

Son muchos los *kernels* que pueden cumplir con el teorema de Mercer [81]. Una clasificación sencilla fue propuesta por Smola en su Tesis Doctoral, separándolos en *kernels* locales y globales [11]. Para los *kernels* locales solamente aquellos datos que están cerca de otros, tienen influencia en los valores del *kernel*, mientras que por el contrario, en los *kernels* globales, todos los puntos, por muy distantes que estén unos de otros, tienen influencia en los valores del *kernel*. Un ejemplo de *kernel* global podría ser el polinomial y de local, el *kernel* de función de base radial. Ambos se presentarán a continuación.

Los *kernels* más utilizados son, con diferencia, los lineales (Eq. 15), los polinomiales de grado q (Eq. 16), los función de base radial, *radial-basis functions* (RBF) (Eq. 17) y los *kernel* función sigmoideal (Eq. 18). Se presentan las 4 ecuaciones que los definen a continuación

$$(EQ. 15) \quad K(x^t, x) = (x^T x^t)$$

$$(EQ. 16) \quad K(x^t, x) = (x^T x^t + 1)^q$$

donde q es el grado del polinomio,

$$(EQ. 17) \quad K(x^t, x) = \exp \left[-\frac{\|x^t - x\|^2}{2\sigma^2} \right]$$

donde x^t es el centro y σ es el parámetro definido por el usuario para el radio y

$$(EQ. 18) \quad K(x^t, x) = \tanh (2x^T x^t + 1).$$

1.3.2 APRENDIZAJE CON MÚLTIPLES KERNELS

El aprendizaje por múltiples *kerneles* o *Multiple Kernel Learning* (MKL) fue desarrollado en la última década para resolver problemas de clasificación binarios. Esta técnica propuesta por Lanckriet et al. [86], es una aproximación intermedia a la fusión de datos [8] que combina kernels convexamente, de manera que preserva la estructura de los datos infiriendo un único modelo [87-89]. El funcionamiento básico de MKL consiste en aprender utilizando un conjunto de *kernels*. Esta combinación consiste en un conjunto de *kernels* distintos y múltiples parametrizaciones distintas de cada uno de esos *kernels* (ver Figura 12). De esta manera se consigue un framework para aprendizaje con múltiples conjuntos de datos distintos como sucede en un análisis de texturas [5,86]. Desde un único espacio de entrada, se generan diferentes espacios de características ϕ que luego tendrán su influencia λ en el kernel combinado.

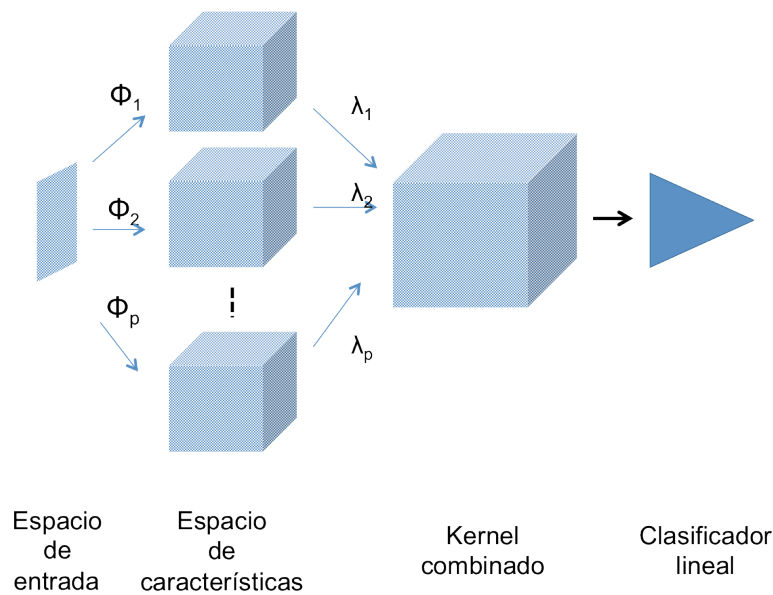


FIGURA 12. MKL

Es posible por tanto, construir nuevos kernels por combinación de otros simples ya sea multiplicando un kernel por un escalar, sumando dos kernels o multiplicando dos kernels sencillos. Estos kernel pueden o no haberse generado con el mismo subconjunto de datos de x . Se podrían tratar diferentes conjuntos de datos de entrada y aplicarles un kernel en concreto, tomando la suma de ambos sería como hacer el producto escalar de los vectores encadenados, y podría generalizarse a cualquier número de kernels, de la forma

$$(EQ. 19) \quad K(x, y) = \sum_{i=1}^m K_i(x, y).$$

Si como proponen [7,90] se tomase la suma ponderada, se podrían aprender los pesos de los datos según la siguiente ecuación

$$(EQ. 20) \quad K(x, y) = \sum_{i=1}^m \eta_i K_i(x, y)$$

sujeto a la restricción de que el peso del kernels $\eta_i \geq 0$. Es en concreto esta sustitución por el peso ponderado de los *kernels* sencillos lo que se conoce como MKL. Es necesario ahora modificar la (Eq. 14) de forma que queda

$$(EQ. 21) \quad L(w, w_0) = -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s \sum_i \eta_i K_i(x^t, x^s) + \sum_t \alpha^t.$$

Al final, dependiendo de la importancia del kernel y de la influencia que ha ejercido para la solución final, el parámetro de peso tendrá un valor más elevado. Si el valor de peso del *kernel* es 0, quiere decir que las variables no son relevantes o la información que contiene puede estar implícita en otros datos.

En el caso de usar esta técnica para fusión de datos heterogéneos, el resultado se muestra en la Figura 13.

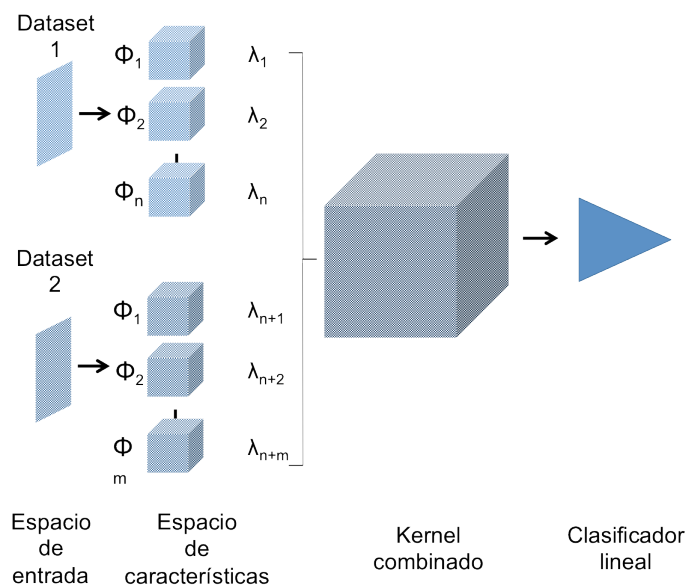


FIGURA 13. MKL PARA FUSIÓN DE DATOS CON DIFERENTES DATASETS HETEROGÉNEOS

Un buen tutorial acerca del uso de SVM y MKL puede encontrarse en [6,12].

1.4 SELECCIÓN DE VARIABLES

Uno de los mayores problemas que aparece en un análisis de texturas es su alta dimensionalidad y la diversidad de los datos que se calculan puesto que pertenecen a grupos de variables de textura muy dispares. Un análisis de textura calcula las variables de una región concreta de la imagen. Algunas de estas variables pueden estar muy correlacionadas y contener, por tanto, información irrelevante. Las técnicas de selección de características surgen para abordar estos problemas, puede verse un esquema general de funcionamiento en la Figura 14. El rendimiento de un algoritmo supervisado puede verse afectado por el número y la relevancia de las variables de entrada. El objetivo de estas técnicas es encontrar el subconjunto de variables de entrada que describa la estructura de los datos de igual manera [91-93] o mejor [94] que los datos originales.

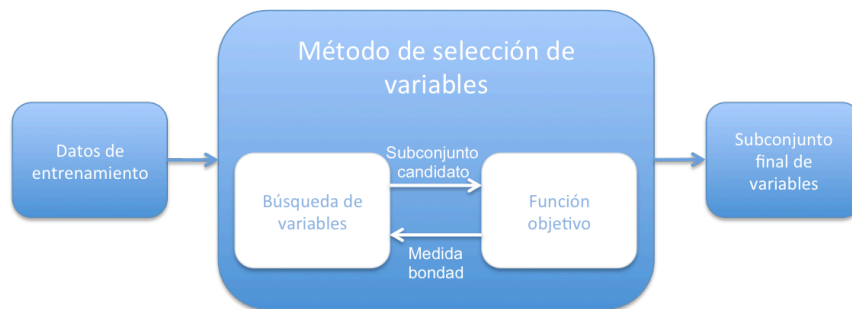


FIGURA 14. TÉCNICAS DE SELECCIÓN DE VARIABLES

Una posible clasificación [5,95-98] de técnicas de selección de variables sería la que las divide en métodos *filter*, *wrapper* o *embedded*.

A grandes rasgos, los métodos *filter* tienen en cuenta la relevancia particular de cada variable en el conjunto de datos n estableciendo un ranking por orden de relevancia [99]. De este modo, las variables con la peor puntuación son eliminadas del conjunto de variables creando un nuevo espacio de variables de entrada k para el clasificador, siendo $k < n$ en cuanto a la dimensionalidad de ambos. Puede verse un esquema de funcionamiento de este tipo de técnicas en la Figura 15. De acuerdo con [95,100] son ejemplos de medidas válidas para establecer el ranking de las variables,

entre otros los siguientes: criterio lambda de Wilks, análisis de componentes principales, técnicas de información mutua, los mapas auto-organizativos o medidas de distancia como: Student's T-test, Wilcoxon sum-rank test, Bhattacharyya o Entropía Relativa.

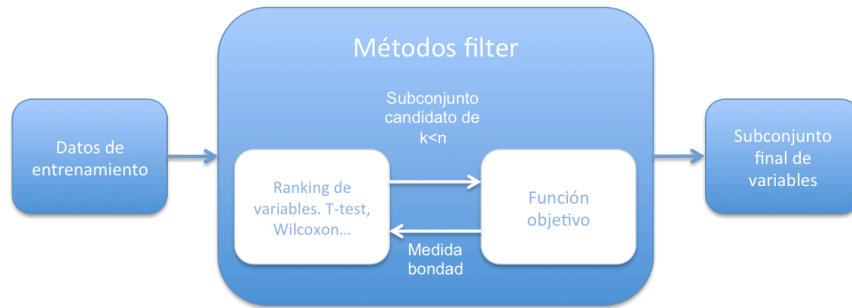


FIGURA 15. TÉCNICAS DE SELECCIÓN DE VARIABLES FILTER

Los métodos *wrapper* [97] evalúan un subconjunto de variables para resolver el problema de clasificación y usan el valor de rendimiento obtenido. De este modo, los subconjuntos son evaluados de acuerdo con el error de clasificación, en el sentido de que cuanto más bajo sea, mejor es el clasificador con esas variables para resolver el problema. Un esquema general de funcionamiento de las técnicas de selección de variables *wrapper* puede verse en la Figura 16. Existen también otro conjunto de técnicas que llevan a cabo una búsqueda local en el espacio de variables, llamadas estrategias *greedy*. Estos algoritmos deterministas incluyen técnicas de selección hacia adelante y hacia atrás. En las técnicas de selección hacia adelante, por ejemplo, el proceso de selección de variables comienza con un conjunto vacío de variables y de acuerdo con el valor de error de clasificación, se van añadiendo nuevas variables. Por el contrario, las técnicas de selección hacia atrás comienzan el proceso iterativo con todas las variables y van eliminando de manera iterativa. Son ejemplo de estos métodos las técnicas basadas en *kernel* como los GA o PSO en combinación con SVM [53,101,102] así como el algoritmo de enfriamiento simulado, *Simulated Annealing* (SA).

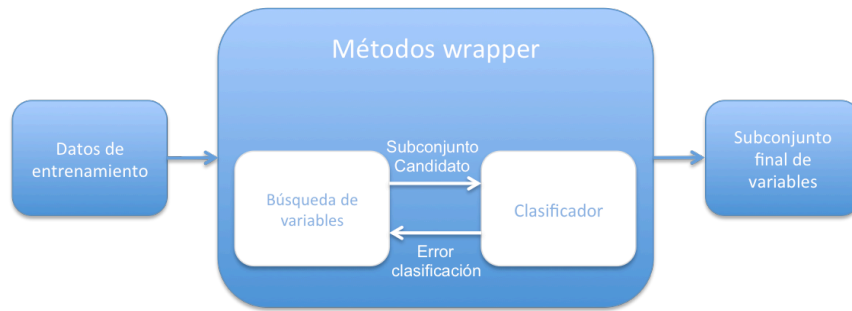


FIGURA 16. TÉCNICAS DE SELECCIÓN DE VARIABLES WRAPPER

La tercera gran categoría son los métodos *embedded*. Estos métodos combinan a la vez el proceso de entrenamiento con la búsqueda en el espacio de variables. Los métodos *nested* son un caso particular de esta categoría donde el proceso de búsqueda es guiado estimando los cambios del rendimiento del clasificador para diferentes subconjuntos en la función objetivo [98]. La técnica conocida como Máquinas de Vectores Soporte-Eliminación Recursiva de Variables, *Support Vector Machines – Recursive Feature Elimination* (SVM-RFE) está incluida en esta categoría.

Las técnicas de búsqueda del mejor subconjunto de variables descartan las variables que menos contribuyen a obtener un modelo que mejore el rendimiento del que utiliza todas las dimensiones. Partiendo de un conjunto de dimensionalidad n , existen 2^n subconjuntos posibles de variables que no pueden ser evaluados todos a menos que n sea pequeño. Esta es la razón por la que se utilizan heurísticas que obtengan la mejor solución (no necesariamente la óptima) en un tiempo razonable (polinomial).

Las técnicas *filter* son más rápidas y muy generales en el sentido de que no son dependientes de ningún clasificador, aunque tienden a seleccionar grandes subconjuntos de características. Por otro lado las técnicas *wrapper* tienen un coste en tiempo asociado al entrenamiento del clasificador y valores de rendimiento dependientes del mismo. A pesar de que se obtiene un buen rendimiento de clasificación y si se emplea de manera conjunta con mecanismos que eviten el sobreajuste, generalizan bien.

1.4.1 MÉTODOS HÍBRIDOS

Se conoce como métodos híbridos a aquellos que surgen de la combinación de diferentes técnicas. Esta Tesis va a considerar el uso conjunto de técnicas de optimización como las presentadas en el capítulo 1.2 con SVM [73] para la selección de variables en problemas de alta dimensionalidad (aproximación *wrapper*). Múltiples estudios se han llevado a cabo al respecto, y consideran que estas técnicas de optimización son las adecuadas para encontrar soluciones óptimas en problemas con más de 40 variables [103]. Estas técnicas bioinspiradas optimizan entonces el rendimiento del clasificador, en concreto son varios los trabajos que encuentran muy acertada la combinación de los mismos con técnicas basadas en kernel [104-113], en campos muy diversos, algunos de ellos en análisis de texturas.

Cada individuo de las técnicas bioinspiradas podría tener por lo tanto una longitud igual al número de variables a utilizar y el proceso de selección de características estaría guiado por una medida de acierto de cada solución. Los individuos vendrían expresados en codificación binaria, de manera que si se debe utilizar la variable para clasificar aparecerá con un uno, y si no la debe de utilizar aparecerá un cero.

Como se puede apreciar en la Figura 17, en una aproximación *wrapper* cada posible subconjunto de variables es evaluado en función de una medida de acierto (se discuten en el capítulo 4.1.3). Esto implica un alto coste computacional dependiendo del problema, por lo que no es aplicable en cualquier entorno [114].

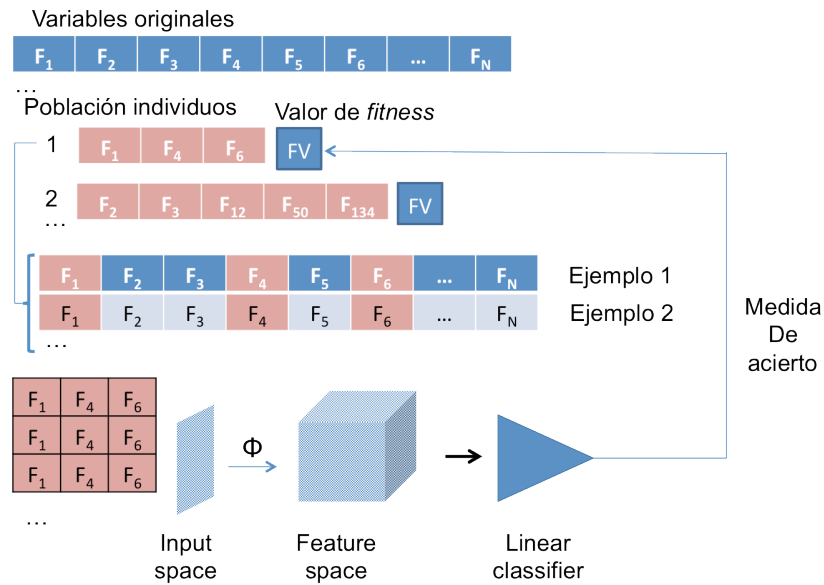


FIGURA 17. TÉCNICAS HÍBRIDAS BIO-INSPIRADAS CON UN CLASIFICADOR (SVM) PARA SELECCIÓN DE VARIABLES

1.4.2 SVM-RFE

Guyon et al. [115] propusieron SVM-RFE, ver Figura 18, para la selección de genes relevantes en la resolución de un problema de clasificación de cáncer. Esta técnica clasifica variables de acuerdo a una función objetivo y va eliminando uno o más genes con las menores puntuaciones mediante un proceso de entrenamiento de SVM con una función de *kernel* lineal. Esta técnica es conocida como selección hacia atrás o *backward*. El criterio de eliminación es el valor de decisión del hiperplano obtenido por el SVM. Así este algoritmo analiza la relevancia de las variables de entrada estimando cambios en la función de coste. Este proceso es repetido hasta que se obtiene el valor de clasificación más alto. Para resolver problemas de clasificación no lineales, Rakotomamonjy et al. modificaron esta técnica [118].

Se considera entonces que la orientación del hiperplano que encuentra el SVM puede dar más información, de forma que para una variable, si el plano es ortogonal, dicha variable es informativa del problema a resolver. Concretando, dado un SVM con un vector de pesos de la forma $\vec{w} = \sum_k \alpha_k y_k \vec{x}_k$, entonces cada variable i tiene un valor de ranking que se calcula como $c_i = (w_i)^2$. De manera operativa, lo que se hace es, mediante un proceso iterativo, eliminar la variable que menos importancia tiene según

el criterio c . Por tanto, en cada iteración del proceso iterativo se entrena un SVM lineal que calcula un valor de c para cada variable.

Se pretende evaluar el error de generalización del modelo, porque lo que este proceso de selección hacia atrás tiene que comprobarse con un conjunto de validación distinto del usado para entrenar el modelo.

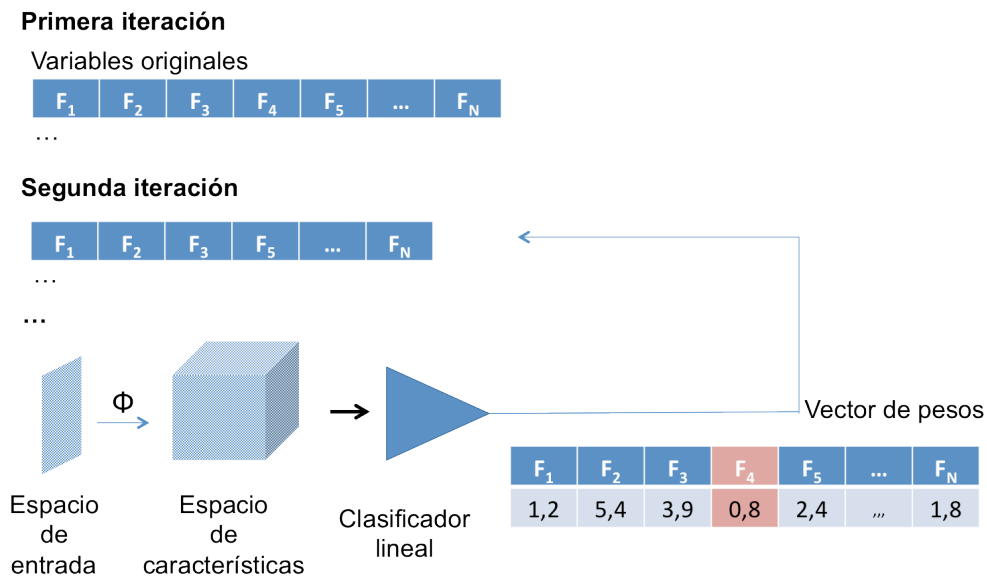


FIGURA 18. SVM-RFE: ELIMINACIÓN DE VARIABLES

A pesar de que reduce el número de variables utilizadas después de la fase de aprendizaje, estos métodos no son capaces de detectar variables redundantes y son además sensibles al ruido. Existen diferentes aproximaciones y modificaciones de este algoritmo para hacer menos costoso computacionalmente el proceso, como eliminar la mitad de las variables en cada iteración o eliminar la raíz cuadrada del número de variables por cada iteración [115,117-119]. Una aproximación muy interesante es la propuesta por Furlanello et. al [120] en la que la eliminación de las variables se hace de acuerdo a la entropía de la distribución de pesos.

1.4.3 MKL CON SELECCIÓN DE VARIABLES

En la apartado 1.3.2 se introdujo MKL. Seoane et . al [121] proponen la modificación de esta técnica mediante la inclusión de un sistema de selección de variables separado para cada conjunto de datos que se incluya en el aprendizaje. De este modo, se produce la integración de datos mediante un algoritmo supervisado como MKL. Para cada

kernel, y siempre en el caso de clasificación binaria, se calcula un ranking de las variables (significancia estadística) que están más estadísticamente alineadas con la clase de pertenencia.

Esta aproximación está desarrollada partiendo del SimpleMKL propuesto en [116], que realiza combinaciones lineales de kernels para minimizar mediante un método iterativo de gradiente descendiente los parámetros correspondientes al SVM (α_i) y kernel (η_i) en (Eq. 21). Esta implementación utiliza una regularización 2-normal que conduce a una solución *sparse* de los coeficientes del kernel dado que la mayoría tienden a cero. En los MKL no *sparse* estos valores no tienen por qué tender a cero. En análisis numérico una matriz *sparse* es una matriz de un gran tamaño en la que la mayoría de sus términos es cero.

Para establecer la significancia estadística dentro de cada conjunto de variables que pertenece a un dataset, se utiliza un t-test entre las variables y las clases del conjunto de entrenamiento. Se trata por tanto de un preprocesado para hacer un ranking de las variables de manera previa, sin tener en cuenta el método de predicción que se utilizará después. Es por tanto un método de selección de características tipo *filter* que construye, a partir de las variables más significativas, un *kernel* base para cada 2, 3 y hasta un número máximo de variables determinado por el usuario. Como *kernels* básicos se implementan los *kernels* polinomial y RBF presentados en el apartado anterior.

La ventaja de esta aproximación *filter* es que permite realizar el proceso de selección de variables de una manera muy rápida ya que son métodos muy simples. Se nombrará como *Feature Selection* MKL (FSMKL).

1.5 FUSIÓN DE DATOS

Hoy en día las técnicas de aprendizaje máquina y de minería de datos son herramientas imprescindibles para la extracción de conocimiento. El problema surge por la gran cantidad de datos de los que se dispone y del posible ruido que incorporan los mismos. Además de las técnicas de selección de variables ya comentadas, surgen las técnicas de fusión de datos que pretenden, mediante la incorporación de fuentes de

datos de origen distinto, aumentar la significancia estadística así como potenciar las posibles correlaciones e interacciones [8].

El concepto surgió como aplicación sobre datos genómicos. La idea subyacente es que mediante la incorporación de más de un conjunto de datos al análisis, se pueden mejorar los resultados para obtener información refinada de más alto nivel [9]. Se puede establecer, a grandes rasgos, una clasificación en tres clases según si se realiza una integración temprana, intermedia o tardía.

En las aproximaciones tempranas se fusionan diferentes conjuntos de datos heterogéneos que son analizados mediante un único método de análisis, de manera que se trata de encontrar una representación común para diferentes conjuntos de datos [122-125]. En una aproximación tardía los conjuntos de datos se tratan por separado haciendo inferencias independientes que luego son fusionadas siguiendo diferentes aproximaciones [126-128]. Por último, las aproximaciones intermedias son aquellas que no pueden ser asignadas de manera clara a ninguno de los grupos anteriores [129].

1.6 IMAGEN DE ELECTROFORESIS BIDIMENSIONAL

Durante el desarrollo de la prestación asistencial diaria resulta cada vez más corriente la realización de varias pruebas que apoyen al diagnóstico basado en imagen; radiografías, resonancias magnéticas, tomografía por emisión de positrones, tomografía computerizada o ecografía, por mencionar algunas.

Poder inferir conocimiento de forma automatizada de dichas pruebas ayuda en el diagnóstico y reduce el tiempo en la toma de decisiones de los facultativos. Desde el comienzo del diagnóstico por imagen, la detección de las regiones de interés (ROI) a estudiar se ha hecho siguiendo diferentes aproximaciones, como por ejemplo el uso de operadores morfológicos, pasando por el estudio de los niveles de gris de la imagen o la detección de estructuras presentes en la imagen.

Un ejemplo de imagen biomédica son las imágenes de electroforesis bidimensional. Se trata de una técnica de alta resolución cuyo objetivo es la separación de mezclas de proteínas altamente complejas. Se somete a una muestra (tejido, célula,

suero) a dos procesos físicos distintos sobre un gel de poliacrilamida. Por un lado se somete a la muestra a una diferencia de potencial en un gel con gradiente de pH en condiciones desnaturizantes de acuerdo con su punto eléctrico. Seguidamente de esta separación por carga, las proteínas son separadas de acuerdo con su masa molecular por electroforesis discontinua en gel de poliacrilamida en presencia de un método de tinción (Coomassie Blue, Sypro Ruby, Plata). Tras la tinción del gel las proteínas aparecen formando manchas circulares (spots).

Este proceso de electroforesis bidimensional además de permitir la representación con diferentes tinciones de las proteínas, produce sobre las imágenes la aparición de mucho ruido. En concreto, de ruido blanco Gaussiano acumulativo [130]. Ha tenerse en cuenta que los spots, se forman por acumulación de proteínas, por lo que su representación se asemeja a una curva Gaussiana. A pesar de haberse realizado trabajos para la reducción de ruido sobre este tipo de imágenes para combatir el ruido inherente al proceso de elaboración de las mismas, como por ejemplo el uso de filtros espaciales, deformaciones elásticas, filtro de mediana modificado de Wiener o filtros basados en wavelet por ejemplo [131-137], lo cierto es que el estudio de estas imágenes y la eliminación de ruido en las mismas aún es un trabajo abierto. El problema de la utilización de estos métodos para la eliminación del ruido radica en que, a pesar de ser simples en concepto y eficientes computacionalmente, producen distorsiones muy severas en los bordes de las proteínas y altera considerablemente la intensidad de los valores de los píxeles internos del spot.

En un análisis de textura, como ya se ha mencionado en esta Tesis Doctoral, los estadísticos de orden igual o superior a dos, son insensibles al ruido Gaussiano [138] como por ejemplo la matriz de coocurrencias de los valores de gris.

La proteómica es el estudio de las propiedades de un tejido, célula o suero que se realiza para obtener una visión global e integrada de enfermedad, procesos fisiológicos o bioquímicos de las células. Con esta técnica, las proteínas se clasifican por su peso molecular (MWt) y su punto isoeléctrico (pI) en un proceso controlado en el laboratorio y con equipamiento de imagen digital. Desde el comienzo de la

investigación en proteómica, esta técnica ha sido la más usada para separar proteínas de una muestra, incluso antes de que la proteómica fuese una realidad en sí misma.

Las principales ventajas de esta técnica residen en su robustez, además de tener una habilidad única para analizar proteínas completas en alta resolución de forma paralela, dejándolas intactas y permitiendo aislarlas completamente [139]. De todas maneras, esta técnica también tiene desventajas, como su muy baja efectividad en el análisis de proteínas hidrofóbicas, así como su alta sensibilidad a los rangos dinámicos (i.e. ratio cuantitativo entre la proteína más rara que se ha expresado en una muestra y la más abundante) y cuestiones relacionadas con la distribución cuantitativa [140].

El resultado final de un proceso de electroforesis bidimensional es una imagen como la mostrada en la Figura 19. Esta imagen en concreto fue utilizada en un trabajo para detectar biomarcadores de proteínas en suero en niños con síndrome fetal alcohólico. El tamaño de la imagen es 512x512 píxels con 8 bit de profundidad de color y 340 micrones/píxel. Esta imagen ha sido tomada de la base de datos pública de Lemkin [141]. Se puede observar en la imagen la posición final que ocupan las proteínas en el proceso de migración.

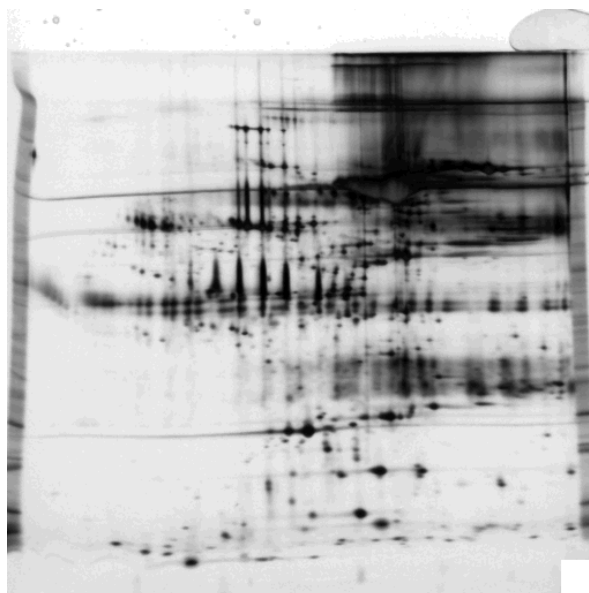


FIGURA 19. IMAGEN DE ELECTROFORESIS BIDIMENSIONAL

2. ESTADO DE LA CUESTIÓN

2.1 ANÁLISIS DE TEXTURAS EN IMAGEN MÉDICA

Desde la introducción del uso de texturas para clasificación de imágenes [15] han sido muchas las aproximaciones que se han presentado en la comunidad científica. El propio Haralick presentó en 1979 [31] varias aproximaciones que utilizaban métodos estadísticos como: funciones de autocorrelación, transformadas ópticas y digitales, densidad de bordes de las texturas, elementos estructurales, coocurrencia de niveles de gris y modelos autoregresivos, además de métodos estructurales como los basados en primitivas complejas y sus relaciones espaciales.

Como ya se ha mencionado, siendo cierto que no hay unanimidad para definir una textura, es comprensible que tampoco exista un acuerdo entre los posibles grupos o agrupaciones de las mismas. Dónde casi todos los autores coinciden es en señalar los tipos estadísticos y estructurales. Hay discrepancias con la consideración de dos grupos más: el basado en modelos [142] y el basado en transformadas.

Por todo lo comentado en apartados anteriores de esta Tesis Doctoral el uso de una aproximación basada en texturas para realizar una segmentación presentaría serias limitaciones debido al efecto borde que se genera en las zonas de transición entre imágenes, precisamente por la presunción de que el valor de textura de cada píxel depende de la distribución de los niveles de gris con sus píxeles vecinos. Más en concreto, este efecto borde aparece de manera más pronunciada cuando los bordes texturales de la forma estudiada son paralelos a la orientación de la célula [143].

En cambio, una aproximación basada en texturas cuando se desea buscar elementos en la imagen en los que sus principales características derivan de su especificidad en la organización espacial de sus elementos tiene grandes visos de lograr buenos resultados.

Esta combinación del conocimiento de texturas con otros algoritmos para la segmentación ha mejorado resultados en otro tipo de imagen médica como ecografía [144,145], ecocardiografía [146], resonancia magnética del cerebro [147], imágenes de teledetección [148], microscopía [149] o mamografías [150,151] entre otras.

Mahapatra et al. [152] evaluaron el incremento en la incidencia de la enfermedad de Crohn mediante una nueva aproximación que distingue regiones normales de patológicas usando variables de bajo nivel como las texturas, intensidad o información de asimetría, con especial énfasis en operadores como *skewness* o *kurtosis*. Un diagnóstico preciso en esta enfermedad en concreto es un reto médico, debido a que se podrían suprimir las actuales pruebas invasivas, evitando sufrimiento a los pacientes, usando técnicas no invasivas como la resonancia magnética.

Markel et al. [153] caracterizaron regiones correspondientes con cáncer de pulmón y órganos del tórax mediante operadores estadísticos de primer, segundo y orden superior. En concreto usaron la combinación de un árbol de decisión con un clasificador *K-nearest neighbor*. Este trabajo se llevó a cabo para definir dianas en radioterapia obteniendo unos valores muy altos de especificidad con los datos proporcionados por tres clínicos.

Weaver et al. [154] proponen un análisis de texturas como técnica no invasiva en el estudio en la cuantificación de la distrofia muscular en el golder retriever. El autor apunta en este trabajo que el estudio de desórdenes musculares está siendo cada vez más utilizado para evaluar la progresión de la distrofia muscular de Duchenne.

Abbey et al. [155] proponen la cuantificación y caracterización con operadores estadísticos no Gaussianos de textura en mamografías digitales, mejorando el diagnóstico y la estimación del estadio del cáncer, permitiendo además el estudio de la respuesta al tratamiento oncológico usando la combinación de distintas imágenes médicas. En este trabajo, los autores proponen que la heterogeneidad espacial del tumor es un factor pronóstico importante y señalan al análisis de texturas como uno de los factores principales que lo explican. [156] detectaron diferencias significativas en textura entre los carcinomas invasivos lobulares y ductal en mamografía.

Existen varios trabajos de Ganeshan et al. [157-159] donde se comparan variables de textura con otras variables presentes en la imagen y con parámetros biológicos [160] para caracterizar regiones correspondientes a diferentes tipos de cáncer en distintos tipos de imagen, evidenciando el potencial de correlación con la enfermedad de los operadores texturales.

En [161] se estudia si el tiempo de respuesta a un determinado fármaco para el tratamiento de un cáncer tiene alguna correlación con la textura del tumor y la progresión de los pacientes después del tratamiento. La conclusión a la que llegan es que la heterogeneidad del mismo es un factor independiente asociado a la progresión temporal del tumor y tiene un gran potencial como biomarcador en imagen para estudiar la respuesta de las metástasis en cáncer renal.

En ocasiones se utiliza el análisis de texturas de manera conjunta con otro tipo de técnicas de postprocesado de imagen dando mejores resultados en el tipado y graduación de un glioblastoma, pudiendo diferenciar los tumores glioneuronales malignos como una subclase de glioma maligno de grados tres y cuatro de la Organización Mundial de la Salud [162].

Una aproximación con redes de neuronas artificiales [163,164] para analizar lesiones de cáncer de pecho mediante un análisis de textura se comparó con dos expertos clínicos utilizando un t-test llegando a la conclusión de que el rendimiento obtenido era comparable al obtenido por los humanos [165].

Las variables de textura están altamente correlacionadas con la severidad de la fibrosis pulmonar [166,167]. En este trabajo, se relacionan datos de imagen con biomarcadores obtenidos en el laboratorio para medir la evolución de la enfermedad.

2.2 SELECCIÓN DE VARIABLES EN ANÁLISIS DE TEXTURAS

Con la técnica de selección de variables presentada en 1.4.2, SVM-RFE, Zacharaki et al. [168] clasifican el grado de tumor cerebral en imágenes de resonancia magnética con datos extraídos de un análisis de textura. Como indicaron los autores, en su trabajo parecen encontrar una capacidad de discriminación entre meningiomas y gliomas de

todos los grados en la sección de realce de una T1ce. Más específicamente en el caso del glioblastoma, sólo con T1ce se consigue un 97.4% de porcentaje de acierto. Mediante una combinación de variables multiparamétricas, diferencian entre glioma primario y metástasis (84.7% de porcentaje de acierto) y entre grados alto y bajo de glioma (87.8% de acierto).

Para la caracterización de un cáncer de próstata, Lopes et al. [169] utilizan con éxito una aproximación basada en SVM y AdaBoost con variables de textura.

En [170] los autores proponen una aproximación basada en la medida de similitud con distancia Bhattacharyya para selección de características que lleven a la clasificación de imágenes de resonancia magnética mediante variables de textura 2D y 3D.

En [171-173] se presentan aproximaciones híbridas de GA con distintos clasificadores (red de neuronas artificiales, redes bayesianas) para la selección de características de textura en mamografías y diferenciar tejido sano de enfermo. Se puede consultar en el apartado 1.4.1 de esta Tesis para métodos híbridos de selección de variables.

En [174] se presenta una modificación del algoritmo presentado en el apartado 1.4.2 pero en lugar de utilizar el peso de las variables en el proceso hacia atrás e ir eliminando las menos interesantes, se hace usando un ranking de variables con un t-test. Las variables de textura son combinaciones lineales de las extraídas originariamente de las imágenes de resonancia magnética.

Para la detección de nodos linfáticos en imágenes de ultrasonidos mediante variables de textura [175] presentan una aproximación híbrida de PSO y redes de neuronas artificiales con la función de Boltzman para la selección de variables más representativas y luego realizan una clasificación multiclase (6 tipos de tejidos) con un SVM de los tejidos.

En [176] se realiza una normalización de imágenes de resonancia magnética del cerebro para identificar esclerosis múltiple. Dentro de este trabajo, se realiza un

ranking de las variables más correlacionadas con la salida del proceso de entrenamiento supervisado mediante un test de Wilcoxon.

En [177] comparan el uso de diversas aproximaciones basadas en kernels para la selección de variables de textura en el diagnóstico del labio, método de diagnóstico de la medicina tradicional china. Concluyen que SVM-RFE supera a otras aproximaciones como SVM-KNN o Naïve Bayes.

2.3 CONSIDERACIONES

Como se ha visto a lo largo de este capítulo, el análisis de texturas ha sido utilizado satisfactoriamente en el tratamiento de diferentes tipos de imagen biomédica. Además se ha indicado cómo diferentes tipos de algoritmos de aprendizaje máquina han sido utilizados para realizar dicho análisis, destacando las técnicas basadas en *kernel* por su capacidad de discriminación, rendimiento global y capacidad de generalización en espacios de alta dimensionalidad.

La clasificación por texturas implica el procesado de un elevado número de características para caracterizar cada región textural de manera adecuada. La información de textura tiene una naturaleza multicanal, es por esto que además los sistemas de segmentación se encuentran con dificultades al tratar con este problema.

Las texturas en el mundo real no son uniformes debido a cambios en la orientación, escala o apariencia visual por lo que existen múltiples descriptores de textura para cada píxel que caracterizan en base a la relación espacial entre píxeles vecinos a las texturas presentes en la imagen.

Muchos de los trabajos que se han revisado trabajan directamente con la GLCM, operador estadístico de segundo orden, dando por hecho que es el operador de texturas que mejor va a funcionar sin hacer ningún tipo de proceso de selección de variables.

En el apartado 2.2 se ha visto cómo mediante el uso de técnicas de selección de características es posible mejorar los resultados obtenidos. Las técnicas de selección de características pueden ser usadas en combinación con técnicas del apartado 2

recibiendo el nombre de técnicas híbridas. La ventaja de utilizar técnicas de selección de características, en general, en lugar de técnicas de extracción de características es que mantienen la interpretabilidad de los resultados obtenidos como se vio anteriormente.

Todas estas consideraciones hacen necesario la realización de un estudio adaptativo que permita la reducción de dimensionalidad del problema y permita la selección de las variables más representativas en el dominio de la imagen biomédica. Además, las técnicas híbridas basadas en kernel aparecen como una de las más indicadas para este tipo de análisis por los resultados publicados en la literatura. Se han presentado, del mismo modo, múltiples tipos de imagen biomédica que han sido analizadas mediante descriptores de textura y se ha visto cómo los resultados obtenidos han sido mejores que utilizando otro tipo de técnicas. Es por esto que se ejemplificarán las propuestas de esta Tesis Doctoral acerca de un análisis de texturas sobre un tipo particular de imagen biomédica como son las imágenes de electroforesis bidimensional sin pérdida de generalidad.

3. HIPÓTESIS DE TRABAJO

Las texturas siempre están presentes en las imágenes, aunque con diferente grado de importancia. Existen múltiples tipos diferentes de variables de textura y de una naturaleza muy distinta que se pueden extraer de una imagen. Como se ha visto en el capítulo anterior, en las imágenes biomédicas su grado de importancia es elevado. Como ejemplo, mediante un análisis de texturas es posible obtener suficiente información para discriminar regiones de interés relacionadas con una determinada patología, permitiendo clasificar dentro de las imágenes entre tejido sano y tejido enfermo. La situación actual hace que el nivel de definición de las imágenes y la cantidad de información que se puede extraer de las mismas sea muy elevado, más si se tiene en cuenta que esta información se extrae a nivel de píxel.

Las técnicas basadas en *kernel* han demostrado una gran eficiencia en problemas de clasificación, siendo además destacable en el caso de dominios de una gran dimensionalidad. Es posible además utilizar este tipo de técnicas para la selección de variables y reducción de dimensionalidad del problema.

Las imágenes biomédicas presentan una serie de características genéricas que permiten, como se ha visto en el capítulo anterior de Estado de la Cuestión, clasificar una imagen mediante un análisis de texturas. El campo de trabajo está abierto, ya que está por formalizar el proceso que permita automatizar el uso de técnicas basadas en *kernel* para sacar partido de toda la información de textura (de diferentes tipos) que se puede extraer de una imagen.

Esto supone el desarrollo de nuevas propuestas para reducción de dimensionalidad e integración de datos de diferentes tipos de textura dado que cada tipo de imagen biomédica en particular puede estar influenciada por un número y tipo de variables de textura distintos. Esto va a permitir la eliminación de variables que aportan poca o ninguna información mejorando la interpretabilidad de las escogidas como solución del problema.

Cuando se realiza un análisis de cualquier tipo de imagen biomédica se está buscando la clasificación de la misma en zonas diferenciadas, por lo que sin pérdida de generalidad, se validará esta propuesta con imágenes de electroforesis bidimensional en las que se busca la clasificación de curvas Gaussianas en un medio en el que de manera inherente a su proceso de creación se añade ruido blanco Gaussiano acumulativo [130] mediante un análisis de texturas, inmunes a ese tipo concreto de ruido [138].

Es por ello que esta Tesis Doctoral plantea la siguiente hipótesis de trabajo sobre la que se fundamenta:

El desarrollo de nuevas de técnicas basadas en kernel para el análisis de texturas, mediante la integración de distintos tipos de datos de textura y la selección de variables más representativas en imágenes de electroforesis bidimensional, mejorará los resultados obtenidos en clasificación y en interpretabilidad de las variables solución al problema.

Se espera que el uso de estas técnicas basadas en *kernel* obtengan el mejor subconjunto de variables de cada tipo distinto de variable de textura estudiado, minimizando el impacto del ruido presente en estas imágenes y mejorando por tanto de manera sensible los resultados obtenidos.

4. PROPUESTA PARA EL ANÁLISIS BASADO EN TEXTURAS EN IMAGEN BIOMÉDICA

Esta Tesis Doctoral propone mejorar los resultados de clasificación en un análisis de texturas en imagen biomédica. Se ha visto en el apartado del Estado de la cuestión que existen diferentes tipos de técnicas basadas en *kernel* que se pueden utilizar para realizar este análisis y que, además, han sido utilizadas con éxito en diferentes tipos de imágenes biomédicas.

Para evaluar las mejoras propuestas, en primer lugar será necesario obtener resultados para los modelos basados en *kernel* del estado de la cuestión para poder cuantificar la mejora obtenida con las propuestas que se presentarán en esta Tesis Doctoral. Es necesario realizar esta cuantificación en condiciones de igualdad y de manera que sea reproducible en futuros experimentos, por lo que se comenzará esta tesis formalizando un diseño experimental que incluye una fase final de selección del mejor modelo mediante la aplicación de test estadísticos.

Se propondrán como mejoras el uso de nuevos kernels y el estudio de nuevas medidas de correlación entre las variables y la clase de pertenencia dentro de una aproximación *filter* para la selección de características. Se espera que estas propuestas mejoren los resultados obtenidos con *kernels* de propósito general como los presentados en la sección 1.3.1.4 y también a combinaciones lineales de *kernels* como los obtenidos con técnicas como la presentada en el apartado 1.3.2. Del mismo modo, con las nuevas medidas propuestas se pretende paliar las debilidades de una medida en una única dimensión como el t-test para medir la diferencia, recurriendo a técnicas que sean capaces de evaluar el solapamiento entre clases [178]. Estas mejoras se integrarán sobre técnicas ya publicadas en la literatura [121] tomadas como base y se espera que se mejoren los resultados obtenidos con las técnicas del estado de la cuestión comentadas en el capítulo de Fundamentos.

Tal y como se ha comentado, ha de tenerse en cuenta que para comprobar el correcto funcionamiento de cualquier nueva técnica propuesta, es inexcusable que se puedan comparar en igualdad de condiciones los resultados obtenidos por ésta con respecto a las técnicas existentes. Es necesario por consiguiente, que los experimentos sean lo más reproducibles posible y que se tenga un conjunto de datos fijos. Es por esto que existen bases de datos públicas en internet como la de la Universidad de California (UCI) [179] para probar algoritmos de aprendizaje máquina, por citar alguna de las más conocidas en este ámbito. En el caso concreto de la imagen biomédica, debería existir o bien un atlas o un *ground truth* para utilizar en este tipo de comparaciones. Por desgracia esta cuestión dista de la realidad en lo relacionado con múltiples tipos de imagen, como se verá más adelante.

Sin embargo, a pesar de ser lo más adecuado, resulta de extrema dificultad poder comparar los resultados en igualdad de condiciones con las técnicas estado de la cuestión ya que los autores suelen ser reacios a compartir información crucial de las pruebas realizadas (por ejemplo los puntos concretos estudiados de la imagen, criterios de selección, etc.). Para lograr que la comparación sea justa, se propondrá un diseño experimental que garantice la comparabilidad de los resultados, lo que ahora mismo no es posible. Este diseño experimental es genérico y se puede aplicar a cualquier tipo de imagen para la comparación de diferentes modelos. Evidentemente ha de tenerse en cuenta que inicialmente se debe realizar el estudio del estado de la cuestión del tipo de imagen en particular y, después, siguiendo el diseño experimental, se podrá refinar el modelo obtenido. Ninguna de las fases de este diseño experimental es obligatoria, aunque sí recomendable.

Se formalizan a continuación las diferentes fases que se proponen para automatizar un análisis de texturas con técnicas basadas en kernel para reducción de dimensionalidad. Los objetivos que se persiguen es que se pueda, por un lado, maximizar la probabilidad de éxito y, por otro, que los resultados de diferentes experimentos sean comparables.

Se presentará en el capítulo 5 Pruebas y Resultados cómo se incluyen las mejoras propuestas en la fase adecuada de este diseño experimental. Además, se

explicará para el caso concreto de las imágenes de electroforesis bidimensional cómo se ha llevado a cabo cada una de las fases y cómo se pretende realizar la comparación en igualdad de condiciones y mantener una futura reproducibilidad de los resultados y del proceso a seguir.

El término *modelo de referencia*, a lo largo de esta Tesis Doctoral, se referirá a que ha sido necesario la generación de diferentes *modelos* de clasificación, ampliamente usados y conocidos en el dominio del aprendizaje máquina, que serán utilizados como *referencia* para comparar los resultados obtenidos con las mejoras propuestas en esta Tesis Doctoral. Es por esto que se generarán los diferentes modelos de referencia, utilizando para ello las técnicas basadas en *kernel* del estado de la cuestión y se calcularán los resultados obtenidos por cada una de ellas como punto de partida. Se seleccionará el mejor modelo y se comparará si los resultados obtenidos, con las mejoras propuestas en esta Tesis Doctoral, son estadísticamente significativas mediante la aplicación de test estadísticos paramétricos/no paramétricos.

Se finalizará el capítulo de Pruebas y Resultados estudiando un tipo específico de *kernel* Mahalanobis para problemas de alta dimensionalidad.

4.1 FORMALIZACIÓN DE LAS FASES DEL DISEÑO EXPERIMENTAL

Dados unos datos, se debe elegir el mejor modelo para resolver el problema. Es necesario para ello formalizar un diseño experimental para conseguir resolver de la manera más eficaz posible el problema con los modelos elegidos y poder estimar el error cometido por los mismos.

Esta Tesis propone la adopción de un diseño experimental que se compone de cuatro fases: extracción, preprocesado de los datos, aprendizaje y selección del mejor modelo, como puede verse en la Figura 20.

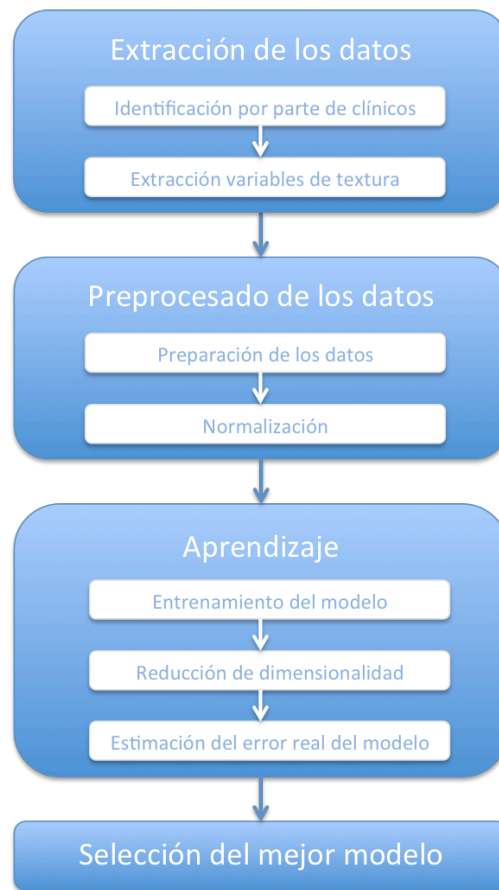


FIGURA 20. FASES DEL DISEÑO EXPERIMENTAL PROPUESTO

Se describirá cada una de las fases del diseño experimental en las secciones siguientes de este capítulo.

4.1.1 EXTRACCIÓN DE DATOS

Durante esta primera fase, los clínicos deben ponerse de acuerdo en la identificación de las regiones de interés que se quieran estudiar dentro de las imágenes. Deben realizar el marcado de las mismas y consensuar en el grupo de trabajo las zonas comunes que mejor representen la problemática concreta.

Este proceso no es sencillo debido a que en el análisis manual de imágenes biomédicas es habitual que exista una gran variabilidad inter- e intra-operador, lo que dificulta su análisis y hace que las técnicas de análisis de imágenes surjan como medio de apoyo para los clínicos.

Una vez se han consensuado las zonas de interés, es necesaria la utilización de algún tipo de herramienta para extraer las variables de textura que se utilizarán

posteriormente. Para ello, se dispone de diferentes herramientas software, entre las que destaca Mazda [180-182]. Existen múltiples trabajos en los que se demuestra la efectividad de dicha herramienta para extraer variables texturales en diferentes tipos de imágenes médicas [4,5,156,183-188].

4.1.2 PREPROCESADO

4.1.2.1 PREPARACIÓN DE LOS DATOS

Un paso primordial que se debe llevar a cabo cuando se pretende abordar el análisis de unos datos es estudiar los mismos. Los datos en estado puro son a menudo difíciles de analizar, por lo que toda fase de análisis suele requerir un estudio previo de los datos con los que se está trabajando y un preprocesado de los mismos. Es necesario comprobar que no existen datos con información incompleta, casos atípicos o ruido en los mismos. Se considera que los datos, una vez procesados, están listos para comenzar el análisis. Frecuentemente se afirma que el 80% del esfuerzo de un análisis de datos, se invierte en preparar los datos de manera correcta para el análisis [189].

En general, y para las técnicas basadas en kernel en particular, es recomendable siempre que sea posible, trabajar con un conjunto balanceado de datos [190,191]. Cuando se entrena con conjuntos desbalanceados de datos, tienden a generarse clasificadores que sobrepredicen la clase mayoritaria. Existen múltiples trabajos en los que se puede apreciar que un clasificador entrenado con un conjunto balanceado de datos ofrece mejores resultados globales comparado con un conjunto desbalanceado [192-194]. En este caso, se propone la utilización de alguna técnica que se encargue de corregirlo como SMOTE [195] (*Synthetic Minority Over-Sampling Technique*) tal y como se explica con más detenimiento en los anexos de esta Tesis Doctoral.

Es necesario estudiar también la correlación entre las variables de entrada, ya que en el caso de que existan variables que estén muy correlacionadas entre si, generalmente es conveniente eliminarlas antes de pasarlas por el modelo ya que podrían introducir ruido en el experimento, no favoreciendo el resultado.

Es común representar la clase de pertenencia de un determinado ejemplo en problemas de clasificación con números enteros positivos que están en el rango que va

desde cero hasta el número de clases totales del problema menos uno. En esta Tesis Doctoral, diferentes técnicas basadas en kernel son utilizadas para resolver el problema de reducir la dimensionalidad de un análisis de texturas. Como se explicó en el apartado 1.3, con estas técnicas se pretende encontrar el hiperplano de separación óptimo que maximice los márgenes de separación entre clases, situando no sólo los ejemplos en el lado correcto del hiperplano, sino haciéndolo con la mayor distancia posible, lo que servirá para mejorar la capacidad de generalización del modelo. Es por esto que en este tipo de problemas se utilizan por tanto las etiquetas $-1/+1$ para diferenciar las dos clases. Recordando que el par $\{x^t, r^t\}$ es el conjunto ordenado de entradas y salidas deseadas donde $r^t = +1$ si $x^t \in C_1$ y $r^t = -1$ si $x^t \in C_2$, se pretende encontrar w y w_0 tal que

$$w^T x^t + w_0 \geq +1 \text{ para } r^t = +1$$

$$w^T x^t + w_0 \leq -1 \text{ para } r^t = -1$$

4.1.2.2 NORMALIZACIÓN

Las variables de un conjunto de datos pueden estar en diferentes escalas y ser de distintos tamaños haciéndolas difíciles de comparar. Es por esto que surgen las técnicas de normalización donde se realiza un re-escalado lineal de las variables de entrada propiciando la eliminación de *outliers* y permitiendo que se puedan observar patrones más claramente dentro de los datos. Con ellas se trata las variables de entrada independientemente, de manera que las nuevas variables generadas tengan media igual a cero y desviación típica igual a uno. Para cada variable x_i se calcula su media (μ_i) y la desviación típica (σ_i) en el conjunto de entrenamiento, y por tanto, las nuevas variables se calculan como

$$(EQ. 22) \quad x_i = \frac{x_i - \mu_i}{\sigma_i}$$

Al aplicar este proceso, los datos tiene una media nula y una desviación típica igual a uno haciendo que los datos sean independientes de la unidad de medida con la que se han calculado, permitiendo de esta manera comparar más fácilmente las variaciones.

4.1.3 APRENDIZAJE

Ante un problema para el que no hay suficiente información que permita el desarrollo de un algoritmo que lo resuelva, pero por el contrario se dispone de un elevado número de ejemplos acerca de la tarea a desarrollar, se pueden utilizar técnicas basadas en inteligencia artificial para el aprendizaje.

Las técnicas de aprendizaje máquina pueden aplicarse a diversos problemas, principalmente de clasificación. En los problemas de clasificación, como el estudiado en esta Tesis Doctoral, se trata de identificar si un determinado ejemplo pertenece a cierta clase o no.

Después de realizar esta fase de entrenamiento, se trataría de evaluar un nuevo caso que no se haya visto antes a una de las clases predefinidas del problema. Esto supone que se debe tener algún tipo de medida de acierto que indique de manera clara cómo de bien se ha realizado esta fase de entrenamiento. Es deseable además que el modelo no sobreentrene, de manera que se ajuste a los datos de los que dispone, ya que ante nuevos datos desconocidos la respuesta del mismo sería mala. Por último, cuanto mayor sea la dimensionalidad de los datos, existen más posibilidades de que se haya introducido ruido o redundancia en los mismos que dificulten la extracción de conocimiento.

Por todo ello se discutirán a continuación, siguiendo lo indicado en la Figura 20, técnicas de validación cruzada que eviten el sobreentrenamiento de un modelo, diferentes medidas de acierto que se pueden utilizar para valorar el rendimiento de un algoritmo y la necesidad del uso de técnicas para la reducción de la dimensionalidad del problema como las comentadas en el apartado 1.4.

4.1.3.1 ENTRENAMIENTO DEL MODELO

Como se ha indicado en el punto anterior, es deseable que no se produzca sobreentrenamiento en la fase de aprendizaje de un clasificador, para lo que se recurre a técnicas de validación cruzada o *cross validation*. La idea de k-fold cross validation surgió a finales de los años sesenta (Mosteller 1968) y consiste en separar los datos en dos grupos: un grupo para entrenamiento y otro para validación. A su vez el conjunto

de datos de entrenamiento se divide en k conjuntos que no se solapan (*k-fold cross validation*). En cada iteración (que varía de 1 a k) uno de estos conjuntos se emplea para validar el resultado obtenido tras el entrenamiento del modelo realizado a partir de los $(k-1)$ conjuntos restantes.

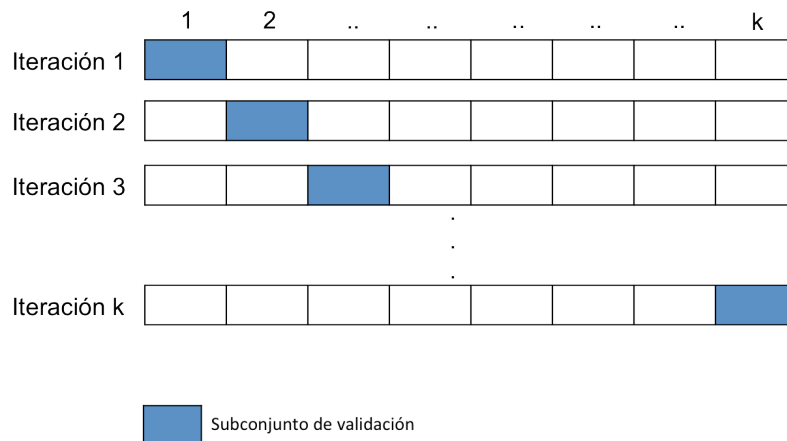


FIGURA 21. MÉTODO DE VALIDACIÓN CRUZADA

Se considera que las técnicas de validación cruzada tienen principalmente dos objetivos:

- Medir el grado de generalización de un algoritmo entrenado. Se mide el rendimiento del modelo entrenado y se estima el rendimiento sobre nuevos datos.
- Comparar el rendimiento de dos o más algoritmos que hayan sido entrenados en las mismas condiciones y determinar cuál es el mejor con el conjunto de datos disponible. De manera alternativa, es posible comparar el rendimiento de dos o más variantes del mismo modelo con distintas parametrizaciones.

El error de entrenamiento no es adecuado para comparar modelos como se ha dicho con anterioridad, por lo que es necesario comprobar si las diferencias que existen en el error de generalización entre los diferentes métodos son significativas o no. Para esto, existen métodos estadísticos de contraste de hipótesis que se verán más adelante en esta Tesis Doctoral. Además, se suele repetir el experimento un número suficiente de veces que permita comprobar que el reparto realizado por la validación cruzada no ha influido en la capacidad de generalización del modelo.

4.1.3.2 REDUCCIÓN DE LA DIMENSIONALIDAD

La conocida como *maldición de la dimensionalidad* [196] implica que cuanto mayor sea la dimensionalidad de los datos de entrada, mayor será el número de ejemplos necesarios para que el aprendizaje proporcione como resultado un buen modelo. En contraposición con esta teoría, existe la de la *bendición de la dimensionalidad* [197] que establece que en problemas de clasificación, la separabilidad entre clases aumenta con la dimensionalidad. El problema de un espacio de entrada de variables de alta dimensionalidad, es que además, puede encontrarse información redundante o correlacionada, que empeore los resultados del modelo.

Es obvio que la reducción de la dimensionalidad de los datos de entrada implica que se reduzca la complejidad del modelo generado, lo que además supone una reducción del tiempo de ejecución y una mejora de la capacidad de generalización global del sistema.

Como se comentó en el apartado 1.4, cuando se quiere reducir la dimensionalidad de un problema, puede seguirse una aproximación para que se elija el mejor subconjunto de variables originales o bien, crear un conjunto de variables a partir de combinaciones de las originales, que representen los datos con el menor número de variables posibles y que conserve la mayor cantidad de información. Esto es lo que se conoce respectivamente como técnicas de selección y extracción de características.

Por tanto, en las técnicas de selección de características, lo que se busca es encontrar las n dimensiones de m , que dan más información descartando las otras $(m - n)$ dimensiones. En las técnicas de extracción de características, lo que se busca es encontrar un nuevo conjunto n , combinación del conjunto original de m dimensiones. Este tipo de aproximación requiere una estrategia de búsqueda para elegir cuáles son las variables candidatas y una función que las evalúe y compruebe la bondad de esa solución.

La extracción de características puede requerir en algunos casos un elevado coste computacional y las variables originales se pierden, por lo que no se puede, por

ejemplo, extraer reglas con significado real acerca de las variables más útiles del análisis de texturas que se ha llevado a cabo.

4.1.3.3 ESTIMACIÓN DEL ERROR REAL DEL MODELO

Cuando se trata de evaluar el rendimiento de un algoritmo de clasificación, existe un elevado número de medidas de acierto bien conocidas para un clasificador de dos clases en la literatura: porcentaje de acierto de clasificación, precisión, sensibilidad, especificidad, medida F, área debajo de la curva Característica Operativa del Receptor (AUROC), índice Youden, valor kappa de Cohen, etc.

Una comparación experimental del rendimiento de estas medidas para problemas de clasificación puede encontrarse en [198] en donde el autor propone que la medida más segura es AUROC. Es mayoritaria la opinión dentro de la comunidad científica de que se trata de una buena medida de acierto del rendimiento de un algoritmo, aunque no para comparar distintos algoritmos [199-201].

En todo caso, la curva ROC es útil para visualizar el rendimiento de un clasificador. Este tipo de curvas es muy utilizada en el ámbito de la toma de decisión médica y desde finales de los años ochenta [202] se ha venido usando de manera incremental en el mundo del aprendizaje máquina, debido sobre todo a que las medidas de acierto sencillas no son adecuadas para comparar el rendimiento de diferentes clasificadores [203-205]. Un buen trabajo introductorio acerca de análisis de resultados basado en la curva ROC puede verse en [206].

Por lo general, en los métodos de selección de variables el error de entrenamiento describe cómo de bien se ajusta el subconjunto de variables al conjunto de entrenamiento y el error de generalización describe cómo de bien el subconjunto de variables separa nuevos datos no conocidos para clasificación. El objetivo es minimizar el error de entrenamiento. Ha de ser tenido en cuenta que el valor de acierto obtenido en entrenamiento suele ser ligeramente superior al obtenido en validación debido a un pequeño grado de sobrentrenamiento asociado a los espacios de alta dimensionalidad.

En el siguiente apartado se comentarán los diferentes métodos de contraste de hipótesis y test estadísticos que se pueden aplicar para seleccionar el mejor clasificador así como los requisitos necesarios para aplicar dichos test de manera correcta.

4.1.4 SELECCIÓN DEL MEJOR MODELO

La selección del mejor modelo ha de hacerse en función del rendimiento obtenido por los mismos. Esto no puede hacerse directamente con las medidas de acierto obtenidas para una ejecución concreta. Dado que se trata de un proceso estocástico, es necesario realizar diferentes ejecuciones de cada modelo para asegurar que los resultados obtenidos no se han visto influenciados por la componente aleatoria intrínseca de este tipo de procesos. Esto quiere decir que se compararán resultados de diferentes modelos que han sido entrenados en diferentes ejecuciones sobre los mismos datos. Para determinar la existencia o no de diferencias significativas entre los diferentes modelos, será necesario realizar un estudio estadístico, en concreto se deben utilizar métodos de contraste de hipótesis.

Los métodos estadísticos de contraste de hipótesis se utilizan para comprar la validez o no de una hipótesis (hipótesis nula). Las etapas de un test de contraste de hipótesis son las siguientes:

1. Definición de la hipótesis nula (H_0)
2. Selección, en función del caso, del test estadístico de contraste que permita evaluar la validez de H_0 . (Paramétrico/No paramétrico)
3. Elección del nivel de significación del test, o probabilidad de rechazar H_0 siendo cierta (α)
4. Cálculo de la probabilidad de obtener una discrepancia mayor de la observada siendo H_0 cierta (p-valor)
5. Comparación del p-valor obtenido con nivel de significación
 - a. Si $p\text{-valor} \leq \alpha$ entonces rechazar H_0
 - b. Si $p\text{-valor} > \alpha$ entonces aceptar H_0

Por tanto, el p-valor obtenido indica el nivel de significación más pequeño a partir del cual se permitiría rechazar la hipótesis nula (H_0). Los métodos de contraste

de hipótesis pueden ser paramétricos o no paramétricos [207]. Y, dependiendo del problema concreto sobre el que se vayan a aplicar, será necesario usar u otros. Se pueden ver en la Tabla 2 los test paramétricos y no paramétricos más usados dependiendo del número de variables u observaciones a comparar.

TABLA 2. TESTS ESTADÍSTICOS MÁS UTILIZADOS DEPENDIENDO DEL NÚMERO DE VARIABLES

Número de variables	Paramétrico	No paramétrico
2	T-test	Wilcoxon
Mayor que 2	ANOVA	Friedman, Friedman alineado, Quade

Para poder utilizar un test paramétrico se puede, o bien suponer que se ha obtenido una muestra aleatoria de una distribución normal o de Gauss, o comprobar la independencia, normalidad y heteroscedasticidad de los datos [208-212] siguiendo el diagrama de flujo mostrado en la Figura 22. Los test paramétricos son más potentes aunque su utilización depende de que se cumplan unas determinadas condiciones y los test no paramétricos, no hacen ninguna suposición acerca de la normalidad de los datos. En caso de que no se cumpla alguna de las condiciones se realizaría un test no paramétrico. La definición de cada una de estas condiciones es la siguiente:

- En estadística, dos eventos son independientes cuando el hecho de que uno ocurra no modifica la probabilidad de que lo haga el otro. La condición de independencia se ve satisfecha con la separación mediante validación cruzada de los datos.
- Una observación es normal cuando su comportamiento sigue una distribución normal o Gaussiana con un cierto valor de media y varianza. Se puede comprobar la normalidad de los datos utilizando el test de Shapiro-Wilk [213]. Se considera uno de los test más potentes para comprobar la normalidad de los datos, sobre todo en muestras pequeñas que no superen un tamaño de valores de treinta.
- La heteroscedasticidad indica la existencia de una violación de la hipótesis de igualdad de varianzas [214]. Para evaluar esta condición, se puede utilizar un test de Bartlett [211,215].

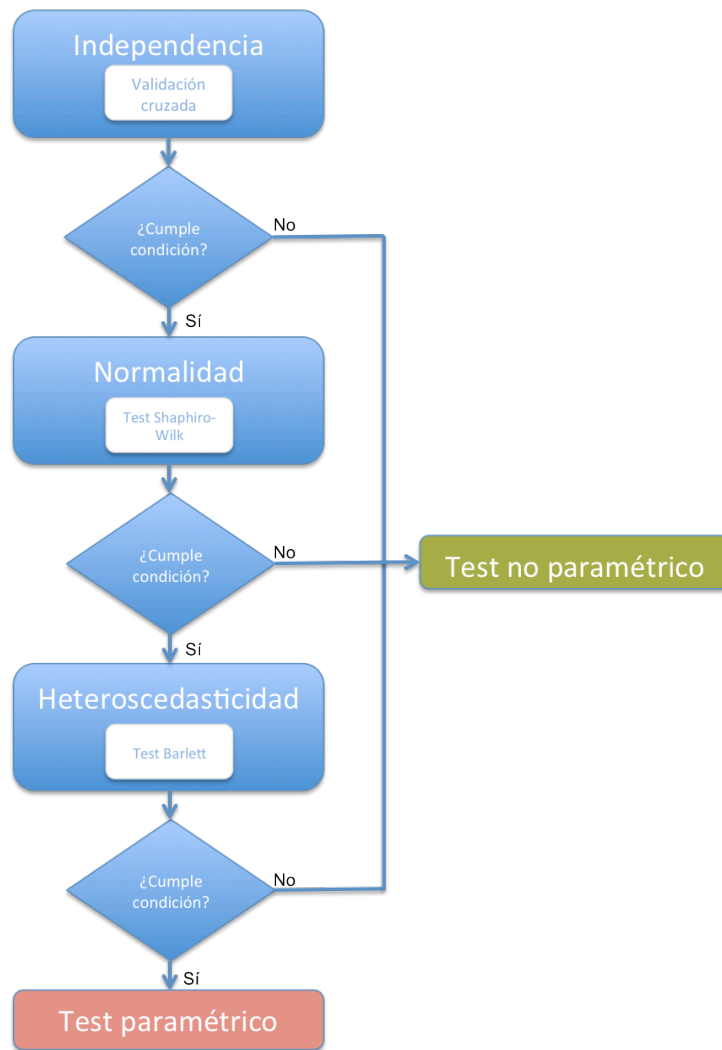


FIGURA 22. TEST PARAMÉTRICO: DIAGRAMA DE FLUJO

Un test estadístico puede utilizarse para comparar resultados entre algoritmos, siendo recomendable que se hayan obtenido siguiendo los mismos criterios: mismo número de ejecuciones, mismo problema, misma medida de error. Es necesario ser cuidadoso en el cumplimiento de las condiciones necesarias para realizar un test paramétrico, dado que puede llevar a conclusiones erróneas de no hacerlo.

Cuando se comparan dos modelos si se cumplen estas condiciones, se aplica el test paramétrico t-test, en caso contrario debería aplicarse el test de Wilcoxon. El t-test supone que ambas distribuciones siguen una distribución normal o de idénticas varianzas y evalúa las diferencias entre las medias (errores medios) de dos modelos. En este caso, la hipótesis nula es $\mu_1 - \mu_2 = 0$. Mientras que el test Wilcoxon evalúa las diferencias entre las medianas sin hacer ningún tipo de suposición, siendo la hipótesis

nula $m_1 - m_2 = 0$. El resultado de estos test puede indicar que los errores son iguales (se acepta la hipótesis nula) y por tanto se elegiría el modelo más sencillo, con menor número de parámetros o puede indicar que los errores son diferentes (se rechaza la hipótesis nula), en este caso se elegiría el modelo con el error más pequeño.

Si se van a comparar más de dos modelos para resolver un problema, no es correcto utilizar simplemente un t-test para comparar cada par de modelos, puesto que la posibilidad de detectar incorrectamente una diferencia significativa aumenta con el número de comparaciones. En este caso, es necesario utilizar un análisis de varianza (ANOVA) o un test de Friedman [216], tal y como se muestra en la Figura 23, dependiendo de si se puede utilizar un test paramétrico o no.

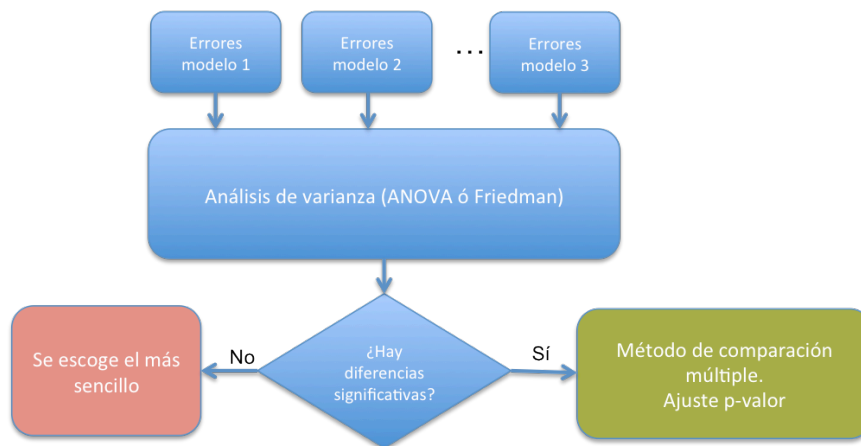


FIGURA 23. DIAGRAMA DE FLUJO DE UN ANÁLISIS DE VARIANZA

ANOVA es por tanto, un test estadístico utilizado para medir las diferencias entre más de dos algoritmos en base a los resultados de cierto medidos como rendimiento al ser aplicados en el mismo conjunto de datos. En este caso, la hipótesis nula que se comprueba es que todos los algoritmos tengan el mismo rendimiento. En caso de rechazo de la hipótesis nula, se realiza un test post-hoc (del latín después de) para encontrar cuál o cuáles difieren y con qué magnitud. Los más habituales son el test de Tukey [217] y el de Dunnett [218].

El test de Friedman establece un ranking entre los algoritmos que se están comparando de manera que el que obtenga un valor más bajo es el que ha obtenido un mejor rendimiento. Una vez utilizado cualquiera de estos dos test de varianza, si el test

concluye que hay diferencias, es necesario utilizar un método de comparación múltiple, conocido como un análisis post-hoc, en caso contrario se elige el más sencillo de los que sean significativamente mejores. Entre los métodos de comparación múltiple para ajustar el p-valor que se pueden utilizar se encuentran los métodos Tukey, Scheffé [219], Bonferroni-Dunn [220] o los más potentes tests de Holm [221], Hochberg [222] y Hommel [223]. En concreto en [212] se concluye que el test de Holm siempre es preferible al ajuste Bonferroni-Dunn, y además, el test de Hochberg se recomienda utilizar de manera conjunta con el test de Holm ya que es aún más potente.

En muchos casos, la condición de normalidad cuando se comparan diferentes ejecuciones de varios clasificadores aplicado a un único problema no se cumple. Y sucede también que la heteroscedasticidad es dependiente del número de algoritmos que se comparen, dado que comprueban las varianzas de todas las poblaciones [212]. Es crucial comprobar si se cumplen las condiciones adecuadas y aplicar el test correcto, ya que no es raro que los test paramétrico y no paramétrico puedan contradecirse.

A pesar de que según el Teorema del Límite Central [224], que dice que la suma de muchas variables aleatorias idénticamente distribuidas tiende a una distribución normal, se debe ser cauto a la hora de proponer un estudio con demasiadas ejecuciones, ya que también es cierto que un número de las mismas elevado tiene efectos negativos sobre el test estadístico. Se propone en [212] que el número de ejecuciones se calcule como $N = a \cdot k$ donde k es el número de algoritmos que se comparan y $a \geq 2$.

En el caso de que el número de algoritmos a comparar no supere los 4 o 5, en [207] se recomienda utilizar, además del test de Friedman, el test de Friedman Alineado o el test de Quade [225]. Difieren del test de Friedman en la forma que calculan el ranking y pueden ofrecer mejores resultados dependiendo de las características del diseño experimental seguido.

4.2 MEJORAS PROPUESTAS EN EL PROCESO DE ANÁLISIS BASADO EN TEXTURAS

El modelo formalizado en esta Tesis Doctoral automatiza las fases del diseño experimental (ver Figura 20) que se deben seguir para realizar un análisis de texturas: extracción de los datos, preprocesado de los datos, aprendizaje y selección del mejor modelo.

Los dos puntos en los que es más factible realizar nuevas propuestas que intenten mejorar los resultados de las técnicas existentes se encuentran en la fase de aprendizaje del diseño experimental (ver Figura 20). Más en concreto durante el proceso de entrenamiento del modelo y durante el proceso de reducción de dimensionalidad.

Durante el proceso de entrenamiento del modelo se propone la inclusión de dos nuevos kernels, combinación de un *kernel* global y otro local, calculados sobre el mismo subconjunto de datos y las posteriores combinaciones lineales de los mismos según la aproximación MKL (recordando que realiza combinaciones lineales de *kernels* sencillos calculados sobre diferentes subconjuntos de datos). Durante el proceso de selección de variables se propone además la utilización de seis nuevas medidas de correlación para la aproximación *filter* de selección de variables. Posteriormente se entrará a explicar con más profundidad los fundamentos matemáticos que respaldan estas modificaciones.

Estas mejoras se integrarán sobre técnicas basadas en kernel para selección de variables ya publicadas en la literatura [121]. Se espera que las aproximaciones propuestas mejoren los resultados obtenidos con otras técnicas estado del arte para la selección de variables con técnicas basadas en *kernel*. La inclusión de dos nuevos kernels que mejoren los resultados obtenidos afectará a cualquier técnica basada en kernel sobre la que se apliquen siguiendo lo que se muestra en la Figura 24, en la que se puede observar que el núcleo de FSMKL y MKL es un SVM, al igual que el del SVM-RFE. Por lo tanto se podrá generalizar que la propuesta de mejora funcionará mejor

para cualquier problema de análisis de texturas en imagen en el que se obtengan los mejores resultados con SVM.



FIGURA 24. PROPAGACIÓN DE LAS MEJORAS ENTRE TÉCNICAS QUE EMPLEAN SVM

Se validarán los resultados obtenidos en el siguiente capítulo de esta Tesis Doctoral.

4.2.1 DESARROLLO DE NUEVOS KERNELS

La propuesta de desarrollo de nuevos kernels se centra en el proceso de entrenamiento del modelo que se realiza dentro de la fase de aprendizaje del diseño experimental propuesto, ver Figura 25.

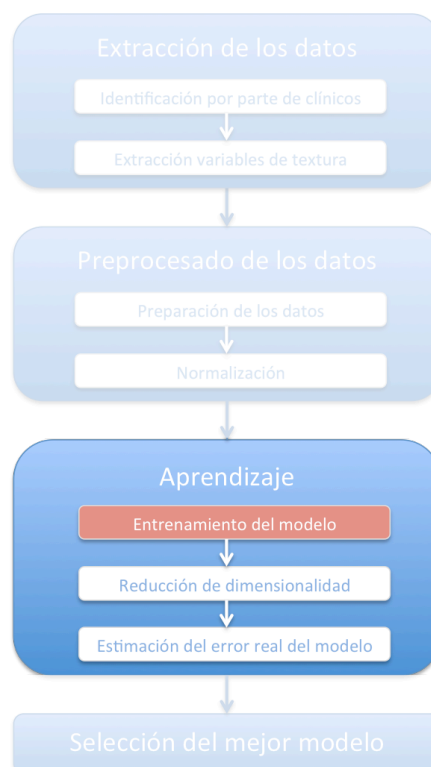


FIGURA 25. FASE DEL DISEÑO EXPERIMENTAL EN LA QUE SE PROPONE LA INCLUSIÓN DE NUEVOS KERNELS

Recordando todo lo expuesto en el capítulo de Fundamentos, dos kernels que están correctamente definidos en el espacio de Hilbert y que cumplan con las condiciones de Mercer [81], pertenecen a la clase de kernels permisibles [6]. La clase de kernels permisibles cumple, entre otras propiedades, que permiten la nueva generación de kernels permisibles a partir de miembros de la clase. A modo de ejemplo si $K_1(x_i, y_i)$ y $K_2(x_i, x_i)$ son dos kernels de la clase de kernels permisibles, los siguientes kernels también lo son

$$(EQ. 23) \quad K_1(x_i, y_i) = cK_1(x_i, y_i) \text{ con la constante } c > 0$$

$$(EQ. 24) \quad K(x_i, y_i) = K_1(x_i, y_i) + K_2(x_i, y_i)$$

$$(EQ. 25) \quad K(x_i, y_i) = K_1(x_i, y_i)K_2(x_i, y_i)$$

$$(EQ. 26) \quad K(x_i, y_i) = f(x_i)K_1(x_i, y_i)f(x_j)$$

$$(EQ. 27) \quad K(x_i, y_i) = \exp(K_1(x_i, y_i))$$

Las características particulares de cada *kernel* tienen lógicamente un gran impacto en los resultados que se obtengan. Que dichos *kernels* tengan una buena capacidad de ajuste a los datos, no quiere decir que con eso se obtenga un mejor valor de generalización para el modelo. Es por esto que la combinación de *kernels* dentro de SVM, en este caso un *kernel* global (lineal) y uno local (RBF) puede llevar a mejorar el error de generalización del modelo.

En los *kernels* globales, los puntos que están muy alejados de los puntos de test tienen un gran efecto en los valores del *kernel*, mientras que en los kernels locales, sólo los puntos que están realmente cerca de los puntos de test tienen un gran efecto en los *kernels* [80]. Por tanto, todos los kernels que están basados en una función de distancia son *kernels* locales, mientras los que están basados en el producto vectorial son *kernels* globales [226].

Esta Tesis propone el uso de dos nuevos *kernels* combinación de un kernel local y otro global, de tal manera que se calcule sobre cada subconjunto particular de datos:

- Un nuevo kernel suma del *kernel* gaussiano y del *kernel* lineal.
- Un nuevo kernel producto del *kernel* gaussiano y del *kernel* lineal.

Dado que los nuevos *kernels* cumplen con las (Eq. 24), (Eq. 25) y lo expuesto en el capítulo 2, estas operaciones están permitidas y lo que se consigue con estas dos nuevas aproximaciones es aumentar la complejidad del kernel generado. Estas combinaciones reescalan la geometría del espacio y por tanto, cambian los pesos relativos asignados a diferentes funciones lineales [73]. Se integrará con la técnica FSMKL, por lo que luego se podrán obtener combinaciones lineales de estos nuevos kernels propuestos.

Combinar un *kernel* local con uno global puede hacer que el *kernel* local reduzca la fluctuación en los valores de predicción del *kernel* global mientras que se incrementa el ajuste del valor de acierto y generalización en el local. En cierta manera, al añadirle un término lineal al *kernel* RBF, se está ayudando al *kernel* a reproducir un comportamiento global de la función. Los kernels locales son sensibles al número de dimensiones [227], por lo que es de esperar que su uso en un sistema de selección de variables obtenga mejores resultados cuanto menor número de variables de ellas seleccione.

Se ha mencionado con anterioridad que el *kernel* RBF es uno de los más utilizados por sus buenos resultados de clasificación y además, los SVM con este kernel han demostrado ser unos clasificadores consistentes [228]. Recordando que se puede definir un kernel RBF como

$$(EQ. 28) \quad K(x_i, y_i) = e^{-d^2(x_i, x_j)}$$

donde la elección más habitual es seleccionar la distancia Euclídea. Tomando como ancho del *kernel* un valor $\sigma > 0$ y siguiendo la ecuación se tiene

$$(EQ. 29) \quad d^2(x_i, x_j) = \frac{1}{\sigma^2} (x_i - x_j)^T (x_i - x_j)$$

Es posible que una medida más ajustada que la distancia Euclídea pueda mejorar los resultados debido a que se trata de una medida que no tiene en cuenta la correlación entre variables por lo que esta Tesis propone el estudio de la distancia de Mahalanobis, que es una generalización de la distancia Euclídea definida por la siguiente ecuación

$$(EQ. 30) \quad d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)}$$

siendo M la matriz de covarianzas. Se ha elegido esta medida debido a que cuanto más correlacionadas están dos variables, más se reduce. Se verá esta nueva medida para incluir en un nuevo *kernel* con más detenimiento en el apartado 5.2.5.

Esta Tesis Doctoral propone además el uso de kernels normalizados y centrados sobre los datos. La operación de normalización de un kernel se realiza para incrementar la independencia de todas las entradas, mientras que el centrado sobre los datos, se realiza con el objetivo de mover el origen del espacio de características al centro de masas de los ejemplos de entrenamiento, puede consultarse [73] para más información.

Ya se ha mencionado que se puede obtener un nuevo *kernel* a partir de diferentes operaciones sobre un kernel correctamente definido. Sea $K(x, z)$ un kernel que se corresponde con el mapeo de características Φ , el nuevo *kernel* normalizado se corresponde con la (Eq. 31),

$$(EQ. 31) \quad x \rightarrow \phi(x) \rightarrow \frac{\phi(x)}{\|\phi(x)\|}$$

de tal forma que para un par de puntos, el kernel normalizado \hat{k} viene dado por la (Eq. 32)

$$(EQ. 32) \quad \hat{k}(x, z) = \langle \hat{\phi}(x), \hat{\phi}(z) \rangle = \left\langle \frac{\phi(x)}{\|\phi(x)\|}, \frac{\phi(z)}{\|\phi(z)\|} \right\rangle = \frac{k(x, z)}{\sqrt{k(x, z)k(z, z)}}$$

Se puede definir el centro de masas como el vector $\phi_S = \frac{1}{\ell} \sum_{i=1}^{\ell} \phi(x_i)$, y la norma queda definida por la (Eq. 33)

$$(EQ. 33) \quad \|\phi(x)\|_2^2 = \frac{1}{\ell^2} \sum_{i,j=1}^{\ell} k(x_i, x_j)$$

de manera que la raíz cuadrada de la norma del centro de masas es igual a la media de las entradas del *kernel*, lo que implica que la suma es mayor o igual que cero, siendo igual cuando el centro de masas está situado en el origen de coordenadas [73]. Si se calcula ahora el cuadrado de la distancia de un punto y su media según la (Eq. 34), queda

$$(EQ. 34) \quad \frac{1}{\ell} \sum_{s=1}^{\ell} \|\phi(x) - \phi_s\|^2 = \frac{1}{\ell} \sum_{s=1}^{\ell} k(x_s, x_s) - \frac{1}{\ell^2} \sum_{i,j=1}^{\ell} k(x_i, x_j)$$

De manera que la media del cuadrado de la distancia de los puntos a su centro de masas es la media de los valores de la diagonal del *kernel* menos la media de todos los valores. Por lo que el centro de masas resuelve el siguiente problema de optimización:

$$(EQ. 35) \quad \min_{\mu} \frac{1}{\ell} \sum_{s=1}^{\ell} \|\phi(x_s) - \mu\|^2$$

por lo que el kernel para centrar los datos se puede obtener con la (Eq. 36)

$$(EQ. 36) \quad \hat{k}(x, z) = k(x, z) - \frac{1}{\ell} \sum_{i=1}^{\ell} k(x, x_i) - \frac{1}{\ell} \sum_{i=1}^{\ell} k(z, x_i) + \frac{1}{\ell^2} \sum_{i,j=1}^{\ell} k(x_i, x_j)$$

4.2.2 MÉTRICAS PARA LA SELECCIÓN DE VARIABLES

La propuesta de nuevas métricas para la selección de variables se centra en el proceso de reducción de dimensionalidad que se realiza dentro de la fase de aprendizaje del diseño experimental propuesto, ver Figura 26.



FIGURA 26. FASE DEL DISEÑO EXPERIMENTAL EN LA QUE SE PROPONE LA INCLUSIÓN DE NUEVAS MÉTRICAS

En Seoane et al. [121] la selección de variables se lleva a cabo mediante una aproximación que utiliza un t-test [229] para hacer un ranking de las variables más correlacionadas con la salida objetivo. El uso de esta medida es mejorable en tanto en cuanto no parece que sea la más ajustada para el problema de un análisis de texturas. Se propone la inclusión de nuevas mejoras en el método *filter* de selección de variables según la Figura 27.

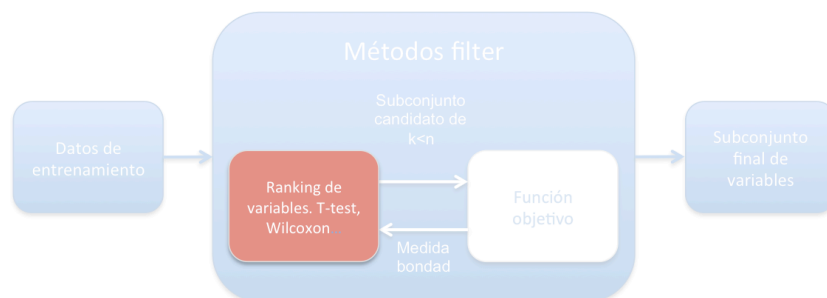


FIGURA 27. MEJORA PROPUESTA EN EL PROCESO DE SELECCIÓN DE VARIABLES

Se espera que estas modificaciones mejoren los resultados debido a que tomar sólo la diferencia en una medida entre las diferentes variables para un proceso de selección de características da información útil que, sin embargo, puede no ser suficiente para discriminar entre variables cuyos valores sean muy cercanos en esa única dimensión. Es por esto que se recurre a métricas de distancia que puedan evaluar el solapamiento entre clases [178].

Esta Tesis Doctoral propone el estudio de las siguientes medidas de distancia para mejorar los resultados: Entropía Relativa, Entropía Relativa Transformada, Bhattacharyya, ROC, Wilcoxon o Fisher. Se describirán a continuación brevemente.

Se considera que el ranking obtenido mediante un t-test para estudiar las diferencias entre dos poblaciones, está más balanceado y es más efectivo [6]. Sin embargo, el valor obtenido mediante la puntuación de Fisher está más influenciado por la magnitud de la diferencia. Es decir, una variable con un buen valor de discriminación será aquella que tiene el valor más grande de separación entre las medias y el menor valor de las desviaciones estándar entre puntos que pertenezcan a las dos clases. Siendo μ_+ y μ_- los valores respectivos para las medias, y σ_+ y σ_- los valores correspondientes a las desviaciones estándar para cada una de las clases (positiva y negativa), se puede definir la distancia Fisher como

$$(EQ. 37) \quad F_{score} = \frac{(\mu_+ - \mu_-)^2}{\sigma_+^2 - \sigma_-^2}$$

Por su parte, la distancia Bhattacharyya supone que existe una distribución Gaussiana o normal entre clases, tiene mucha relación con el error de un clasificador Bayesiano y puede ser definida [178] entre dos clases como

$$(EQ. 38) \quad B_{1,2} = \frac{1}{8} (m_1 - m_2)^T \left(\frac{S_1 + S_2}{2} \right)^{-1} (m_1 - m_2) + A$$

siendo S_i la matriz de covarianzas, m_i la media de la clase y donde A se define como

$$(EQ. 39) \quad A = 0.5 \ln \frac{0.5(|S_1 + S_2|)}{\sqrt{|S_1| + |S_2|}}$$

Puede verse como una aproximación asintótica de una distancia de Mahalanobis y su interpretación geométrica es un arco de circunferencia máxima entre dos puntos de una esfera [230].

La distancia Wilcoxon, también conocida como distancia Mann-Whitney es una medida de distancia que se calcula para estudiar la correlación entre las variables y la clase, usando el conocido test no paramétrico de Wilcoxon. Se utiliza como alternativa no paramétrica al t-test cuando no se puede suponer la normalidad de los datos.

Es posible estudiar, y por lo tanto medir, mediante la curva ROC la capacidad de discriminación de una variable con respecto de la clase de pertenencia.

La distancia entropía relativa es una medida de separación entre clases. Una medida de distancia no debe ser dependiente sólo de la diferencia entre las medias, sino que debe serlo también entre las varianzas. La distancia podría tener valores elevados para datos con medias iguales y aún así, que las varianzas fuesen muy diferentes. Esta medida permite, por tanto, separar entre clases cuyas medias coincidan [178] y viene dada por la siguiente ecuación

$$(EQ. 40) \quad E_{i,j} = \frac{1}{2} \left(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2} - 2 \right) + \frac{1}{2} (\mu_i - \mu_j)^2 \left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2} \right)$$

siendo $\sigma_{i,j}$ las varianzas y $\mu_{i,j}$ las medias.

Por último, con la distancia entropía relativa transformada al igual que sucedía con la media anterior, se obtiene una relación directa entre la entropía o divergencia y el error de Bayes, que en este caso es el error mínimo que se obtiene en la categorización de ejemplos entre clases. Está claro que esto es lo que se busca cuando se pretende utilizar una medida de separación entre clases. Sin embargo, esta relación directa de la entropía con el error de Bayes no se puede aplicar para distribuciones más generales [178], de manera que pequeñas diferencias en las medias pueden producir grandes cambios en la entropía [231,232], que sin embargo, no influyen en los errores de clasificación. En todo caso, para solucionar esto, se propone una medida de entropía relativa transformada que se define como la (Eq. 41)

$$(EQ. 41) \quad \hat{E}_{i,j} = 2 \left(1 - \exp \left(- \frac{\frac{1}{2} \left(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2} - 2 \right) + \frac{1}{2} (\mu_i - \mu_j)^2 \left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2} \right)}{8} \right) \right)$$

5. PRUEBAS Y RESULTADOS

En este capítulo se presentarán las pruebas desarrolladas y los resultados obtenidos siguiendo el diseño experimental propuesto en el capítulo anterior. Se va a comprobar cuantitativamente lo acertado de las mejoras que se han propuesto, para lo que será necesario establecer un marco comparativo que permita evaluarlas contra las técnicas del estado de la cuestión, de tal manera que una vez se conozca la técnica del estado de la cuestión basada en *kernel* que mejores resultados obtiene, se apliquen sobre ella y se estudie su impacto estadístico.

Existe una gran dificultad para replicar los resultados obtenidos por las diferentes aproximaciones consideradas y comentadas en el estado de la cuestión, debido en gran parte a que no existe un *ground truth* o un atlas para las imágenes de electroforesis bidimensional en el que se indiquen las coordenadas de las proteínas utilizadas por otros métodos. Se propone precisamente en esta tesis seguir el diseño experimental descrito en el capítulo anterior, lo que permitirá una comparativa justa de los resultados obtenidos frente a los proporcionados por las técnicas basadas en *kernel* expuestas en el estado de la cuestión para selección de texturas en imagen biomédica. Como se ha visto, este diseño experimental incluye la fase de extracción de los datos en la que se seleccionan, mediante un proceso que se explicará a continuación, las proteínas. Dichas proteínas servirán para generar un conjunto de datos común a estudiar por todas las técnicas que se probarán.

Se obtendrán inicialmente los resultados para los modelos de referencia utilizando para ello las técnicas que mejores resultados ofrecen a la vista de la bibliografía y que han sido descritas previamente: SVM y MKL. Estos modelos de referencia servirán para comparar los resultados obtenidos una vez realizada la selección de variables y la aplicación de las mejoras propuestas con respecto a los obtenidos a partir de toda la información disponible. Como se indicó al comienzo del capítulo 5, el término *modelo de referencia* indica que ha sido necesaria la generación de diferentes *modelos* de clasificación basados en *kernel*, ampliamente usados y conocidos

en el dominio del aprendizaje máquina, utilizados como *referencia* para comparar los resultados.

Se generarán además los resultados para las técnicas *wrapper* bioinspiradas para selección de variables que han dado los mejores resultados en el estudio de la bibliografía: GA-SVM, PSO-SVM. También, dentro del proceso de selección de variables, se generarán los modelos de referencia para una aproximación *embedded* como SVM-RFE y una aproximación *filter* como FSMKL que se han mostrado igualmente como las más eficaces a la hora de realizar esta tarea. El objetivo de estas pruebas será determinar cuál de las técnicas ofrece los mejores resultados con los datos disponibles para así aplicar las mejoras propuestas sobre dicha técnica.

Por lo tanto, gracias a la obtención de estos modelos se podrá realizar una comparación con los resultados obtenidos por la propuesta presentada en esta Tesis Doctoral así como valorar estadísticamente la significancia de los mismos.

Las implementaciones de los diferentes métodos se han desarrollado en Matlab, excepto para la de SVM-RFE que se ha utilizado el paquete estadístico R [233] y la librería Caret [234]. Los test estadísticos se han hecho también con R, así como las gráficas de las curvas ROC utilizando el paquete pROC [235] para R. Los test no-paramétricos se realizaron según [207,210,212,214]. Las pruebas se han realizado sobre los mismos datos, requisito indispensable para poder aplicar los test estadísticos comentados en el apartado 4.1.4.

Los experimentos descritos en esta Tesis Doctoral comparan diferentes ejecuciones de algoritmos de aprendizaje máquina basados en kernels, con y sin selección de características, para resolver el problema de identificación de proteínas en imágenes de electroforesis bidimensional.

Para comparar las medidas de acierto entre las diferentes técnicas se utilizará un proceso de validación cruzada con la AUROC como medida de acierto. Se repetirá este proceso 10 veces para asegurar que el reparto del proceso de validación cruzada no afecta a los resultados obtenidos, siguiendo la regla de oro mencionada en el apartado anterior y propuesta por [212]

Se utilizará un método de contraste de hipótesis para comprobar si existe diferencia estadística significativa entre las distintas ejecuciones de los diferentes algoritmos sobre el mismo conjunto de datos. Las diferencias son consideradas significativas cuando el p -valor es menor a 0.05 (nivel de confianza es mayor del 95%).

Por todo lo comentado anteriormente la estrategia que se va a seguir para el desarrollo de las pruebas en esta Tesis Doctoral es la siguiente.

- El personal clínico seleccionará las proteínas para ser estudiadas.
- Se utilizará Mazda para la extracción de las variables de textura de las ROIs seleccionadas.
- Se generará el dataset.
- Se realizarán 10 experimentos con cada modelo de los comentados anteriormente.
- En cada experimento se separarán los datos con un 10-fold *cross validation* para poder comparar el rendimiento de generalización de todos los modelos.
- Por cada modelo y experimento se calculará el valor de generalización del modelo en forma de porcentaje de acierto y AUROC. Se dará además información del número de variables finales y el tiempo que necesito el modelo para llegar a ellas.
- Se compararán de manera visual los experimentos de cada modelo mediante las curvas ROC.
- Se compararán de manera visual todos los modelos utilizando una curva ROC que acumule los resultados de los diez experimentos realizados.
- Se comprobará que tipo de test se puede realizar: paramétrico o no paramétrico. Se comprobará las condiciones de independencia, normalidad y heteroscedasticidad.
- Se realizará un test paramétrico/no paramétrico para comprobar si hay diferencias significativas entre los modelos seleccionados de las diferentes aproximaciones y se seleccionará el mejor en base a dicho test estadístico.
- Con la mejor técnica resultante de este proceso experimental, se realizarán nuevas pruebas para las que se propondrá el uso de nuevos kernels y de nuevas

medidas de distancia para estudiar la correlación entre las variables de textura y la variable de salida.

5.1 OBTENCIÓN DE MODELOS DE REFERENCIA

5.1.1 EXTRACCIÓN DE DATOS

Tal y como se ha especificado en los capítulos previos, trabajar con imágenes de electroforesis bidimensional es una tarea muy complicada debido a que no hay un *ground truth* común, a pesar de que es posible encontrar imágenes publicadas en Internet [141,236]. Otro aspecto que hace que sea difícil trabajar con este tipo de imágenes, desde el punto de vista de una aproximación de visión artificial, es que tanto las proteínas como el ruido de fondo presente en las imágenes es intrínseco a las mismas, parece seguir una distribución Gaussiana [130]. Existe además una gran variabilidad inter- e intra-operador en el análisis manual de estas imágenes [237].

Hunt et al. [238] determinaron que el número mínimo de imágenes requerido para cumplir con las condiciones exigidas por un estudio de proteómica es de 7 u 8. En esta Tesis Doctoral se han utilizado diez imágenes de electroforesis bidimensional con diferentes tipos de tejidos y condiciones experimentales. Las imágenes son las utilizadas por G.-Z Yang (*Imperial College of Science, Technology and Medicine*, Londres) en trabajos previos [239,240] de emparejamiento de imágenes. Para cada una de las imágenes, dos clínicos distintos realizaron un proceso de marcado dos veces escogiéndose al final para cada imagen 100 regiones de interés comunes a los dos clínicos y consensuadas entre ellos, con el objetivo de construir un conjunto de datos compuesto por 1000 ejemplos y 274 variables de texturas. La variabilidad inter- e intra-operador en concreto, separado por clínico/proceso, con media y desviación estándar del número de proteínas seleccionadas por cada clínico y cada uno de los dos procesos, para las imágenes utilizadas es la mostrada en la Tabla 3.

TABLA 3. VARIABILIDAD INTER- E INTRA-OPERADOR DEL PROCESO DE MARCADO DE PROTEÍNAS

		IMÁGENES									
	Proc	1	2	3	4	5	6	7	8	9	10
Cli1	1	404	545	539	545	445	539	565	307	539	565
	2	433	551	527	512	412	533	579	306	533	551
Cli2	1	397	481	541	497	431	511	539	297	505	556
	2	401	475	512	505	429	523	545	300	471	523
Media		408,75	513	529,75	514,75	429,25	526,50	557	302,50	512	548,75
Desviación		16,41	40,56	13,35	21,07	13,52	12,26	18,04	4,79	31,09	18,11

Las regiones de interés que caracterizan cada proteína, son ligeramente más grandes que la superficie más oscura de la misma. El ojo humano no es capaz de identificar los cambios sutiles que se producen en el nivel de gris a nivel de píxel. Por tanto, una vez que el clínico ha seleccionado lo que considera la región de interés para una proteína, se selecciona para estudiar una región ligeramente más ancha. Esta suposición se toma ya que las variables de textura que pueden caracterizar a la proteína al realizar un análisis de texturas mediante técnicas de visión artificial, pueden estar tanto en los valores de gris más cercanos al negro, como a los valores más cercanos al blanco. Como se dijo con anterioridad, las proteínas semejan la forma de una curva Gaussiana en donde, idealmente, el centro de la proteína debería encontrarse en el centro de la zona más oscura. Esta aproximación previene la pérdida de información causada por no tener en consideración los valores más bajos de la representación de la proteína invertida (valores de gris más cercanos al blanco), lo más habitual en visión artificial, es representar con valor 0 al negro y con 1 al blanco.

Lo comentado anteriormente se muestra de manera gráfica en la Figura 28: en a) se puede ver una proteína de un proceso real de electroforesis bidimensional, en b) y en pseudocolor se puede ver la representación de las intensidades de los niveles de gris de la proteína en donde los niveles más oscuros de la misma se corresponden con los valores entre rojo y amarillo, en c) se puede ver también en pseudocolor los valores de niveles de gris que tendría una proteína que pudiese ser representada con una curva Gaussiana perfecta, en d) se puede ver la proteína marcada como una curva Gaussiana.

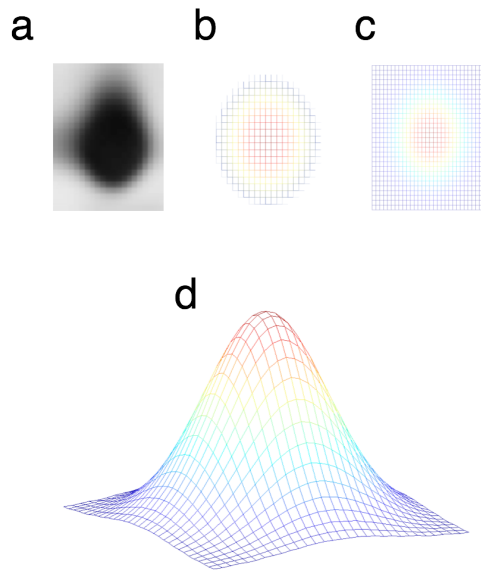


FIGURA 28. REPRESENTACIÓN IDEAL DE UNA PROTEÍNA CON FORMA DE CAMPANA DE GAUSS

Existen diferentes aproximaciones [241-244] basadas en el estudio de la forma de las proteínas, o que hacen uso de operadores habituales como sucesivas umbralizaciones de la imagen, derivadas, Laplacianas, algoritmos *watershed* o determinación de máximos locales para identificar proteínas. Las mayores desventajas que tienen asociadas estos métodos es que son sensibles al ruido de alta frecuencia y la imagen, en general, es suavizada antes de hacer el proceso de detección automáticamente. Este suavizado reduce la resolución de la imagen, por lo que el acierto en la detección de proteínas de reducido tamaño o solapadas, se ve afectado.

Estas aproximaciones basadas en la forma de las proteínas, tienen dificultades dado que el aspecto que puede tener una proteína real en una imagen de electroforesis bidimensional es como el mostrado en la Figura 29: en a) puede verse la representación en tres dimensiones de una proteína marcada por un clínico, en b) y en c) puede verse el aspecto de la proteína al realizar cortes en el eje horizontal y vertical por el centro de la misma. Puede verse que mientras que en c) la proteína puede aproximarse a una curva Gaussiana, no sucede lo mismo en b). También puede verse como el centro de la proteína no se encuentra en la zona con valores más cercanos al negro.

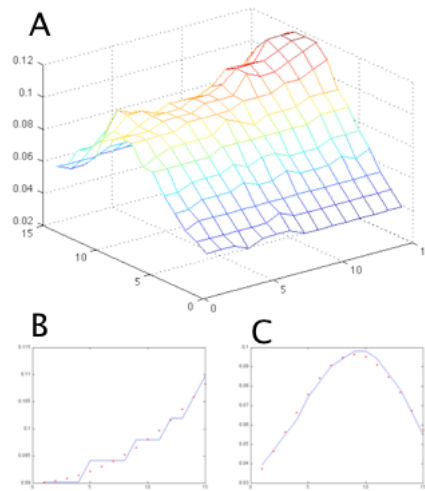


FIGURA 29. ASPECTO REAL DE UNA PROTEÍNA RESULTADO DE UN PROCESO DE ELECTROFORESIS BIDIMENSIONAL

Por lo que respecta al empleo de texturas, las variables de textura fueron calculadas con un software especial llamado Mazda [180-182]. Existen múltiples trabajos en los que se demuestra la efectividad de la herramienta para extraer variables texturales en diferentes tipos de imágenes médicas [4,5,156,183-188]. Puede verse un resumen de las variables utilizadas en la Tabla 4.

TABLA 4. VARIABLES DE TEXTURA EXTRAÍDAS CON MAZDA

Grupo	VARIABLES	Num. de Var.
Histograma	Media, varianza, skewness, kurtosis, percentiles 1%, 10%, 50%, 90% y 99%	9
Gradiente Absoluto	Media, varianza, skewness, kurtosis y porcentaje de píxeles con gradiente distinto de cero	5
Matriz de Longitud de Carreras	Longitud de carrera no-uniforme, no-uniformidad de niveles de gris, énfasis de carrera larga, énfasis de carrera corta, fracción de imagen en carreras	20
Matriz de Coocurrencia	Segundo momento angular, contraste, correlación, suma de cuadrados, momento diferencial inverso, suma promedio, suma varianzas, suma entropías, entropía, diferencia varianzas, diferencia entropías	221
Modelo Autoregresivo	Theta: vector de parámetros del modelo (4 parámetros) Sigma: desviación estándar del ruido	5
Wavelet	Energía de los coeficientes wavelet en sub-bandas en sucesivas escalas; máximo de 4 escalas, cada una con 4 parámetros	14

Puede consultarse un manual de uso de la herramienta en [245]. A continuación se presentará una breve introducción con el significado de cada una de las variables mencionadas en la Tabla 4.

5.1.1.1 VARIABLES DE TEXTURA BASADAS EN EL HISTOGRAMA

El histograma de los niveles de gris de una imagen es una función que se utiliza para contar el número de píxeles observados con un nivel de gris específico. Conforman los operadores de primer orden, presentados en el capítulo 2 ya que se calculan sin tener en cuenta relaciones entre píxeles vecinos. Las variables que se calculan son:

- *Media*: el promedio de los niveles de intensidad de la imagen.
- *Varianza*: describe cuánto de lejos se encuentran los valores de la media. Determina la aspereza de la imagen.
- *Skewness*: describe la simetría del histograma con respecto de la media.
- *Kurtosis*: describe lo plano que es un histograma. Un valor negativo indica una distribución plana y uno positivo describe un pico en la distribución, es decir, indica como de uniforme es una distribución de los niveles de gris con respecto a una distribución normal.

5.1.1.2 VARIABLES DE TEXTURA BASADAS EN EL GRADIENTE

Un gradiente se puede definir como el cambio direccional de los niveles de gris de una imagen, de manera que un valor alto de gradiente representa grandes cambios entre los niveles de gris entre los tonos claros y oscuros y un valor bajo, cambios suaves.

- *Media del gradiente absoluto*: mide la media de las variaciones en los niveles de gris en la imagen.
- *Varianza del gradiente absoluto*: describe lo alejados que están los valores de la media.
- *Gradiente skewness y kurtosis*: medidas utilizadas para comprobar la simetría del gradiente.

5.1.1.3 VARIABLES DE TEXTURA BASADAS EN LA MATRIZ DE LONGITUD DE CARRERA

La matriz de longitud de carreras de niveles de gris o *run-length matrix* contiene información acerca del número de veces que un nivel de gris aparece en la imagen con

una determinada longitud de carrera. Estas matrices son calculadas, al igual que las GLCM en diferentes ángulos.

- Énfasis largo y corto de carrera: miden la proporción de carreras con longitudes largas y cortas. Si la imagen ha sido suavizada, el énfasis largo de carrera tendrá un valor alto.
- No-uniformidad de los niveles de gris: se utiliza para calcular como de uniforme son las carreras entre los niveles de gris. Tiene valores bajos cuando la distribución es uniforme.
- Longitud de carrera no-uniforme: mide la distribución de los niveles de gris entre las carreras largas.
- Fracción de imagen en carreras: describe la fracción de píxeles de la imagen que son parte de cualquiera de las carreras de la matriz.

5.1.1.4 VARIABLES DE TEXTURA BASADAS EN LA MATRIZ DE COOCURRENCIA

La GLCM se obtiene a partir de funciones de probabilidad condicionadas entre dos píxeles en función de una distancia y un ángulo concretos. Se utiliza para medir las dependencias espaciales que caracterizan una textura. Esta matriz es cuadrada y de lado el número de niveles de gris de la imagen. Para cada una de las direcciones se obtiene una matriz. No es una matriz simétrica, para que lo sea habitualmente se le suma su matriz traspuesta y se divide por dos, por lo que GLCM es incapaz de detectar rotaciones de 180° y es por lo que sólo se calcula con ángulos que van de 0° a 180° , los cuatro más usuales ya fueron mencionados anteriormente. Si se promedian las 4 matrices obtenidas, el operador resultante es invariante a la rotación.

Sean (x_1, y_1) y (x_2, y_2) un par de píxeles en la imagen separados por una distancia d con un ángulo α . Si los niveles de gris de cada par de píxeles son i y j respectivamente, se puede definir la probabilidad de que los niveles de gris i y j estén presentes en la imagen a una distancia d y con un ángulo α como $P(i, j, d, \alpha)$. La matriz de GLCM se obtiene a partir de estas funciones de probabilidad condicional.

A partir de GLCM se calculan diferentes valores:

- Segundo momento angular: cuanto más homogénea sea la imagen, el valor será más alto. Este valor también es conocido como Energía.
- Contraste: medida de las variaciones locales de intensidad en la imagen.
- Correlación: medida de las dependencias lineales de los niveles de gris de la imagen.
- Suma de cuadrados: mide la varianza en la GLCM.
- Momento diferencial inverso: mide la homogeneidad de la imagen. Cuánto más suavizada esté la imagen, mayor será el valor.
- Suma promedio: este operador calcula la media de la suma del valor de intensidad de gris de dos píxeles en la imagen.
- Suma varianzas: se calcula la varianza de la suma promedio.
- Entropía: mide el grado de desorden de la imagen. El mayor valor de entropía se obtiene cuando todas las probabilidades de aparición son iguales.
- Suma de entropías.
- Diferencia de varianzas
- Diferencia de entropías

Se puede consultar en los Anexos las fórmulas matemáticas que se utilizan para el cálculo de cada una de estas variables.

5.1.1.5 VARIABLES DE TEXTURA BASADAS EN MODELOS AUTOREGRESIVOS

Este tipo de operador de texturas asume que existe una interacción local entre píxeles en la imagen y describe cada valor de intensidad como una suma ponderada de los valores de los píxeles vecinos. Mazda utiliza para cada región de interés cinco parámetros: los coeficientes de los 4 píxeles vecinos (Θ , θ) y el error estándar del ruido (Σ , σ). Las texturas con una alta rugosidad, tendrán valores muy similares en sus píxeles vecinos, al contrario de lo que sucede con texturas más finas.

5.1.1.6 VARIABLES DE TEXTURA BASADAS EN WAVELET

Un análisis wavelet presenta la imagen como un conjunto independiente de canales de frecuencia orientados espacialmente. Se toma la señal de la imagen y se aplica la

transformada wavelet a través de un filtro paso bajo y un paso alto, con lo que la señal es reducida y descompuesta para incrementar la resolución en frecuencia. Mazda calcula la energía en sub-bandas de la wavelet Haar. Los operadores de textura basados en wavelet son independientes de factores como son las rotaciones o la resolución de la imagen.

5.1.2 PREPROCESADO

En el caso de un análisis de texturas, dónde el conjunto de los datos de entrada es tan heterogéneo, cobra especial importancia esta etapa, ya que las variables pueden tener valores muy distintos (de varias órdenes de magnitud) y es necesario normalizar los datos.

En el caso de las técnicas de fusión de datos (ver sección 1.5), no se eliminan las variables correlacionadas, ya que es ésta correlación la que permite sacar más información útil de las relaciones entre variables, para alcanzar un conocimiento de un nivel superior, por lo que para el problema del análisis de variables de textura propuesto, no se han eliminado las variables correlacionadas.

5.1.3 APRENDIZAJE

Se ha propuesto en la formalización del diseño experimental que dentro de esta fase se realizan el entrenamiento del modelo, la reducción de la dimensionalidad y la estimación del error cometido. Por cuestiones de claridad y dado el volumen de pruebas realizadas, se estudiará cada una de las técnicas por separado. De este modo, para cada una de ellas se entrenará el modelo y se estimará el error real cometido.

Como punto de partida para realizar las comparaciones y evaluar el correcto funcionamiento del proceso de selección de variables de las diferentes técnicas basadas en kernel, se realizarán pruebas con los modelos de referencia SVM y MKL. Se mostrarán después los resultados obtenidos de las pruebas desarrolladas con dos aproximaciones híbridas con metaheurísticas bioinspiradas y técnicas basadas en kernel (GA/PSO-SVM) y dos técnicas específicas basadas en *kernel* (SVM-RFE y FSMKL) para selección de características. Finalmente se mostrará una comparativa entre todas las técnicas y se elegirá la mejor en base a test estadísticos de comparación.

5.1.3.1 TÉCNICAS BASADAS EN KERNEL

5.1.3.1.1 SVM

Se han hecho diferentes experimentos en busca de los mejores parámetros con el kernel RBF y polinomial explicados en el capítulo 2. El dominio de los parámetros para esta técnica se muestra en la Tabla 5.

TABLA 5. SVM: DOMINIO DE LOS PARÁMETROS

Elemento	Dominio
Función de kernel	Lineal y RBF
Sigma del kernel RBF	De 0.5 a 20
Valor de C para el SVM	De 0.01 a 100
Orden del polinomio	De 2 a 10
Método para encontrar el hiperplano	Programación cuadrática

Los mejores resultados se han encontrado realizando diferentes experimentos dirigidos en un proceso de validación cruzada, con el kernel RBF, 2 para el valor de sigma y una C de 100. Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 6, donde se presentan los valores de porcentaje de acierto y AUROC por cada uno de los 10 experimentos realizados. Cada experimento se ha desarrollado con un proceso de *10-fold cross validation*.

TABLA 6. SVM: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación
1	83.50	89.68
2	83.40	89.87
3	83.30	89.95
4	83.70	89.92
5	83.30	89.74
6	83.00	89.76
7	82.80	89.95
8	83.00	89.74
9	82.90	90.01
10	83.20	89.92

Se representan a continuación en la Figura 30 las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo. Las curvas se superponen debido a la estabilidad del método haciendo difícil apreciar las diez curvas individuales. Esta representación se realiza con el ánimo de poder comparar los 5 modelos con una única curva ROC para cada uno de ellos de manera gráfica, facilitando la comprensión de los resultados. La selección del mejor modelo se hará en base a la aplicación de test estadísticos *post-hoc* para comparar los resultados

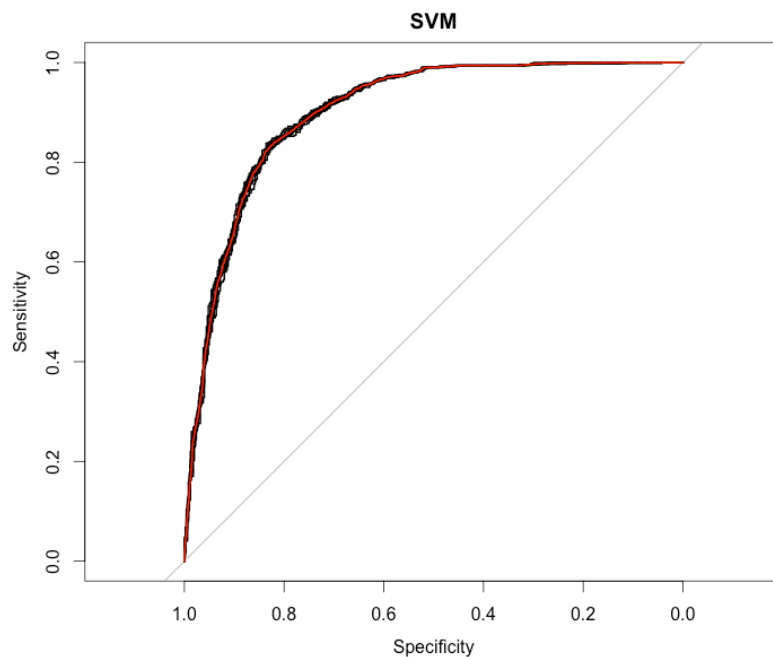


FIGURA 30. SVM: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.1.3.1.2 MKL

Esta técnica busca las mejores soluciones haciendo combinaciones lineales del *kernel* RBF y polinomial explicados en el capítulo 2. El dominio de los parámetros para esta técnica se muestran en la Tabla 7.

TABLA 7. MKL: DOMINIO DE LOS PARÁMETROS

Elemento	Dominio
Sigma del kernel RBF	De 0.5 a 20
Orden del kernel polinomial	De 1 a 4
Valor de C para el SVM	10,100
Criterio de parada	Duality gap
Valor de parada	0.001

Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 8, se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución y número de kernels de la solución final. Cada experimento se ha desarrollado con un proceso de *10-fold cross validation*.

TABLA 8. MKL: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels
1	82.90	89.17	3.10	3
2	82.70	89.38	2.58	2
3	82.70	89.17	2.72	2
4	82.30	89.29	3.28	3
5	83.30	89.23	2.87	3
6	82.80	89.16	3.05	2
7	82.70	89.26	2.63	3
8	83.10	89.33	2.93	3
9	82.60	89.24	2.72	2
10	82.90	89.28	2.63	2

Los parámetros que selecciona para cada uno de estos *kernels* combinación lineal de uno local y otro global son, en el rango entre (1-2) para el valor de sigma en el kernel local y en el rango entre (1-2) para el grado del kernel global. Esta técnica identifica a las variables de wavelet (catorce variables) como las más representativas del problema dado que los 2-3 *kernels* siempre las toman como el grupo de variables con más peso en la clasificación. Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

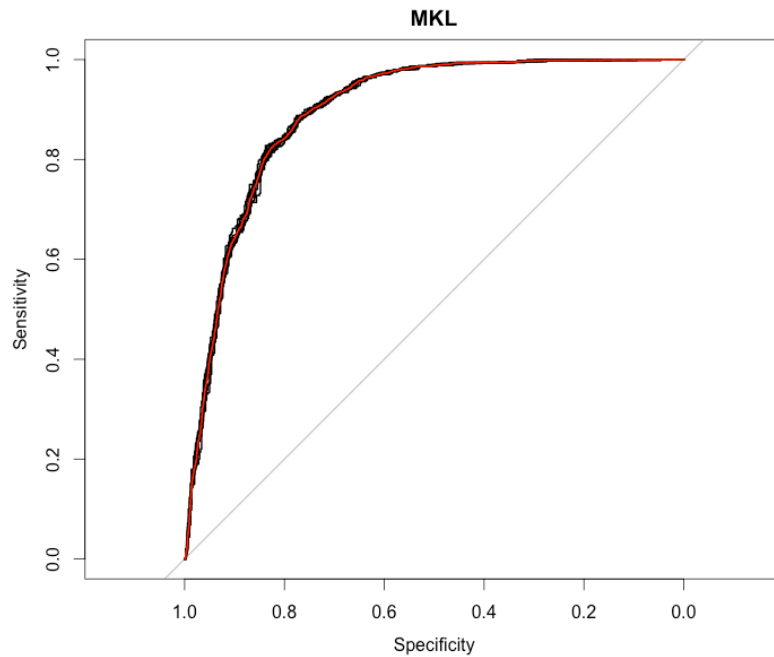


FIGURA 31. MKL: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.1.3.2 REDUCCIÓN DE DIMENSIONALIDAD

5.1.3.2.1 ALGORITMO GENÉTICO

Este tipo de técnica pertenece al grupo de metaheurísticas bioinspiradas que pueden utilizarse para selección de variables. El dominio de los parámetros probados para esta técnica, que se muestra en la Tabla 9, se corresponde al dominio clásico de la literatura.

TABLA 9. GA: DOMINIO DE LOS PARÁMETROS

Elemento	Dominio
Tamaño de la población	De 100 a 250
Elitismo	Del 0 al 2%
Probabilidad de cruce	Del 80 al 98%
Probabilidad de mutación	Del 0.01 al 5%
Operadores de cruce	Un punto, dos puntos, <i>scattered</i> , aritmético, uniforme y heurístico
Función de selección	Ruleta y torneo
Función de mutación	Gaussiana

Se realizaron diferentes experimentos dirigidos, en un proceso de validación cruzada, y la combinación final que resultó dar los mejores resultados tenía una población de 250 individuos, sin elitismo, con una probabilidad de cruce del 95%, una probabilidad de mutación del 2% con operador de cruce *scattered*, selección por torneo y mutación uniforme. La implementación utilizada está basada en el toolbox de optimización de MATLAB, para cualquier aclaración con alguno de los parámetros en concreto puede ser consultada la web oficial del programa en (<http://www.mathworks.es>).

Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 10, donde se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución y número de variables seleccionadas por el modelo en cada uno de ellos. Cada experimento se ha desarrollado con un proceso de *10-fold cross validation*.

TABLA 10. GA: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Núm. de variables
1	89.70	94.29	2104.06	59
2	89.20	94.42	1992.90	62
3	90.30	94.09	1376.19	65
4	89.70	93.88	2136.20	43
5	89.00	94.41	2629.67	50
6	90.10	95.26	1957.21	48
7	90.00	94.49	1749.62	59
8	88.70	93.75	1541.73	66
9	89.20	94.44	2104.38	54
10	90.60	95.18	1951.07	57

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

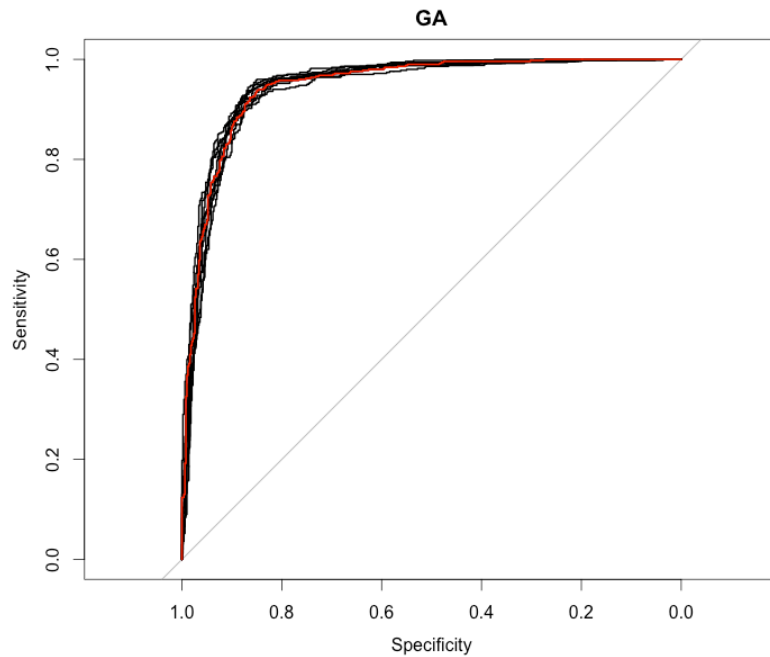


FIGURA 32. GA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS.

5.1.3.2.2 OPTIMIZACIÓN POR ENJAMBRE DE PARTÍCULAS

El algoritmo de optimización por enjambre de partículas pertenece al grupo de metaheurísticas bioinspiradas que pueden utilizarse para selección de variables. Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 11, se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución en validación en el proceso de *10-fold cross validation* y número de variables usadas al finalizar el proceso de selección.

TABLA 11. PSO: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Núm. de variables
1	84.30	88.27	194.39	103
2	81.80	86.85	202.12	122
3	86.40	90.47	384.51	107
4	84.20	88.40	155.57	124
5	83.20	87.66	196.27	128
6	88.80	93.78	848.68	89
7	85.40	90.25	387.86	117
8	80.30	86.65	122.12	122
9	84.60	90.50	352.76	109
10	83.40	88.56	281.31	109

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

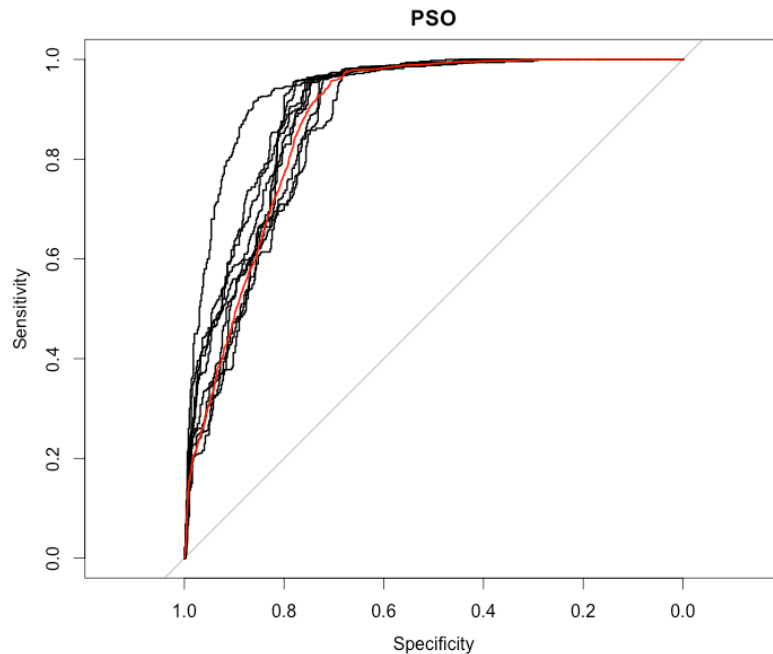


FIGURA 33. PSO: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

Se puede observar como la aleatoriedad en el inicio de las poblaciones de esta técnica genera curvas más separadas entre experimentos.

5.1.3.2.3 SVM-RFE

Dentro de la comunidad científica, es ampliamente utilizada la técnica SVM-RFE para la selección de variables relevantes. Esta técnica utiliza un proceso iterativo y un modelo basado en kernel para la reducción de la dimensionalidad según se explicó con detenimiento en 1.4.2. Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 12, donde se presentan los valores de porcentaje de acierto, AUROC y número de variables usadas al finalizar el proceso de selección. Cada experimento se ha desarrollado con un proceso de *10-fold cross validation*. En este caso no se presenta el tiempo de ejecución del método debido a que las pruebas fueron ejecutadas en el centro de Supercomputación de Galicia, siendo una ventaja por la configuración hardware en comparación con los otros equipos de prueba.

TABLA 12. SVM-RFE: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Núm. de variables
1	93.00	93.43	103
2	93.00	93.43	138
3	89.00	89.78	113
4	93.00	93.02	145
5	92.00	92.61	164
6	90.00	91.05	114
7	93.00	93.23	109
8	93.00	93.83	175
9	93.00	93.02	107
10	89.00	90.98	108

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

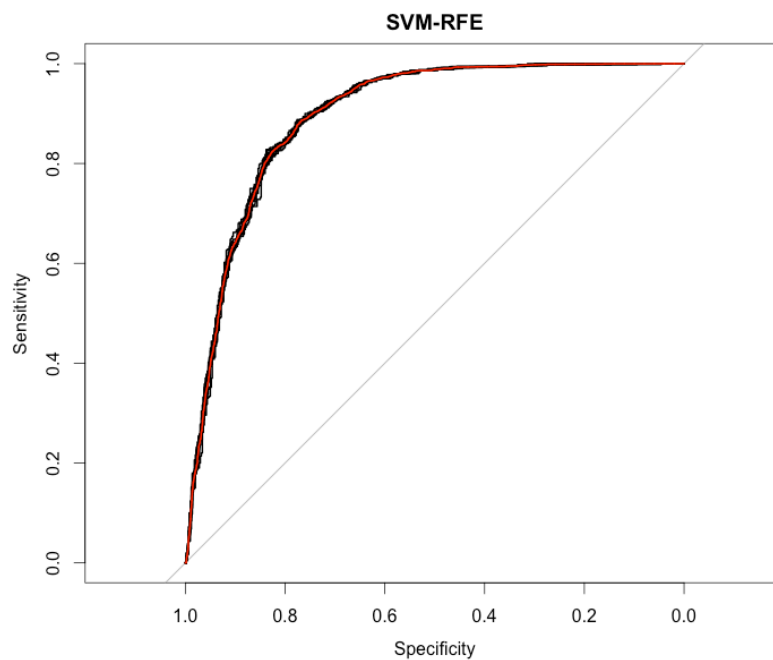


FIGURA 34. SVM-RFE: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.1.3.2.4 FSMKL

Esta técnica busca las mejores soluciones haciendo combinaciones lineales del kernel RBF y polinomial explicados en el capítulo de Fundamentos, a la vez que se queda con las variables más relevantes mediante un proceso que establece un ranking de las

variables más correlacionadas con la clase de salida utilizando un t-test. El dominio de los parámetros para esta técnica se muestra en la Tabla 13.

TABLA 13. FSMKL: DOMINIO DE LOS PARÁMETROS

Elemento	Dominio
Sigma del kernel RBF	De 0.5 a 20
Orden del kernel polinomial	De 1 a 3
Valor de C para el SVM	10,100
Criterio de parada	Duality gap
Valor de parada	0.01

Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 14, se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución en validación en el proceso de *10-fold cross validation*, número de kernels usados y número de variables usadas al finalizar el proceso de selección.

TABLA 14. FSMKL: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	89.70	95.55	126.76	16	31
2	89.70	95.49	157.04	18	22
3	89.70	95.60	150.12	20	35
4	89.50	95.44	131.59	22	22
5	89.30	95.69	148.56	20	23
6	89.30	95.49	126.40	28	24
7	89.40	95.60	142.92	22	22
8	89.40	95.07	143.01	39	35
9	89.60	95.55	119.34	15	25
10	89.30	95.59	119.89	18	22

Los parámetros que selecciona para cada uno de estos *kernels* combinación lineal de uno local y otro global son, en el rango entre (0.5-1) para el valor de sigma en el kernel local y en el rango entre (1-2) para el grado del kernel global.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

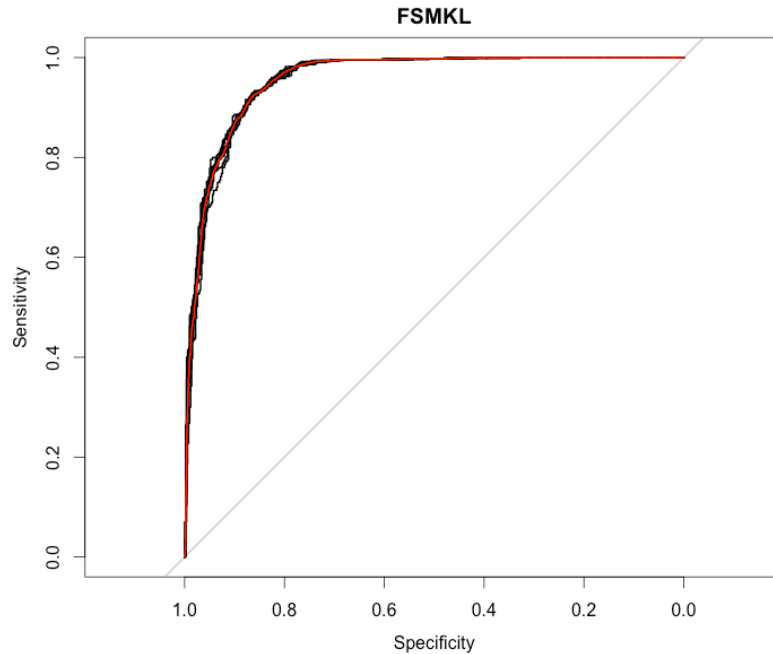


FIGURA 35. FSMKL: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.1.4 SELECCIÓN DEL MEJOR MODELO

Se han probado seis técnicas para resolver el problema de análisis de texturas, de las cuáles una técnica no hace selección de características (SVM), dos son aproximaciones híbridas con metaheurísticas bioinspiradas y técnicas basadas en kernel (GA/PSO-SVM) y tres son técnicas basadas en *kernel* (MKL, SVM-RFE y FSMKL) para selección de características. Se resumen los datos obtenidos en la Tabla 15 y en la Figura 37.

Los resultados muestran que las metaheurísticas bioinspiradas tienen más variabilidad entre experimentos que las técnicas basadas en kernels. El número de variables que seleccionan está entre 60-100, y son las que más tiempo necesitan para encontrar una solución. Los resultados parecen mostrar que terminan en mínimos locales y que no encuentran la mejor solución.

Por otro lado, las técnicas basadas en *kernel* son las que encuentran las variables en el menor tiempo pero, mientras que SVM-RFE necesita en torno a 100 variables para

solucionar el problema, destaca FSMKL que lo hace solamente con cerca de 25 variables de media entre experimentos y MKL que utiliza 14 variables.

Se muestran las ROC calculadas en los apartados anteriores en la Figura 36. Se utilizarán simplemente para dar una aproximación visual a los resultados obtenidos con los diferentes clasificadores, pero la significancia y la selección del mejor modelo se hará en base a la aplicación de los distintos test estadísticos que deban ser aplicados.

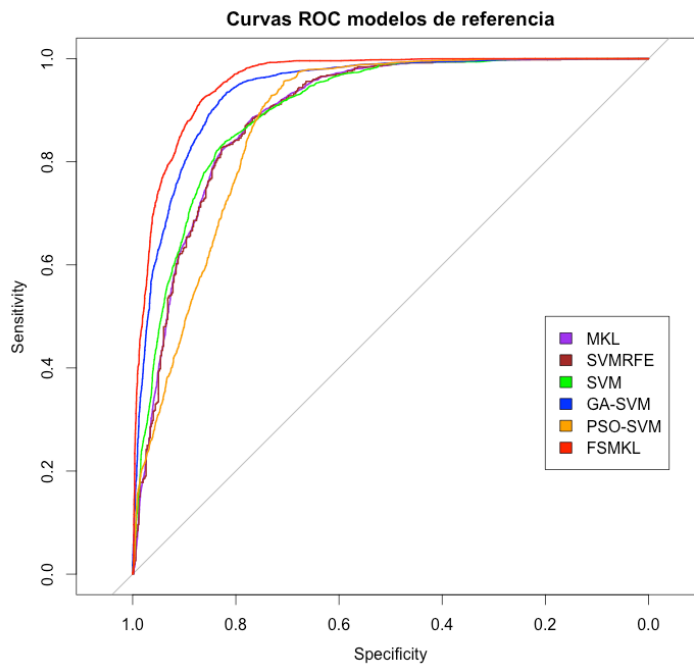


FIGURA 36. CURVAS ROC DE LOS MODELOS DE REFERENCIA

Las curvas ROC mostradas en las secciones anteriores indicaban que el comportamiento durante los 10 experimentos dirigidos realizados para SVM, MKL, SVM-RFE y FSMKL es muy estable. Se puede observar además que la curva ROC calculada con los datos de los 10 experimentos se corresponde con una medida media de ellos, por lo que no se está añadiendo ni quitando información al utilizar esa curva para comparar los modelos. En cuanto a los métodos bioinspirados, son los que parecen comportarse de peor manera, siendo más inestables los resultados obtenidos entre los diferentes experimentos realizados. Particularmente PSO es el que peor se comporta de los dos.

Como se ha explicado en el apartado 4.1.4, es necesario comprobar si se cumplen las condiciones necesarias para aplicar un test paramétrico y poder comparar

las distintas ejecuciones de cada uno de los algoritmos y determinar si existe significancia estadística entre ellos ya que se ha dicho previamente que no se pueden utilizar las curvas ROC para discriminar entre diferentes técnicas.

Recordando, es necesario comprobar que se cumplan las condiciones de independencia, normalidad y heteroscedasticidad.

En este caso, la condición de independencia es obvia ya que se van a comparar diferentes ejecuciones de cada uno de los algoritmos con semillas de inicio generadas aleatoriamente. Para comprobar la condición de normalidad se aplicará el test de Shapiro-Wilk y para comprobar la condición de heteroscedasticidad el test de Bartlett. Una vez se determine si han cumplido todas estas condiciones se podrá aplicar un test paramétrico ANOVA o un test no paramétrico Friedman o variantes, dependiendo de diferentes criterios, como el número de modelos a comparar por ejemplo. Los resultados de acierto y AUROC de validación agrupados a lo largo de los 10 experimentos de los seis modelos son los mostrados en la Tabla 15 y en la Figura 37 se resumen los valores de AUROC gráficamente.

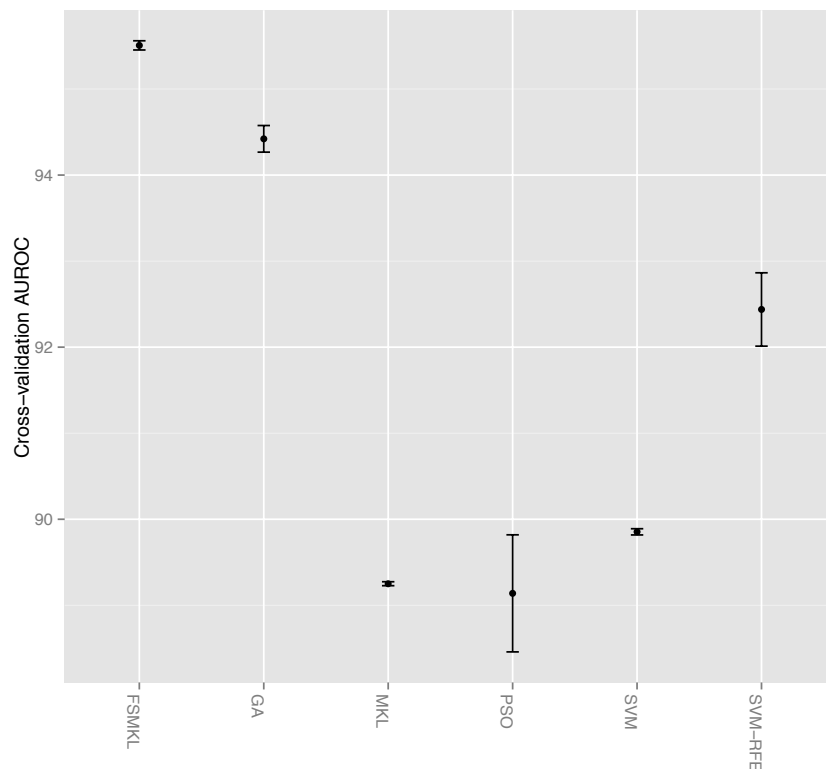


FIGURA 37. MODELOS DE REFERENCIA: RESULTADOS

TABLA 15. MODELOS DE REFERENCIA: RESULTADOS

Exp.	SVM		MKL		GA-SVM		PSO-SVM		SVM-RFE		FSMKL	
	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC
1	83.50	89.68	82.90	89.17	89.70	94.29	84.30	88.27	93.00	93.43	89.70	95.55
2	83.40	89.87	82.70	89.38	89.20	94.42	81.80	86.85	93.00	93.43	89.70	95.49
3	83.30	89.95	82.70	89.17	90.30	94.09	86.40	90.47	89.00	89.78	89.70	95.60
4	83.70	89.92	82.30	89.29	89.70	93.88	84.20	88.40	93.00	93.02	89.50	95.44
5	83.30	89.74	83.30	89.23	89.00	94.41	83.20	87.66	92.00	92.61	89.30	95.69
6	83.00	89.76	82.80	89.16	90.10	95.26	88.80	93.78	90.00	91.05	89.30	95.49
7	82.80	89.95	82.70	89.26	90.00	94.49	85.40	90.25	93.00	93.23	89.40	95.60
8	83.00	89.74	83.10	89.33	88.70	93.75	80.30	86.65	93.00	93.83	89.40	95.07
9	82.90	90.01	82.60	89.24	89.20	94.44	84.60	90.50	93.00	93.02	89.60	95.55
10	83.20	89.92	82.90	89.28	90.60	95.18	83.40	88.56	89.00	90.98	89.30	95.59

Se realizarán a continuación los diferentes estudios estadísticos necesarios para saber si alguno de los modelos propuestos es estadísticamente mejor que las otras técnicas basadas en kernel que se han utilizado en esta Tesis Doctoral. Se propone el

estudio de seis técnicas basadas en kernel: SVM, MKL, GA-SVM, PSO-SVM, SVM-RFE y FSMKL.

El test de Shapiro Wilk revela con resultados ($W=0.8938$ y $p\text{-valor } 7.854E-05$) que los datos no siguen una distribución normal, puesto que $p\text{-valores} < 0.05$ rechazan la hipótesis nula. Se presenta la gráfica Q-Q de la confrontación entre los cuartiles de los datos observados y las distribuciones normales en la Figura 38. Es un método gráfico para el diagnóstico de diferencias y para comprobar la no-normalidad de una distribución de una variable en una población.

El test de Barlett revela con resultados ($K\text{-squared} = 113.34$, 5 grados de libertad, $p\text{-value}=2.2E-16$) que los datos no cumplen con la condición de heteroscedasticidad, puesto que $p\text{-valores} < 0.05$ rechazan la hipótesis nula.

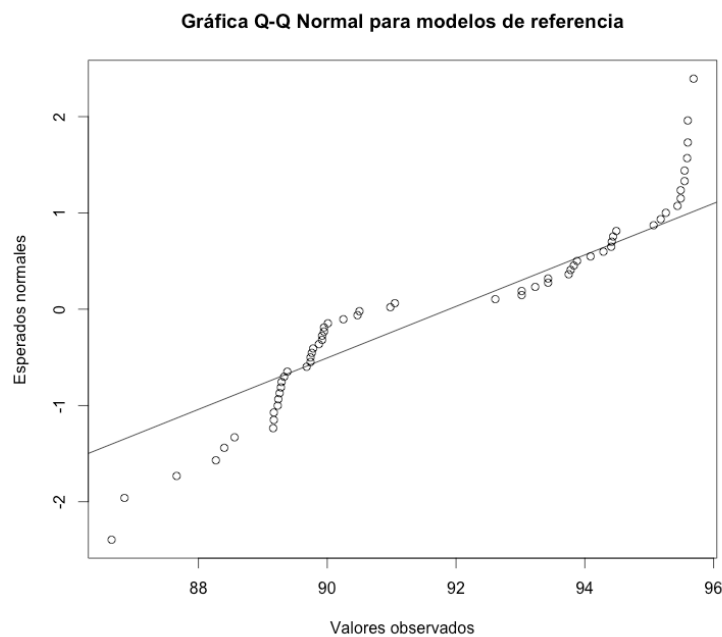


FIGURA 38. GRÁFICA QQ: MODELOS DE REFERENCIA

Dados estos resultados lo que se deberá aplicar entonces es un test no paramétrico. Particularmente un test de Friedman Alineado, dado que el número de algoritmos a comparar es seis y ofrece mejores resultados que la versión normal de Friedman [207,210,212,214]. El ranking obtenido por el test (orden creciente) es el mostrado en la Tabla 16, dando como ganador a FSMKL.

TABLA 16. TEST FRIEDMAN ALINEADO: RANKING DE LOS ALGORITMOS

Algoritmo	Ranking
FSMKL	5.8
GA-SVM	15.60
SVM-RFE	26.5
SVM	39.29
PSO-SVM	47.10
MKL	48.7

El resultado del test revela que los datos son significativamente diferentes con distribución siguiendo una distribución χ^2 : 8.39, de 5 grados de libertad y p-valor=1.350E-17. El p-valor < 0.05, por lo que se hacen los métodos de comparación múltiple de Holm, Hochberg y Hommel para comprobar si alguno de los modelos es significativamente peor que FSMKL ajustando los p-valores, los resultados se muestran en la Tabla 17, donde la columna p-valor muestra el valor obtenido por el test de Friedman y la columna Holm/Hochberg/Hommel muestra el p-valor obtenido con dichos ajustes.

TABLA 17. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL

Algoritmo	p-valor no ajustado	Ajuste Holm/Hochberg/Hommel
SVM	3.98E-8	0.01
MKL	1.27E-7	0.0125
GA-SVM	1.72E-5	0.016
PSO-SVM	0.0080	0.025
SVM-RFE	0.20	0.05

El test de Holm y el test de Hommel rechazan la hipótesis que tiene un p-valor ≤ 0.05 mientras que el test de Hochberg rechaza las hipótesis que tienen un p-valor ≤ 0.025 . Se muestran los resultados gráficamente en la Figura 39.

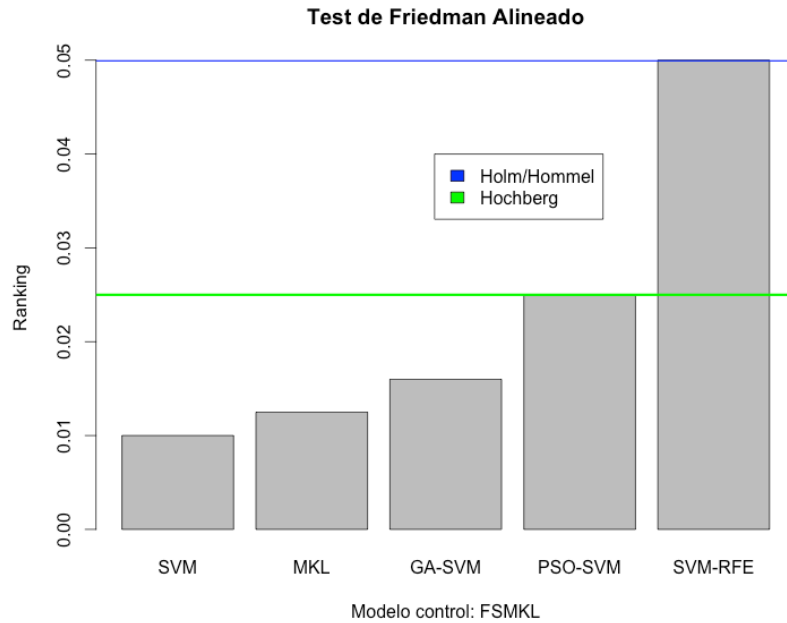


FIGURA 39. TEST DE FRIEDMAN ALINEADO: MODELOS DE REFERENCIA

Los resultados estadísticos muestran claramente que FSMKL es significativamente mejor que las otras técnicas basadas en kernel utilizadas, por lo que se va proceder a hacer un análisis de las variables de textura que están presentes en la mejor solución que encontró este modelo. Se ha de tener en cuenta que los ajustes de Holm, Hommel y Hochberg ofrecen una potencia parecida para ajustar el p-valor. El ajuste de Hochberg establece que SVM-RFE es peor estadísticamente que FSMKL, aunque no de manera significativa.

Como se explicó en el capítulo 2, a cada kernel de la solución ofrecida por FSMKL se le asigna un valor de peso que mide cuánto de importante es para la solución final. Este se construye haciendo combinaciones de kernel como combinaciones lineales de los mismos. Como se puede apreciar en la Figura 40 los dos primeros kernels (kernel 161 y kernel 439) explican el 56.56% de la variación en el FSMKL, estos kernels tienen las mismas dos variables. Con la inclusión del tercer kernel, en orden de importancia decreciente, se explica el 64,00% de la variación. A partir de este tercer kernel, la importancia de los restantes kernel es menor y participan casi de manera marginal en el resultado final.



FIGURA 40. FSMKL: IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA

Se presenta en la Figura 41 una gráfica en la que para cada una de las variables que forman parte de todos los *kernel* solución del problema, se calcula su importancia particular multiplicando el número de veces que uno de los kernel la ha seleccionado como solución del problema, por el sumatorio de la importancia de los kernels en los que ha aparecido. De tal manera que puede evaluarse la importancia e influencia de cada variable particular, discriminando variables que aunque aparezcan mucho en la solución, pueden hacerlo en *kernels* con influencia marginal en el resultado final. Se puede ver en la figura la importancia de la variable Momento Diferencial Inverso con diferentes distancias y ángulos.

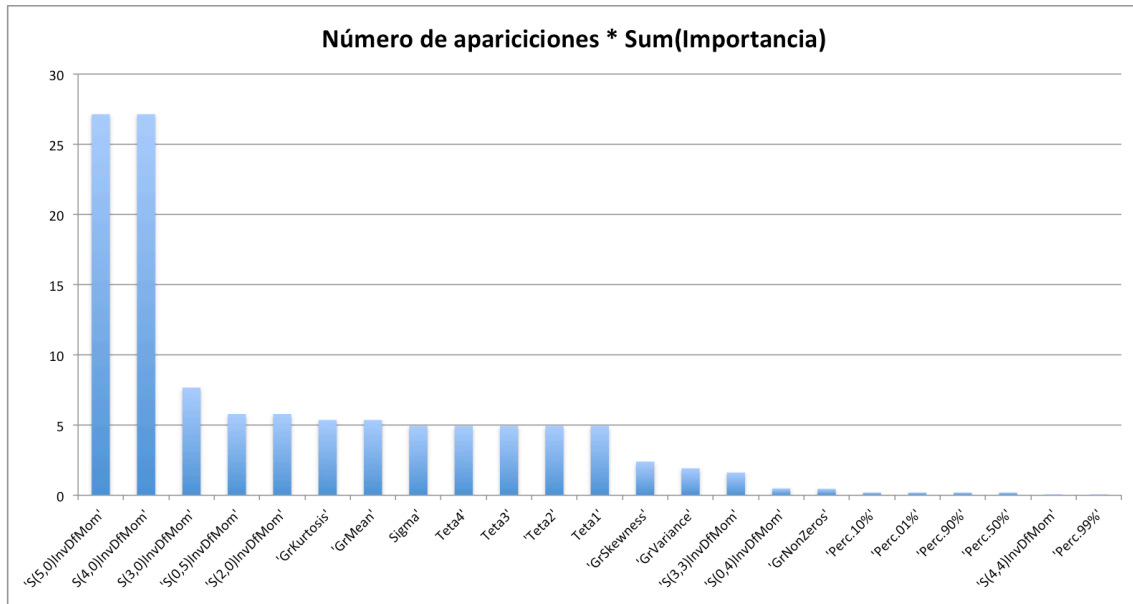


FIGURA 41. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHO KERNEL EN LA SOLUCIÓN FINAL

En la TABLA 18 se presentan agrupadas por grupo de pertenencia las variables que han sido seleccionadas por el FSMKL en los diferentes *kernels* solución. Las variables derivadas del GLCM tienen entre paréntesis información de la distancia y el ángulo en el que fueron medidas con respecto al píxel estudiado. Puede verse como de los 6 grupos de variables de textura que se han extraído originalmente y presentados en la Tabla 4 esta técnica ha descartado a dos de ellos (Matriz de Longitud de Carreras y Wavelet). El grupo que presenta el mayor número de variables es el compuesto por la Matriz de Coocurrencia, siendo además el grupo cuyas variables tienen más influencia en el resultado final. Es más, las variables que pertenecen al grupo de la matriz de coocurrencia es el Momento Diferencial Inverso, calculado con diferentes opciones de longitud y ángulo que como se puede apreciar en la Figura 41 es la variable más importante.

TABLA 18. FSMKL: VARIABLES FINALES

Grupo	Variables	Número
Histograma	Perc.10%, Perc.01%, Perc.50%, Perc.90%, Perc.99%	5
Gradiente absoluto	GrKurtosis, GrMean, GrSkewness, GrVariance, GrNonZeros	5
Matriz de Coocurrencia	S(5,0)InvDfMom, S(4,0)InvDfMom, S(3,0)InvDfMom, S(0,5)InvDfMom, S(2,0)InvDfMom, S(3,3)InvDfMom, S(0,4)InvDfMom, S(4,4)InvDfMom	8
Modelo Autoregresivo	Theta1, Theta2, Theta3, Theta4, Sigma	5

Esta Tesis Doctoral propone FSMKL como la técnica más robusta para la selección de variables de textura en imagen de electroforesis bidimensional. A continuación se realizará un análisis exhaustivo de esta técnica proponiendo mejoras y modificaciones que sirvan para incrementar de manera estadísticamente significativa los resultados en clasificación obtenidos. Para ello se propondrán nuevos kernels, nuevas distancias para estudiar la correlación entre variables como fase previa del método *filter* FSMKL y se estudiará un *kernel* Mahalanobis parsimonioso específico para problemas de alta dimensionalidad.

5.2 RESULTADOS OBTENIDOS CON LAS MEJORAS PROPUESTAS

Los resultados presentados en la apartado anterior muestran como el mejor método basado en kernels para la selección de variables en un análisis de texturas es FSMKL. Este método hace uso de los kernels RBF y polinomial, presentados en el capítulo 2, y busca combinaciones lineales entre ellos, con diferentes conjuntos de variables y configuraciones.

En las secciones siguientes se muestran los resultados obtenidos tras aplicar el modelo y mejoras propuestos en esta Tesis Doctoral.

5.2.1 APRENDIZAJE

Las fases de extracción y preprocesado de los datos se mantienen igual que en el capítulo anterior, por lo que se omite volver a repetir los pasos realizados por cuestiones de claridad. A continuación se muestran los resultados obtenidos por el sistema propuesto en esta Tesis Doctoral. De nuevo se muestran las fases de aprendizaje separadas por cada técnica propuesta, pudiendo comprobar como entrena cada modelo, el funcionamiento del proceso de reducción de dimensionalidad y la estimación del error real cometido.

Como se ha comentado con anterioridad, esta Tesis Doctoral propone el uso de dos nuevos kernels y la aplicación de diferentes medidas de distancia para estudiar la correlación de los datos con la etiqueta de salida. Por este motivo, en las secciones que siguen a continuación, se separará por este criterio para dar claridad al trabajo

presentado. También es cierto que se trata de técnicas para la reducción de dimensionalidad, por lo que se integra de nuevo en la fase general de aprendizaje.

5.2.1.1 NUEVOS KERNELS APLICADOS EN EL SISTEMA PROPUESTO

Se presentarán en este apartado las pruebas realizadas con los nuevos kernels propuestos así como los resultados obtenidos. Una vez se realicen los test estadísticos para analizar la significancia de los resultados, se hará un estudio de las variables encontradas por la mejor solución por parte del *kernel* que haya resultado ser significativamente mejor que el resto.

5.2.1.2 NUEVO KERNEL SUMA DE GAUSSIANO Y LINEAL

Para esta aproximación se va a utilizar el nuevo *kernel* que surge de la suma del Gaussiano y Lineal. El dominio de los parámetros para esta técnica se muestra en la Tabla 19.

TABLA 19. DOMINIO DE LOS PARÁMETROS UTILIZADOS CON FSMKL CON NUEVO KERNEL SUMA

Elemento	Dominio
Sigma del kernel RBF	De 0.5 a 20
Valor de C para el SVM	10,100
Criterio de parada	Duality gap
Valor de parada	0.01

Los resultados de los diez experimentos llevados a cabo se muestran en la , se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución, número de kernels y número de variables escogidas por el modelo en cada experimento. Cada experimento se ha desarrollado con un proceso de *10-fold cross validation*.

TABLA 20. FSMKL-SUMA: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	89.70	95.47	168.78	21	20
2	90.00	95.52	161.22	23	21
3	90.00	95.65	160.07	18	21
4	89.70	95.50	152.67	25	21
5	89.70	95.59	156.86	27	22
6	90.00	95.66	153.79	22	23
7	90.00	95.70	168.11	42	21
8	90.20	95.38	168.68	26	20
9	90.10	95.49	165.43	26	20
10	90.00	95.55	168.08	19	21

Los parámetros que selecciona para cada uno de estos kernels son de orden muy variable y más alto que los obtenidos inicialmente con esta técnica, con valores en el rango entre (0.5-20) para el valor de sigma en el *kernel* local.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

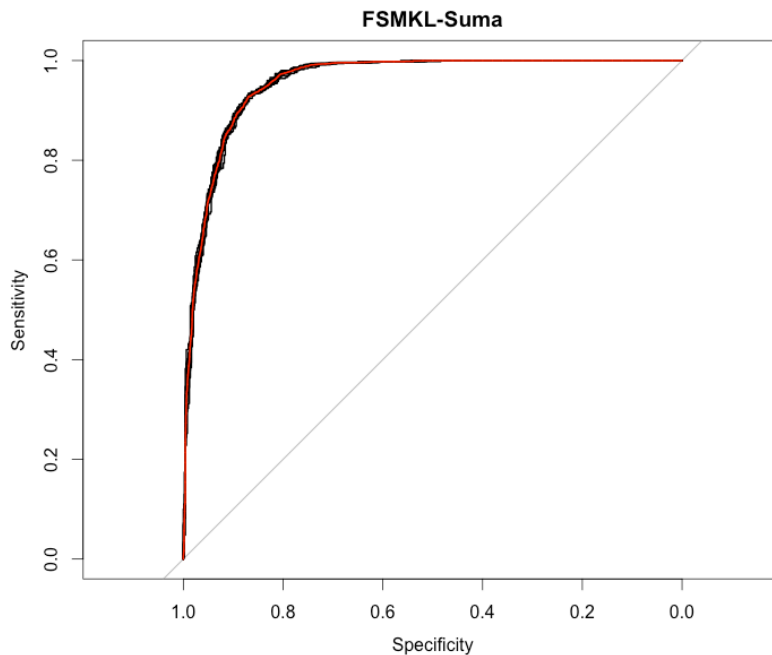


FIGURA 42. FSMKL-SUMA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

La inclusión del nuevo *kernel* en el FSMKL no ha variado su estabilidad, mejorando incluso los resultados. El número de variables encontradas como solución al problema también ha bajado ligeramente, estando ahora más cercano a las 20 variables. El número de *kernels* utilizado permanece estable.

5.2.1.3 NUEVO KERNEL PRODUCTO DE GAUSSIANO Y LINEAL

Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 21, se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución, número de *kernels* y número de variables escogidas por el modelo en cada experimento.

TABLA 21. FSMKL-PRODUCTO: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	90.80	95.65	96.37	86	21
2	90.50	95.67	83.75	64	17
3	90.50	95.76	93.81	54	22
4	90.80	95.65	89.93	73	22
5	90.70	95.71	89.59	66	23
6	90.60	95.64	82.02	101	33
7	90.60	95.83	91.46	68	22
8	90.40	95.57	92.43	67	21
9	90.60	95.68	91.03	79	27
10	90.50	95.68	86.91	60	22

Los parámetros que selecciona para cada uno de estos kernels son de orden muy variable y más elevado que los encontrados inicialmente con esta técnica, en el rango entre (0.5-20) para el valor de sigma en el kernel local.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

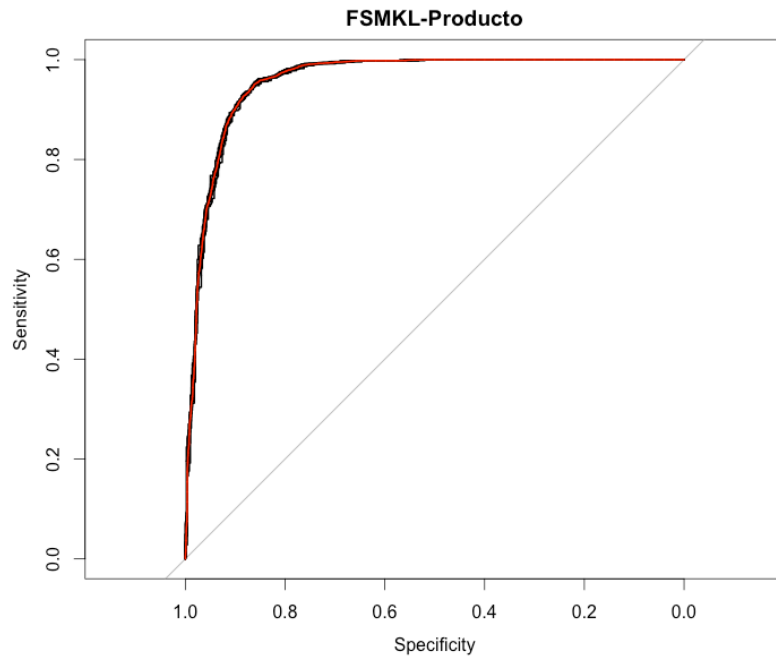


FIGURA 43. FSMKL-PRODUCTO: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

Al igual que en el caso anterior el número de variables se mantiene estable mientras que, por otro lado, se ve incrementado el número de *kernels* necesarios para solucionar el problema. El tiempo de procesamiento también se ha visto reducido con este nuevo *kernel* respecto a las pruebas presentadas anteriormente.

5.2.2 KERNELS PROPUESTOS: SELECCIÓN DEL MEJOR MODELO

De nuevo y como se ha explicado en el capítulo 4.1.4, es necesario comprobar si se cumplen las condiciones necesarias para aplicar un test paramétrico y poder comparar las distintas ejecuciones de cada uno de los algoritmos y determinar si existe significancia estadística entre ellos.

Se muestran las ROC calculadas en los apartados anteriores en la Figura 44. Se utilizarán simplemente para dar una aproximación visual a los resultados obtenidos con los diferentes clasificadores, pero la significancia y la selección del mejor modelo se hará en base a la aplicación de los distintos test estadísticos que sea necesario aplicar.

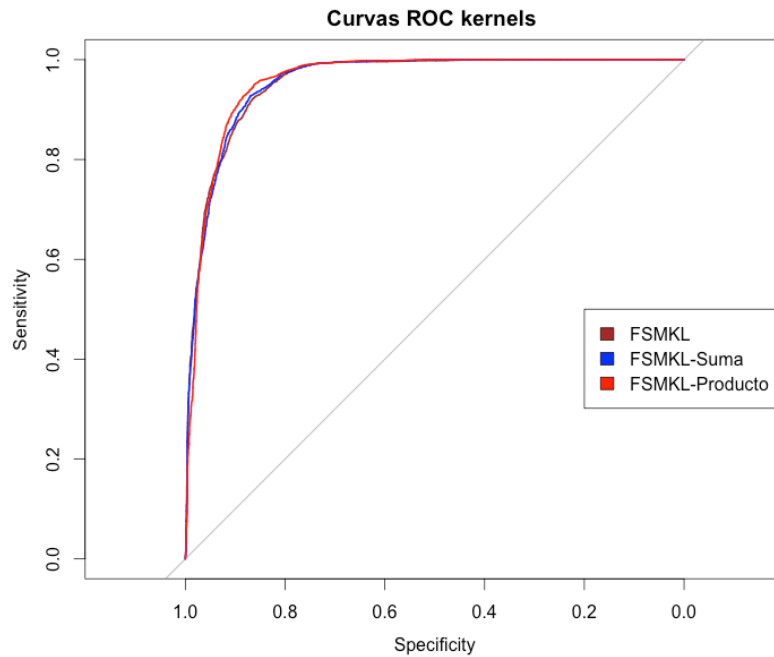


FIGURA 44. NUEVOS KERNELS: CURVAS ROC

Recordando lo dicho previamente, es necesario comprobar que cumplan las condiciones de independencia, normalidad y heteroscedasticidad. En este caso, la condición de independencia es obvia ya que se va a comparar diferentes ejecuciones de cada uno de los algoritmos con semillas de inicio en el reparto generadas aleatoriamente. Para comprobar la normalidad se aplicará el test de Shapiro-Wilk y para comprobar la heteroscedasticidad el test de Bartlett. Una vez se determine si han cumplido todas estas condiciones se estará en disposición de aplicar un test paramétrico ANOVA o un test no paramétrico Friedman o variantes.

Se realizarán a continuación los diferentes estudios estadísticos necesarios para saber si alguno de los modelos propuestos es estadísticamente mejor entre los dos nuevos kernels para el análisis de texturas que surgen de combinar la suma de un kernel RBF y uno lineal y por otro lado, de multiplicar elemento a elemento un kernel RBF y uno lineal, además del mejor modelo presentado en el apartado 5.1.4. Los resultados a lo largo de 10 experimentos con 10-cv de los tres modelos a comparar son los mostrados en la Tabla 22 y en la Figura 45.

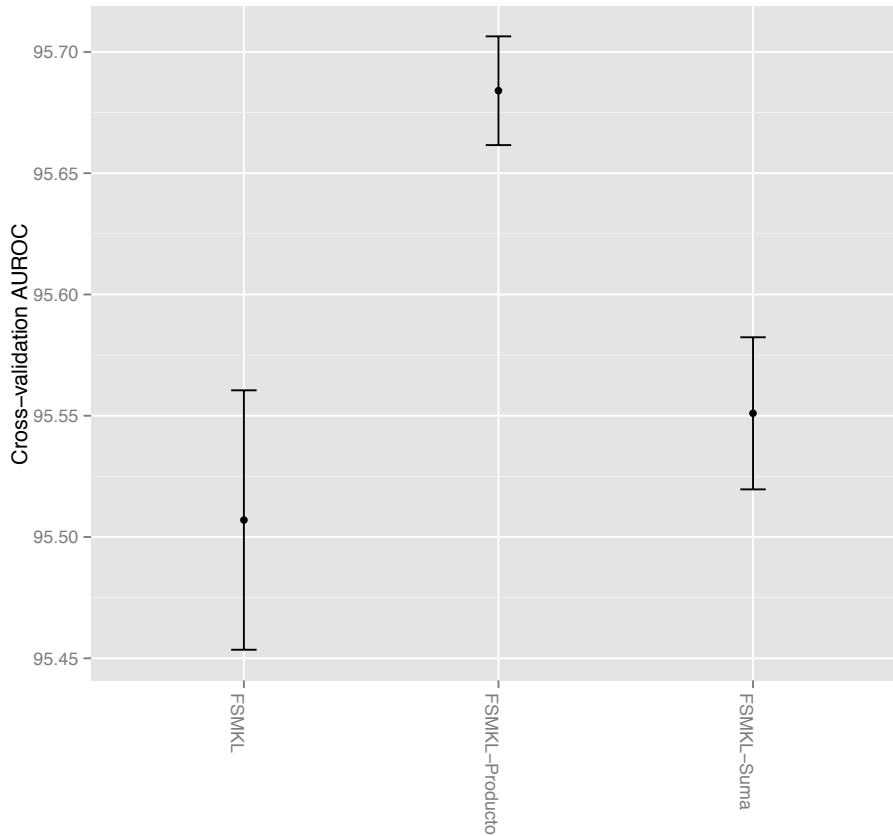


FIGURA 45. NUEVOS KERNELS PROPUESTOS: RESULTADOS

TABLA 22. NUEVOS KERNELS PROPUESTOS: RESULTADOS

Exp.	FSMKL		FSMKL-Suma		FSMKL-Producto	
	Acierto Validación	AUROC Validación	Acierto Validación	AUROC Validación	Acierto Validación	AUROC Validación
1	89.70	95.55	89.70	95.47	90.80	95.65
2	89.70	95.49	90.00	95.52	90.50	95.67
3	89.70	95.60	90.00	95.65	90.50	95.76
4	89.50	95.44	89.70	95.50	90.80	95.65
5	89.30	95.69	89.70	95.59	90.70	95.71
6	89.30	95.49	90.00	95.66	90.60	95.64
7	89.40	95.60	90.00	95.70	90.60	95.83
8	89.40	95.07	90.20	95.38	90.40	95.57
9	89.60	95.55	90.10	95.49	90.60	95.68
10	89.30	95.59	90.00	95.55	90.50	95.68

El test de Shapiro Wilk revela con resultados ($W=0.8835$ y p -valor 0.003395) que los datos no siguen una distribución normal, ya que p -valores < 0.05 rechazan la hipótesis nula. Se presenta el gráfica Q-Q de la confrontación entre los cuartiles de los datos observados y las distribuciones normales en la Figura 46.

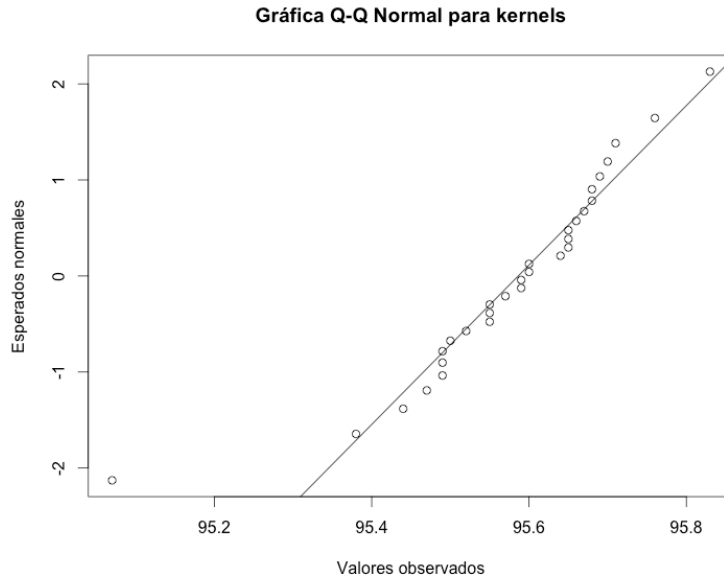


FIGURA 46. GRÁFICA QQ: NUEVOS KERNELS

El test de Barlett revela con resultados ($K\text{-squared} = 6.568$, 2 grados de libertad, $p\text{-valor}=0.03748$) que los datos no cumplen con la condición de heteroscedasticidad ya que $p\text{-valores} < 0.05$ rechazan la hipótesis nula.

Dado que el número de algoritmos a comparar es tres, se puede aplicar un test no paramétrico de Quade que obtiene mejores resultados [207,210,212,214]. El ranking ordenado obtenido con el test es el mostrado en la Tabla 23, dando por ganador a FSMKL-Producto.

TABLA 23. TEST NO PARAMÉTRICO QUADE: RANKING DE LOS ALGORITMOS

Algoritmo	Ranking
FSMKL-Producto	1.07
FSMKL-Suma	2.20
FSMKL	2.71

El resultado del test revela (valor estadístico Quade: 11.34, con distribución siguiendo una distribución F, de 2 grados de libertad y $p\text{-valor}=6.49E-4$) que los datos son significativamente diferentes puesto que se ha obtenido un $p\text{-valor} < 0.05$.

Se realizan los test de Holm, Hochberg y Hommel de comparación múltiple para ajustar los p-valores. Se comprobará si alguno de los algoritmos, FSMKL o FSMKL-Suma es significativamente peor que FSMKL-Producto, los resultados se muestran en Tabla 24.

TABLA 24. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL

Algoritmo	p-valor	Holm/Hochberg/Hommel
FSMKL	0.0011	0.025
FSMKL-Suma	0.024	0.05

El test de Hochberg rechaza aquellas hipótesis que tengan un p-valor < 0.05 y Hommel rechaza todas las hipótesis nulas. Se muestran los resultados gráficamente en la Figura 47.

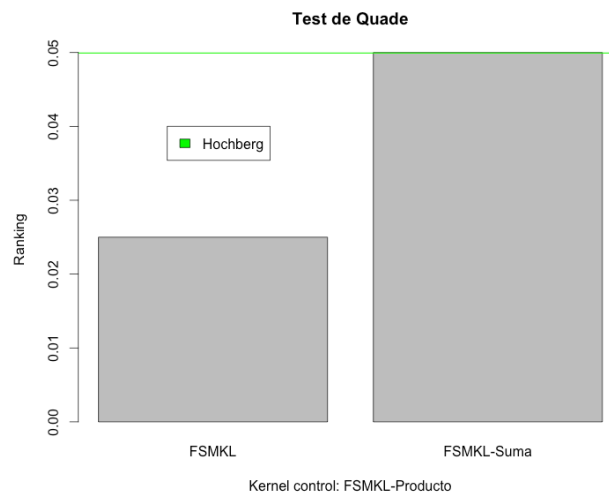


FIGURA 47. TEST NO PARAMÉTRICO QUADE: NUEVOS KERNELS

Se puede concluir que el nuevo *kernel*, FSMKL-Producto propuesto en esta Tesis para el análisis de texturas es significativamente mejor que los otros.

Como se explicó en el capítulo 2 a cada kernel de la solución, se le asigna un valor de peso que mide cuánto de importante es para la solución final. Este se construye con el nuevo *kernel* propuesto. Como se puede apreciar en la Figura 48, los once primeros kernels explican el 61,90% de la variación en el FSMKL, estos kernels

tienen once variables que se corresponden con distintos ángulos y distancias de análisis de la variable derivada de la GLCM, *Momento Diferencial Inverso*.

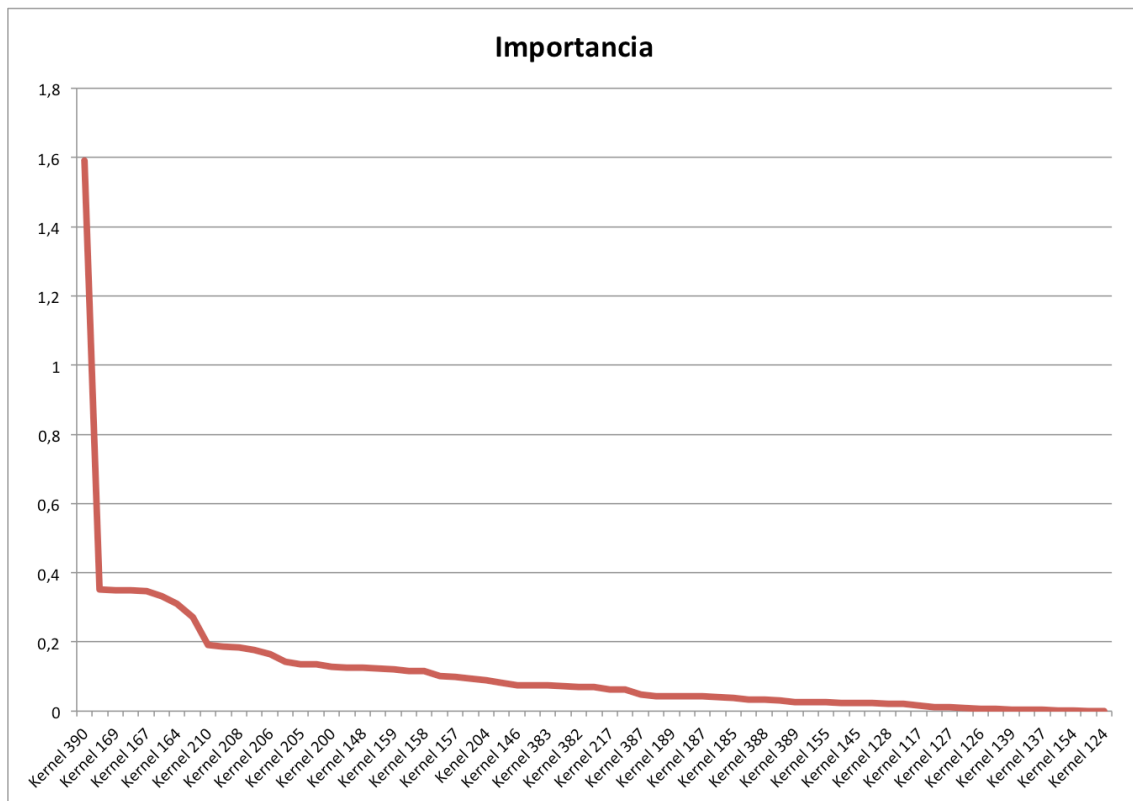


FIGURA 48. IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA

Se presenta en la Figura 49 una gráfica en la que para cada una de las variables que forman parte de todos los *kernel* solución del problema, se calcula su importancia particular multiplicando el número de veces que uno de los *kernel* la ha seleccionado como solución del problema, por el sumatorio de la importancia de los *kernels* en los que ha aparecido. De tal manera que puede evaluarse la importancia e influencia de cada variable particular, discriminando variables que aunque aparezcan mucho en la solución, pueden hacerlo en *kernels* con influencia marginal en el resultado final.

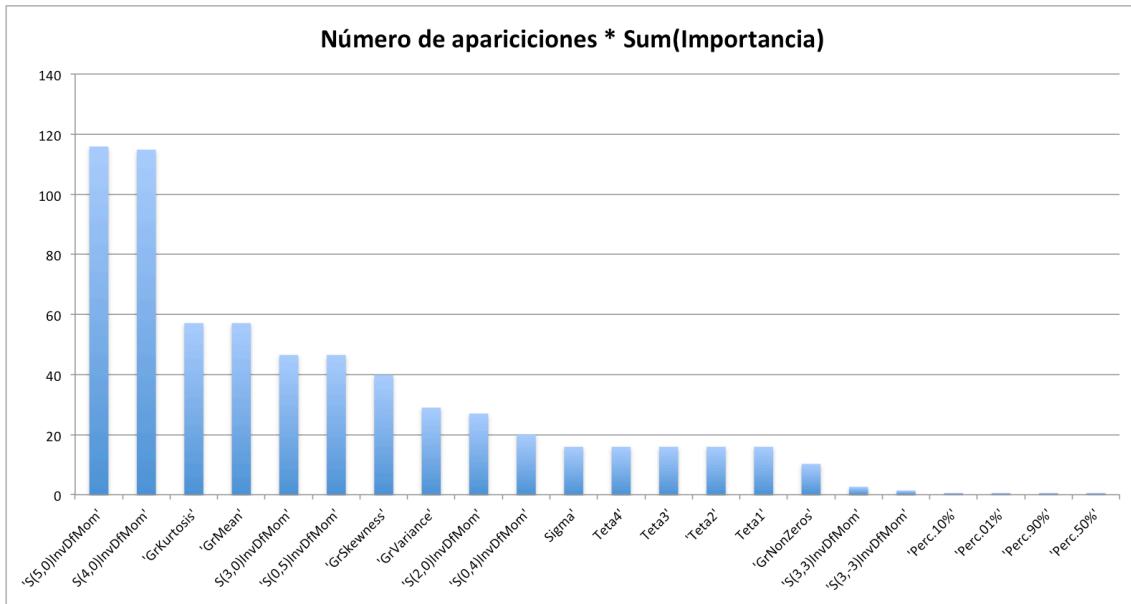


FIGURA 49. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHO KERNEL EN LA SOLUCIÓN FINAL

En la TABLA 25 se presentan agrupadas por grupo de pertenencia las variables que han sido seleccionadas por el FSMKL en los diferentes kernels solución. Las variables derivadas del GLCM tienen entre paréntesis información de la distancia y el ángulo en el que fueron medidas con respecto al píxel estudiado. Puede verse como de los 6 grupos de variables de textura que se han extraído originalmente y presentados en la Tabla 4, esta técnica ha descartado a dos de ellos (Matriz de Longitud de Carreras y Wavelet). El grupo que presenta el mayor número de variables es el compuesto por la Matriz de Coocurrencia, siendo además el grupo cuyas variables tienen más influencia en el resultado final. Es más, la variable que pertenecen al grupo de la matriz de coocurrencia es el Momento Diferencial Inverso, calculado con diferentes opciones de longitud y ángulo que como se puede apreciar en la Figura 49 la variable más importante, aparecen también las variables de Gradiente Absoluto con un valor alto de peso en la solución.

TABLA 25. FSMKL-PRODUCTO: VARIABLES FINALES SELECCIONADAS

Grupo	Variables	Número
Histograma	Perc.10%, Perc.01%, Perc.50%, Perc.90%	4
Gradiente absoluto	GrKurtosis, GrMean, GrSkewness, GrVariance, GrNonZeros	5
Matriz de Coocurrencia	S(5,0)InvDfMom, S(4,0)InvDfMom, S(3,0)InvDfMom, S(0,5)InvDfMom, S(2,0)InvDfMom, S(3,3)InvDfMom, S(0,4)InvDfMom, S(3,-3)InvDfMom	8
Modelo Autoregresivo	Theta1, Theta2, Theta3, Theta4, Sigma	5

5.2.3 APRENDIZAJE

Se propondrán en el presente apartado nuevas distancias para estudiar la correlación entre variables como fase previa del método *filter* FSMKL.

5.2.3.1 NUEVAS MÉTRICAS DE SELECCIÓN DE VARIABLES

A continuación se presentará un estudio con diferentes métricas de distancia en la correlación entre las variables y la clase de pertenencia, en concreto: Fisher, Bhattacharyya, Wilcoxon, ROC, Entropía relativa y Entropía relativa transformada con el objetivo de mejorar los resultados obtenidos.

Los resultados de los diez experimentos de cada una de las medidas llevados a cabo se muestran en las Tablas 26-32 donde se presentan los valores de porcentaje de acierto, AUROC, tiempo de ejecución, número de *kernels* y número de variables escogidas por el modelo en cada experimento. Cada experimento es un proceso *10-fold cross validation*.

5.2.3.1.1 DISTANCIA DE FISHER

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 26. DISTANCIA FISHER: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	89.60	95.51	107.47	16	31
2	89.70	95.49	123.24	18	22
3	89.70	95.61	116.99	20	35
4	89.50	95.45	102.22	22	22
5	89.30	95.69	121.72	20	23
6	89.30	95.55	101.66	19	24
7	89.40	95.60	110.09	22	22
8	89.40	95.08	108.37	39	35
9	89.60	95.55	97.54	15	25
10	89.30	95.59	102.38	18	22

Los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5, 1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

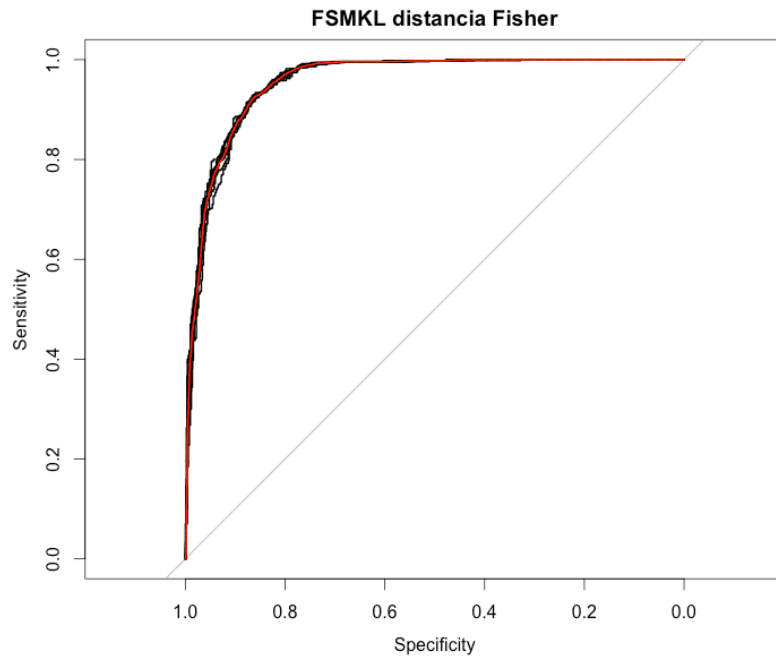


FIGURA 50. DISTANCIA FISHER: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.2.3.1.2 DISTANCIA DE BHATTACHARYYA

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 27. DISTANCIA BHATTACHARYYA: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	89.80	95.75	68.53	24	34
2	89.80	95.94	56.07	24	34
3	90.30	95.75	59.13	12	34
4	89.70	95.75	66.42	33	37
5	89.70	95.79	60.77	17	35
6	89.40	95.76	62.60	27	35
7	90.20	95.97	58.02	17	35
8	90.00	95.64	65.49	24	34
9	89.50	95.82	71.26	14	38
10	89.90	95.55	55.15	26	35

Los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5, 1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global. Se representan a continuación en la Figura 51 las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

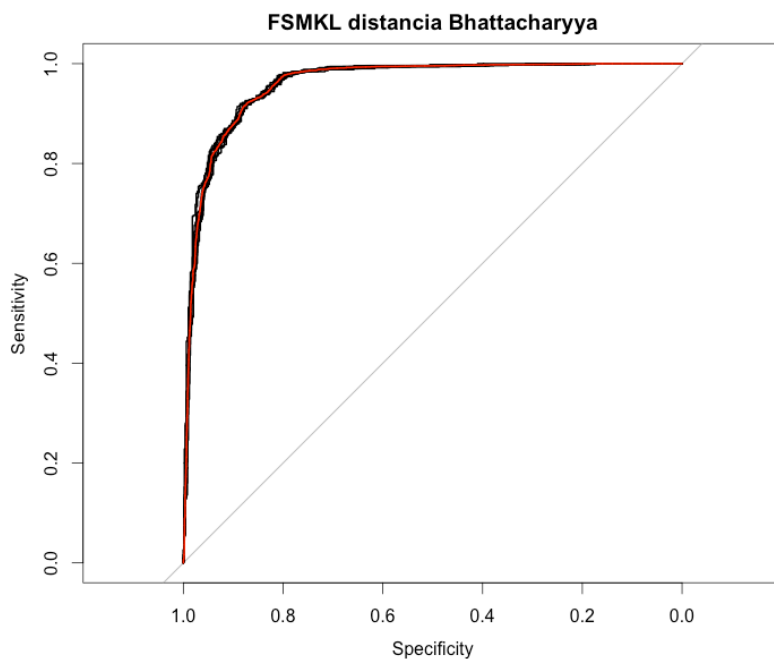


FIGURA 51. DISTANCIA BHATTACHARYYA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

Los tiempos de ejecución con esta distancia se han reducido con respecto a otras aproximaciones.

5.2.3.1.3 DISTANCIA DE WILCOXON

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 28. DISTANCIA WILCOXON: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	89.60	95.56	77.97	19	32
2	89.30	95.44	82.30	31	32
3	89.60	95.57	77.08	19	32
4	89.80	95.60	82.60	18	32
5	89.50	95.60	74.62	17	32
6	89.60	95.45	91.94	40	32
7	89.50	95.62	82.81	33	32
8	89.50	95.10	93.63	30	32
9	89.60	95.41	70.74	29	32
10	89.80	95.49	77.52	27	32

Los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5, 1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global.

Se representan a continuación en la Figura 52 las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

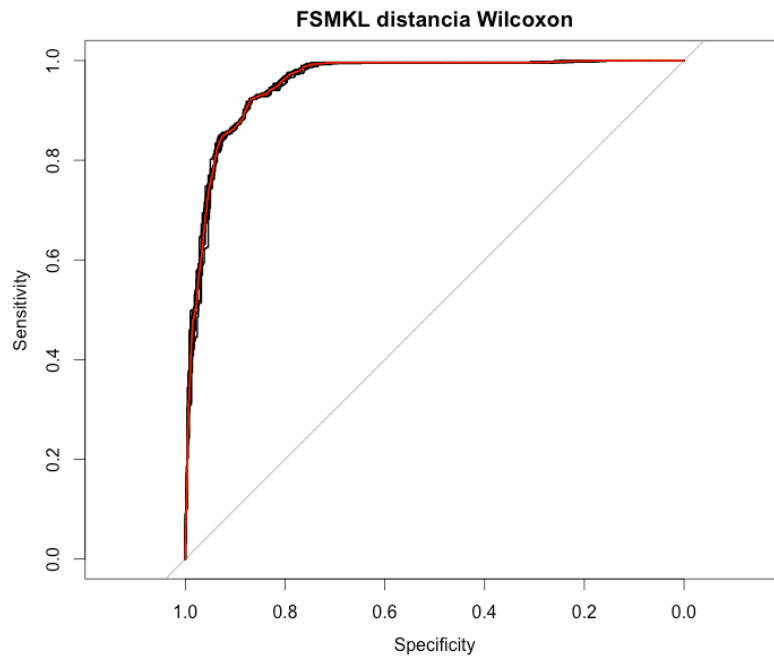


FIGURA 52. DISTANCIA WILCOXON: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

Con esta distancia de Wilcoxon en los diez experimentos inicializados con distintas semillas aleatorias se llega siempre a las mismas 32 variables. Estas variables son las mostradas en la Tabla 29.

TABLA 29. DISTANCIA WILCOXON: VARIABLES SELECCIONADAS

Grupo	VARIABLES	Número
Histograma	Skewness, Variance	2
Gradiente absoluto	GrMean, GrVariance, GrSkewness, GrKurtosis, GrNonZeros	5
Matriz de Coocurrencia	S(5,0)InvDfMom, S(3,-3)InvDfMom, S(1,0)AngScMom, S(1,1)AngScMom, S(1,-1)AngScMom, S(2,0)AngScMom, S(2,2)AngScMom, S(2,-2)AngScMom, S(3,0)AngScMom, S(0,3)AngScMom, S(3,3)AngScMom, S(3,-3)AngScMom, S(4,0)AngScMom, S(0,4)AngScMom, S(4,4)AngScMom, S(4,-4)AngScMom, S(5,0)AngScMom, S(0,5)AngScMom, S(5,5)AngScMom, S(5,-5)AngScMom	20
Modelo Autoregresivo	Theta1, Theta2, Theta3, Theta4, Sigma	5

5.2.3.1.4 DISTANCIA DE ROC

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 30. DISTANCIA ROC: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	90.00	95.51	81.81	20	42
2	90.10	95.62	66.06	25	40
3	90.30	95.59	76.99	41	40
4	90.10	95.75	78.23	24	38
5	90.30	95.64	78.92	63	38
6	90.20	95.90	87.48	32	39
7	90.20	95.55	75.16	51	40
8	90.40	95.59	68.14	17	38
9	90.30	95.61	82.64	14	39
10	90.70	95.63	73.48	48	40

De nuevo con este *kernel*, los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5-1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

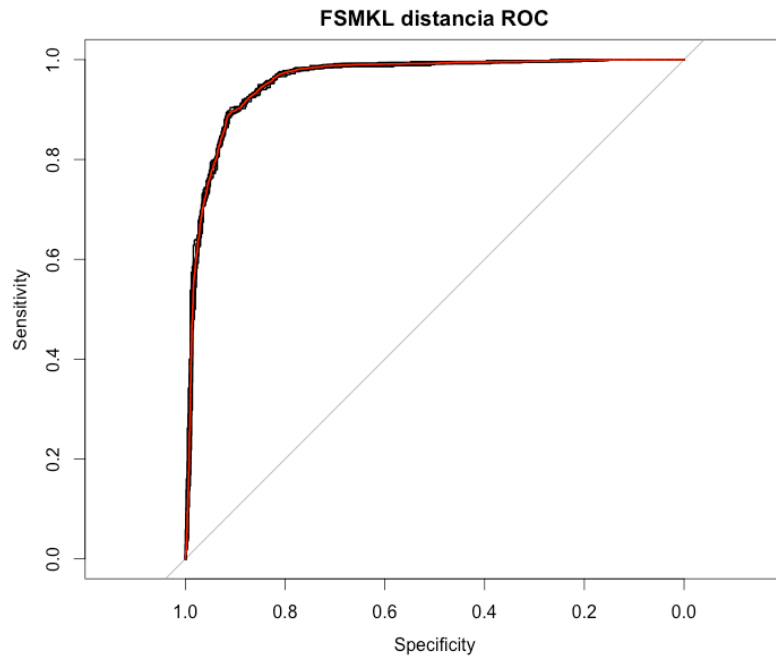


FIGURA 53. DISTANCIA ROC: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.2.3.1.5 DISTANCIA ENTROPÍA-RELATIVA

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 31. DISTANCIA ENTROPÍA RELATIVA: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	90.20	95.73	81.69	15	33
2	90.10	95.71	79.64	16	32
3	89.90	95.73	74.13	19	34
4	90.10	95.74	69.81	18	34
5	90.30	95.73	77.45	12	36
6	90.10	95.67	85.04	16	36
7	90.20	95.77	75.05	22	36
8	90.10	95.73	82.31	23	35
9	90.30	95.80	85.85	20	37
10	90.10	95.74	74.77	16	34

Los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5-1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

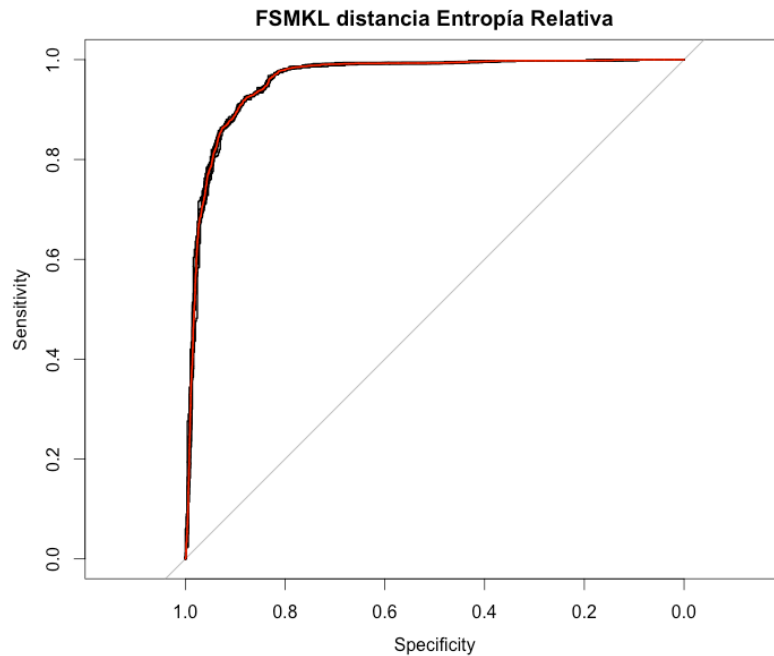


FIGURA 54. DISTANCIA ENTROPÍA RELATIVA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.2.3.1.6 DISTANCIA ENTROPÍA-RELATIVA TRANSFORMADA

Se presentan los resultados obtenidos con esta medida para el estudio de correlación entre las variables y la clase de pertenencia.

TABLA 32. DISTANCIA ENTROPÍA RELATIVA TRANSFORMADA: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)	Número de Kernels	Núm. de variables
1	90.10	95.70	78.10	15	33
2	90.10	95.57	84.46	9	32
3	89.90	95.78	77.27	23	33
4	90.10	95.72	75.56	18	34
5	90.30	95.71	72.13	15	37
6	90.10	95.60	82.35	13	36
7	90.20	95.76	77.02	20	36
8	90.10	95.78	79.83	23	35
9	90.30	95.79	90.43	18	37
10	90.10	95.70	74.59	19	34

Los parámetros que selecciona para cada uno de estos kernels son de orden bajo y estable, en el rango entre (0.5-1) para el valor de sigma en el kernel local y de (1) para el grado del kernel global.

Se representan a continuación las curvas ROC de cada uno de los diez experimentos y en color rojo, la curva ROC que se corresponde con los diez experimentos al mismo tiempo.

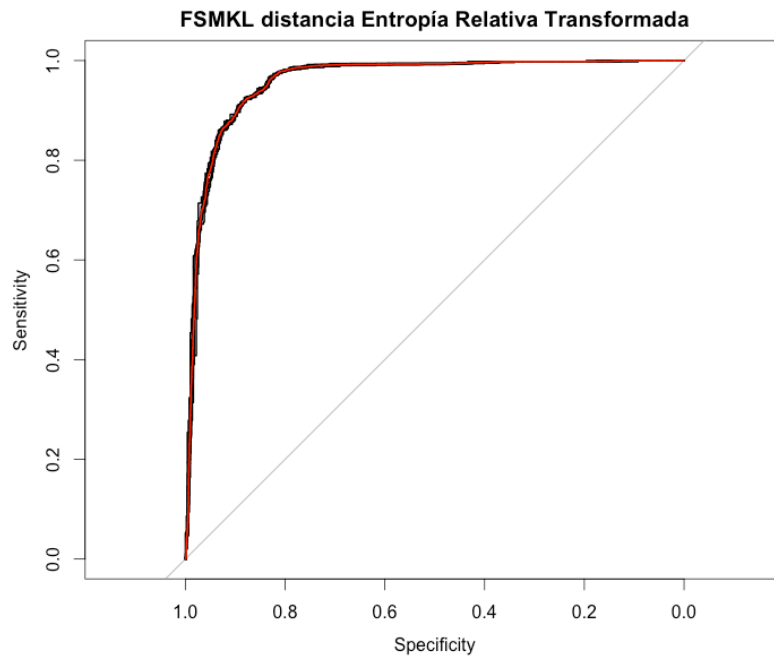


FIGURA 55. ENTROPÍA RELATIVA TRANSFORMADA: CURVAS ROC CORRESPONDIENTES A LOS 10 EXPERIMENTOS

5.2.4 NUEVAS MÉTRICAS: SELECCIÓN DEL MEJOR MODELO

De nuevo y como se ha explicado en el capítulo 4.1.4, es necesario comprobar si se cumplen las condiciones necesarias para aplicar un test paramétrico y poder comparar las distintas ejecuciones de cada uno de los algoritmos y determinar si existe significancia estadística entre ellos.

Recordando lo dicho, es necesario comprobar que cumplan las condiciones de independencia, normalidad y heteroscedasticidad. En este caso, la condición de independencia es obvia ya que se va a comparar diferentes ejecuciones de cada uno de los algoritmos con semillas de inicio en el reparto generadas aleatoriamente. Para comprobar la normalidad se aplicará el test de Shapiro-Wilk y para comprobar la

heteroscedasticidad el test de Bartlett. Una vez se determine si han cumplido todas estas condiciones se está en disposición de aplicar un test paramétrico ANOVA o un test no paramétrico como Friedman.

Se muestran las ROC calculadas en los apartados anteriores en la Figura 56. Se utilizarán simplemente para dar una aproximación visual a los resultados obtenidos con los diferentes clasificadores, pero la significancia y la selección del mejor modelo se hará en base a la aplicación de los distintos test estadísticos que sea necesario aplicar.

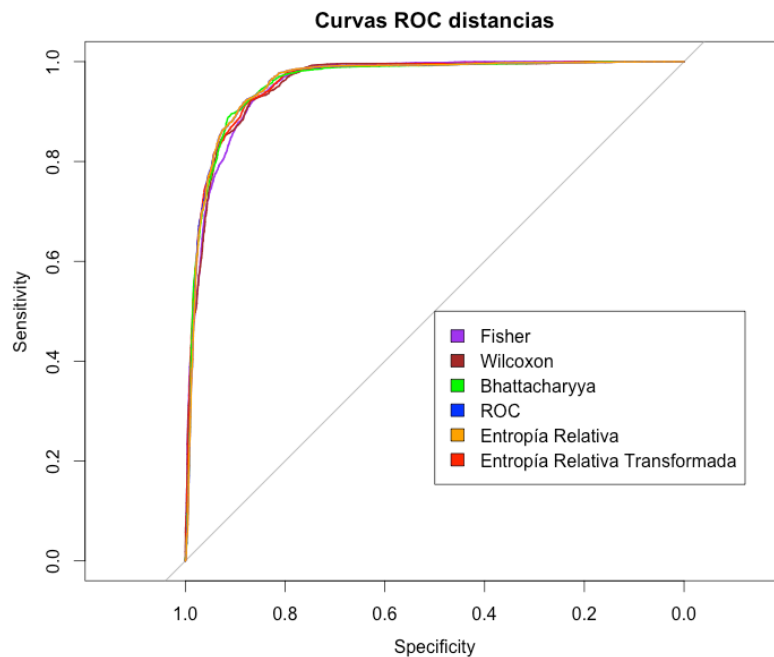


FIGURA 56. CURVAS ROC DE LOS MODELOS CON LAS DISTINTAS DISTANCIAS PROPUESTAS

En el apartado 5.1.4 se comprobó que el FSMKL es estadísticamente mejor que las otras técnicas basadas en kernel que se han utilizado en esta Tesis Doctoral. Se propone el estudio de seis nuevas medidas de correlación entre variables: Fisher, Bhattacharyya, Wilcoxon, ROC, Entropía Relativa y Entropía relativa transformada a través de test estadísticos para buscar la significancia de los resultados. Los resultados agrupados a lo largo de 10 experimentos de los siete modelos a comparar son los mostrados en la Tabla 33 y en la Figura 57.

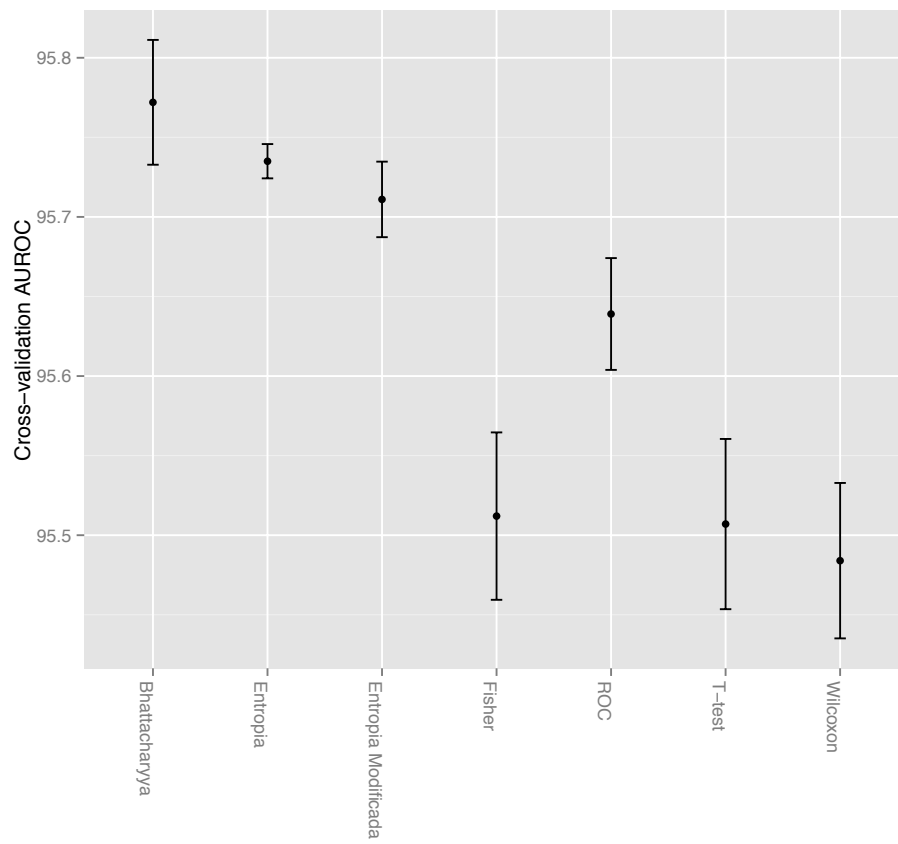


FIGURA 57. NUEVAS MEDIDAS DE DISTANCIA: RESULTADOS

TABLA 33. NUEVAS DISTANCIAS PROPUESTAS: RESULTADOS

Exp	t-test		Fisher		Bhattacharyya		Wilcoxon		ROC		Entropía Relativa		Entropía Relativa Transformada	
	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC	Acierto	AUROC
1	89.70	95.55	89.60	95.51	89.80	95.75	89.60	95.56	90.00	95.51	90.20	95.73	90.10	95.70
2	89.70	95.49	89.70	95.49	89.80	95.94	89.30	95.44	90.10	95.62	90.10	95.71	90.10	95.57
3	89.70	95.60	89.70	95.61	90.30	95.75	89.60	95.57	90.30	95.59	89.90	95.73	89.90	95.78
4	89.50	95.44	89.50	95.45	89.70	95.75	89.80	95.60	90.10	95.75	90.10	95.74	90.10	95.72
5	89.30	95.69	89.30	95.69	89.70	95.79	89.50	95.60	90.30	95.64	90.30	95.73	90.30	95.71
6	89.30	95.49	89.30	95.55	89.40	95.76	89.60	95.45	90.20	95.90	90.10	95.67	90.10	95.60
7	89.40	95.60	89.40	95.60	90.20	95.97	89.50	95.62	90.20	95.55	90.20	95.77	90.20	95.76
8	89.40	95.07	89.40	95.08	90.00	95.64	89.50	95.10	90.40	95.59	90.10	95.73	90.10	95.78
9	89.60	95.55	89.60	95.55	89.50	95.82	89.60	95.41	90.30	95.61	90.30	95.80	90.30	95.79
10	89.30	95.59	89.30	95.59	89.90	95.55	89.80	95.49	90.70	95.63	90.10	95.74	90.10	95.70

El Test de Shaphiro Wilk revela con resultados ($W=0.903$ y $p\text{-valor } 5.11E-05$) que los datos no siguen una distribución normal, ya que $p\text{-valores} < 0.05$ rechazan la hipótesis nula. Se presenta la gráfica Q-Q de la confrontación entre los cuartiles de los datos observados y las distribuciones normales en la Figura 37 es un método gráfico para el diagnóstico de diferencias y comprobar la no-normalidad de una distribución de una variable en una población.

El test de Barlett revela con resultados ($K\text{-squared} = 22.4604$, 6 grados de libertad, $p\text{-value}=0.0009989$) que los datos no cumplen con la condición de heteroscedasticidad, ya que $p\text{-valores} < 0.05$ rechazan la hipótesis nula.

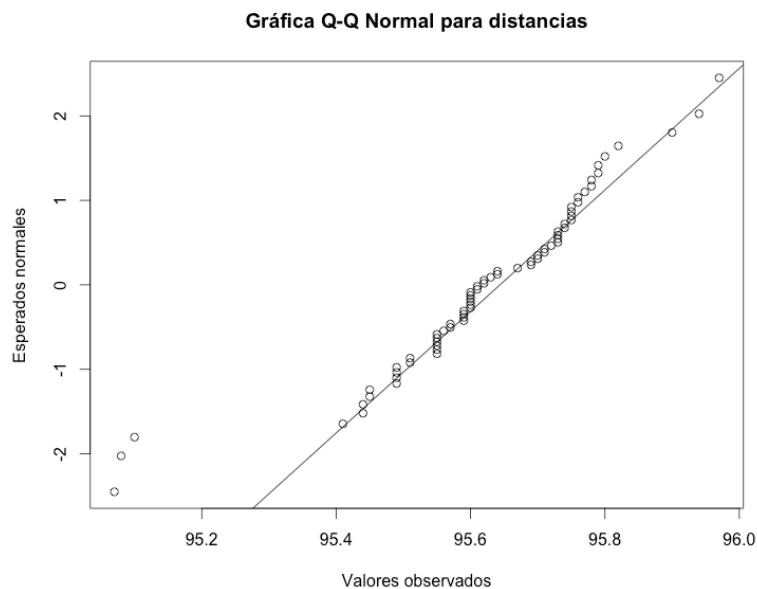


FIGURA 58. GRÁFICO QQ: NUEVAS DISTANCIAS PROPUESTAS

Dado que el número de algoritmos a comparar es seis, se puede aplicar un test no paramétrico de Friedman Alineado dado que tiene mejores resultados que la versión normal de Friedman [207,210,212,214]. El ranking ordenado obtenido con el test es el mostrado en la Tabla 34. La distancia ganadora es Bhattacharyya.

TABLA 34. TEST NO PARAMÉTRICO DE FRIEDMAN ALINEADO: RANKING DE LOS ALGORITMOS

Algoritmo	Ranking
Bhattacharyya	14.39
Entropía	18.59
Entropía Transformada	22.84
ROC	33.8
Fisher	51.5
Ttest	52.15
Wilcoxon	55.19

El resultado del test revela con p-valor < 0.05 que los datos son significativamente diferentes, con distribución siguiendo una distribución χ^2 : 8.715, de 6 grados de libertad y p-valor=1.90E-7. Se hace el test de comparación múltiple de Holm, Hochberg y Hommel pero comprobar si alguna de las distancias es significativamente peor que la distancia de Bhattacharyya ajustando los p-valores. Los resultados se muestran en la Tabla 35, donde la columna p-valor muestra el valor obtenido por el test de Friedman y la columna Holm/Hochberg/Hommel muestra el p-valor obtenido con dichos ajustes.

TABLA 35. RESULTADO DE LOS AJUSTES DE HOLM, HOCHBERG Y HOMMEL

Algoritmo	p-valor no ajustado	Ajuste
		Holm/Hochberg/Hommel
Wilcoxon	7.36E-6	0.0083
T-test	3.35E-5	0.001
Fisher	4.57E-5	0.0125
ROC	0.033	0.016
Entropía	0.35	0.025
Entropía transformada	0.64	0.05

El test de Holm y el test de Hommel rechazan la hipótesis que tiene un p-valor ≤ 0.016 mientras que el test de Hochberg rechaza las hipótesis que tienen un p-valor ≤ 0.0125 . Se muestran los resultados gráficamente en la Figura 59

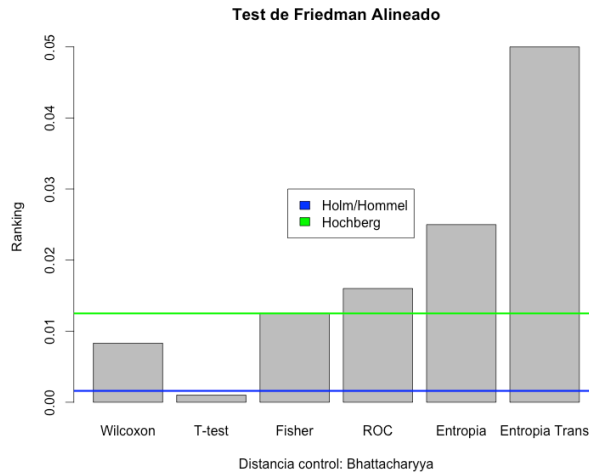


FIGURA 59. TEST NO PARAMÉTRICO DE FRIEDMAN ALINEADO: NUEVAS DISTANCIAS PROPUESTAS

La distancia de Bhattacharyya es significativamente mejor que las distancias de T-test, Wilcoxon y Fisher. La distancia de Bhattacharyya es mejor que la distancia ROC, Entropía Relativa y Entropía Relativa Transformada aunque no de manera estadísticamente significativa.

Como se explicó en el capítulo 2 a cada *kernel* de la solución, se le asigna un valor de peso que mide cuánto de importante es para la solución final. Este se construye con el nuevo *kernel* propuesto. Como se puede apreciar en la Figura 60, los dos primeros kernels explican el 58,12% de la variación en el FSMKL y tienen las mismas 19 variables.



FIGURA 60. IMPORTANCIA DE CADA KERNEL EN LA CONSTRUCCIÓN DE LA SOLUCIÓN DEL PROBLEMA

Se presenta en la Figura 61 una gráfica en la que para cada una de las variables que forman parte de todos los *kernel* solución del problema, se calcula su importancia particular multiplicando el número de veces que uno de los kernel la ha seleccionado como solución del problema, por el sumatorio de la importancia de los kernels en los que ha aparecido. De tal manera que puede evaluarse la importancia e influencia de cada variable particular, discriminando variables que aunque aparezcan mucho en la solución, pueden hacerlo en *kernels* con influencia marginal en el resultado final. Se puede observar en la Figura 61 que las primeras 16 variables siempre aparecen juntas, en un total de 9 de los 17 *kernels*.

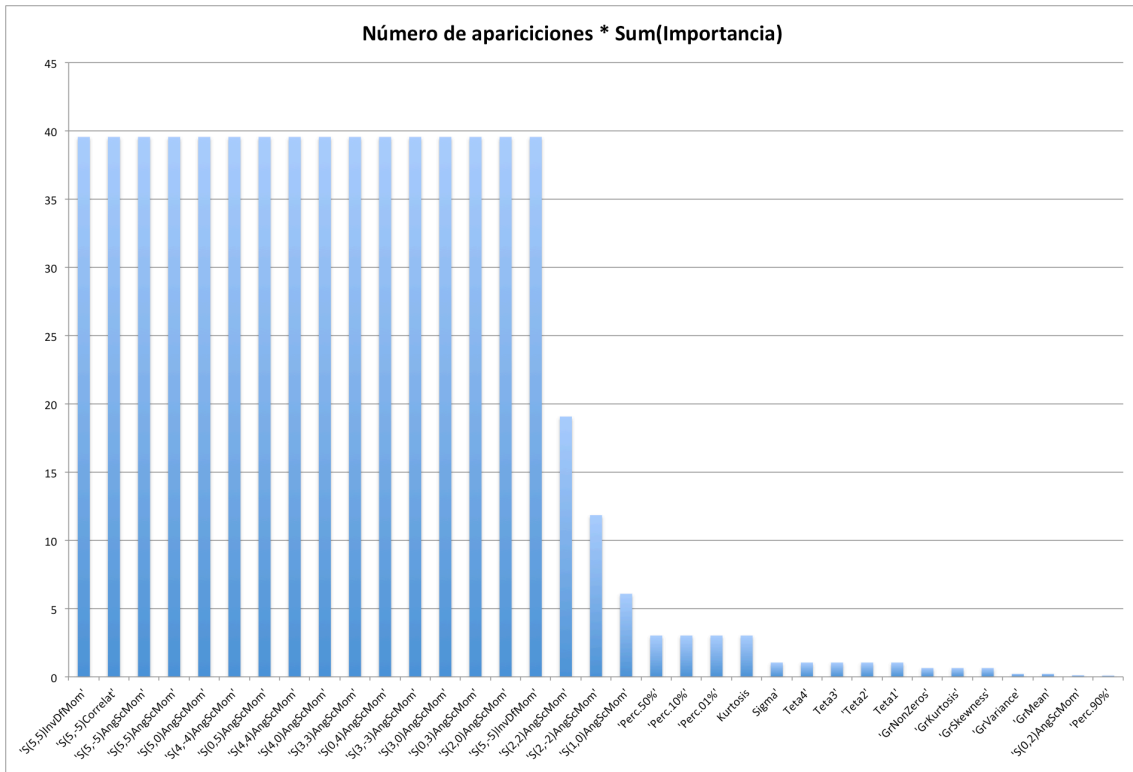


FIGURA 61. IMPORTANCIA DE CADA VARIABLE EN FUNCIÓN DEL NÚMERO DE APARICIONES QUE HA TENIDO ENTRE TODOS LOS KERNELS Y LA IMPORTANCIA DE DICHO KERNEL EN LA SOLUCIÓN FINAL

En la Tabla 36 se presentan agrupadas por grupo de pertenencia las variables que han sido seleccionadas por el FSMKL en los diferentes kernels solución. Puede verse como de los 6 grupos de variables de textura que se han extraído originalmente y presentados en la Tabla 4, esta técnica ha descartado a dos de ellos (Matriz de Longitud de Carreras y Wavelet). El grupo que presenta el mayor número de variables es el compuesto por la Matriz de Coocurrencia, siendo además el grupo cuyas variables tienen más influencia en el resultado final. Es más, de las variables que pertenecen al grupo de la matriz de coocurrencia es el Momento Angular Segundo la que aparece mayor número de veces, y con mayor peso en la solución del problema, calculado con diferentes opciones de longitud y ángulo.

TABLA 36. DISTANCIA BHATTACHARYYA: VARIABLES FINALES

Grupo	Variables	Número
Histograma	Perc.10%, Perc.01%, Perc.50%, Perc.90%, Kurtosis	5
Gradiente absoluto	GrKurtosis, GrMean, GrSkewness, GrVariance, GrNonZeros	5
Matriz de Coocurrencia	S(5,5)InvDfMom, S(5,-5)Correlat, S(5,-5)AngScMom, S(5,5)AngScMom, S(5,0)AngScMom, S(4,-4)AngScMom, S(0,5)AngScMom, S(4,4)AngScMom, S(4,0)AngScMom, S(3,3)AngScMom, S(0,4)AngScMom, S(3,-3)AngScMom, S(3,0)AngScMom, S(0,3)AngScMom, S(2,0)AngScMom, S(5,-5)InvDfMom, S(2,2)AngScMom, S(2,-2)AngScMom, S(1,0)AngScMom, S(0,2)AngScMom	20
Modelo Autoregresivo	Theta1, Theta2, Theta3, Theta4, Sigma	5

Esta Tesis Doctoral propone a las aproximaciones *filter* con las distancias Bhattacharyya y Entropía Relativa como más robustas para la selección de variables de textura en imagen biomédica.

5.2.5 KERNEL MAHALANOBIS

Como ya se ha adelantado es posible que el kernel RBF mejore los resultados que obtiene al utilizar una medida de distancia para la separación de los datos que no sea la euclídea. En concreto, se han realizado pruebas con la medida de Mahalanobis. Más particularmente, se estudiará una implementación específica del *kernel* Mahalanobis para espacios alta dimensionalidad propuesto por Fauvel et. al [246].

El cálculo de la distancia de Mahalanobis requiere conocer la matriz de covarianzas inversa. Este cálculo en espacios de alta dimensionalidad puede llegar a ser muy costoso computacionalmente, por lo que Fauvel et. al proponen el uso de un modelo estadístico parsimonioso, conocido como modelo de análisis discriminante de alta dimensionalidad o *high dimensional discriminant analysis* (HDDA), para calcular subespacios específicos y poder así calcular la matriz de covarianzas inversa, de manera explícita y estable. Los parámetros del *kernel* son estimados como en [247]. En espacios de alta dimensionalidad los ejemplos tienden a tener una distancia constante entre ellos [248], por lo que según [249,250] podría concluirse que la distancia euclídea no es la más adecuada para estos espacios.

Según esto, el *kernel* Mahalanobis regularizado se construye sustituyendo la distancia Mahalanobis calculada con HDDA para la clase c

$$(EQ. 42) \quad d_{\Sigma_c}^2(x, z) = \sum_{i=1}^{\hat{p}_c} \left(\frac{1}{\hat{\lambda}_{ci}} - \frac{1}{\hat{b}_c} \right) \|\hat{q}_{ci}^t(x - z)\|^2 + \frac{\|x - z\|^2}{\hat{b}_c}$$

siendo λ_i el vector de eigenvalores, q_{ci} el vector de eigenvectores y b_c los últimos $d - p_c$ eigenvalores. HDDA asume que los primeros p_c eigenvalues son diferentes y que el resto son iguales. Por lo tanto, sustituyendo la distancia euclídea por (Eq. 42) en (Eq. 17), y cambiando los eigenvalores $(\hat{\lambda}_{ci}, \hat{b}_c)$ por los hiperparámetros $(\sigma_{ci}^2, \sigma_{c\hat{p}_c+1}^2)$ previamente optimizados, nos queda definido el *kernel* Mahalanobis parsimonioso según

$$(EQ. 43) \quad k_m(x, z|c) = \exp \left(-\frac{1}{2} \left(\sum_{i=1}^{\hat{p}_c} \frac{\|\hat{q}_{ci}^t(x - z)\|^2}{\sigma_{ci}^2} + \frac{\|x - z\|^2}{\sigma_{c\hat{p}_c+1}^2} \right) \right)$$

Se recomienda la lectura de [246,251,252] para una mejor comprensión acerca de HDDA y de esta implementación en concreto del *kernel* Mahalanobis. Este *kernel* está diseñado para problemas de alta dimensionalidad con un número mínimo de variables igual a 200.

Los resultados de los diez experimentos llevados a cabo se muestran en la Tabla 37, se presentan los valores de porcentaje de acierto, AUROC y tiempo de ejecución de cada experimento siguiendo un proceso de *10-fold cross validation*.

TABLA 37. KERNEL MAHALANOBIS PARSIMONIOSO: RESULTADOS DE LOS 10 EXPERIMENTOS

Exp.	Acierto Validación	AUROC Validación	Tiempo (min)
1	83.27	84.30	147.49
2	80.80	82.30	148.20
3	81.00	82.11	157.69
4	82.03	85.32	170.68
5	81.97	83.23	169.20
6	83.11	84.01	144.79
7	80.76	82.97	162.05
8	81.41	83.33	158.41
9	81.97	85.02	168.51
10	80.32	81.30	165.61

Estos resultados son peores que cualquiera de los presentados previamente en este Tesis Doctoral por los demás métodos, por lo que se puede concluir que este *kernel* no es adecuado para el análisis de textura. Es posible que el número de variables no sea lo suficientemente alto para sacar todo el rendimiento a esta técnica, ya que los autores proponen utilizar con problemas que cuenten con un mínimo de 200 variables.

5.3 DISCUSIÓN

A la vista de los resultados presentados en el apartado 5.1.4, el mejor modelo para generalizar información textural es el FSMKL. Valorando tanto el porcentaje de acierto como el número de variables que selecciona en función del peso de los kernel que genera el modelo, se puede concluir que la mejor solución la encuentra con la combinación del *kernel* RBF y el polinomial.

En la Figura 40 se puede ver como el 56,56% de la variación en el FSMKL puede ser explicada con dos kernel compuestos por dos variables y el 64,00% con 3 kernels (el tercer kernel incluye una variable distinta solamente de los dos primeros). Lo que es más interesante de esta solución es el número de veces que aparece la variable de Momento Diferencial Inverso (MDI), puede verse en la Figura 41.

En el apartado 5.2.2 se compararon los resultados de los nuevos kernels propuestos en esta Tesis Doctoral, siendo el FSMKL-Producto significativamente mejor que el FSMKL-Suma y la combinación lineal de *kernels* RBF y polinomial. En la Figura 48, los once primeros kernels explicaban el 61,90% de la variación en el FSMKL y dichos kernels tienen once variables que se corresponden con distintos ángulos y distancias de la variable Momento Diferencial Inverso (MDI).

EL MDI es una medida de homogeneidad local y se define según la siguiente ecuación

$$(EQ. 44) \quad MDI = \sum_{i=0}^{N_G} \sum_{j=0}^{N_G} \frac{1}{1 + (i - j)^2} P(i, j)$$

Esta medida se calcula mediante la matriz de GLCM y debido al factor de peso del denominador, se obtienen pequeñas contribuciones de áreas no homogéneas donde

($i \neq j$). El resultado es un valor de MDI bajo para imágenes no homogéneas y un valor relativamente alto para las variables homogéneas.

En el apartado 5.2.4 se presentaron los resultados obtenidos con las nuevas medidas de distancia para estudiar la correlación entre las variables de entrada y la clase de salida propuesta en esta Tesis Doctoral. Se presentó la Figura 60, en la que los dos primeros kernels explicaban el 58,12% de la variación en el FSMKL y tenían las mismas 19 variables. En la Figura 61 se presentó una gráfica en la que las primeras 16 variables siempre aparecían juntas y en un total de 9 de los 17 kernels que eran solución al problema. Dichas variables se corresponden con distintos ángulos y distancias de la variable Momento Angular Segundo (MAS).

El MAS es una medida de uniformidad en la imagen, de tal manera que aplicado sobre una imagen homogénea, se obtendrá una GLCM con pocos valores, pero muy altos. Se calcula con la siguiente ecuación

$$(EQ. 45) \quad MAS = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\}^2$$

Tanto en la búsqueda del mejor método, como en el estudio de nuevos *kernels* y distancias propuestos en esta Tesis Doctoral, aparece de manera recurrente y con diferentes valores de importancia las variables pertenecientes al grupo Histograma, Gradiente Absoluto y Autoregresivo

Esto quiere decir que según el FSMKL y a nivel general, las variables que tienen más peso, son MDI y MAS que son las que más información le dan al mejor modelo para generalizar soluciones y clasificar proteínas en imágenes de electroforesis bidimensional mediante información procedente de un análisis de texturas. Ambas son medidas para cuantificar la homogeneidad en la imagen [15].

6. CONCLUSIONES

En esta Tesis Doctoral se plantea el uso de técnicas basadas en *kernel* para el análisis de espacios de entrada de alta dimensionalidad y con conjuntos de datos heterogéneos en un escenario como el del análisis de texturas en imagen biomédica. A lo largo del desarrollo de esta Tesis Doctoral y partiendo de los objetivos marcados, se han evaluado diferentes técnicas basadas en *kernel* para resolver el problema, tanto de forma directa como de manera combinada con técnicas de optimización bioinspiradas como los algoritmos genéticos y la optimización por enjambre de partículas.

Este tipo de técnicas han demostrado que pueden resolver este tipo de problemas y, además, el uso combinado de ellas con técnicas de selección de variables las hacen unas herramientas muy potentes.

En concreto, en esta Tesis Doctoral se han evaluado aproximaciones basadas en *kernel* para la selección de variables como FSMKL (aproximación *filter*), GA y PSO con SVM (aproximación *wrapper*) y SVM-RFE (aproximación *embedded*). Los resultados obtenidos por la aproximación *filter* han resultado ser significativamente mejores que el resto en base a la comparación hecha con diferentes test estadísticos. También, como era de esperar, el tiempo necesario para llevar a cabo un experimento con esta aproximación ha resultado ser el más reducido.

Esta Tesis Doctoral propone nuevos kernels para el análisis de textura que mejoran los resultados obtenidos con *kernels* convencionales como RBF y polinomial y con combinaciones lineales de *kernels* como los obtenidos con MKL de estos *kernels*. De los propuestos, uno de los *kernels* resultó ser significativamente mejor que el resto en base a la comparación hecha con diferentes test no-paramétricos. Esta Tesis ha estudiado la viabilidad del uso de un *kernel* Mahalanobis parsimonioso y ha determinado que no es adecuado para resolver un análisis de texturas.

Esta Tesis Doctoral ha propuesto, para el algoritmo de selección de variables *filter* referencia, la modificación y utilización de diferentes medidas de distancia para establecer el ranking de las variables. Cuatro de las medidas propuestas resultaron ser significativamente mejores que el resto, en base a la comparación hecha con diferentes test no-paramétricos. Esta Tesis Doctoral propone la distancia de Bhattacharyya como la mejor para ser utilizada en un análisis de texturas.

La técnica aquí propuesta muestra cómo es posible realizar un análisis de texturas sobre imágenes de geles de electroforesis bidimensional para clasificar regiones de interés correspondientes con proteínas y no proteínas. Se trata de una tarea muy complicada debido a la elevada inter- e intra-variabilidad entre diferentes clínicos a la hora de marcar manualmente las zonas a estudiar. Como conclusión, con este tipo de datos, se obtiene que de todo el espacio de variables de entrada, de manera destacada las variables de textura que permiten clasificar entre proteína y no proteína, son el *momento diferencial inverso* y el *segundo momento angular*, que son medidas de la homogeneidad de la imagen. Son operadores estadísticos de segundo orden que se calculan de la matriz de coocurrencia de los niveles de gris.

Como se mencionó con anterioridad, existe una controversia entre los operadores de segundo orden y orden superior. Hay autores que consideran a los de orden superior iguales o mejores mientras que otros autores los consideran menos eficientes. Esta Tesis propone a los operadores de segundo orden como las mejores para discriminar proteínas en un análisis de texturas en imágenes de geles de electroforesis bidimensional y, además, establece a la variable Momento Diferencial Inverso, derivada de la matriz de coocurrencia de niveles de gris, propuesta por Haralick, como la más importante para hacerlo.

Finalmente, el diseño experimental propuesto se ha mostrado como una aportación crucial para comparar en igualdad de condiciones diferentes técnicas sobre un mismo conjunto de datos. Es además una propuesta que se puede generalizar y tomar como buena práctica de trabajo en cualquier tipo de experimentación que se esté realizando, puesto que las fases que se incluyen no son obligatorias y permiten adaptarse al dominio concreto del problema.

7. CONCLUSIONS

In this thesis, the use of *kernel*-based techniques is considered for the analysis of high-dimensionality input spaces and of heterogeneous data sets in a scenario such as the one of texture analysis in biomedical imaging. Throughout the development of this thesis and starting from the set objectives, different *kernel*-based techniques have been assessed in order to solve the problem, both directly and combined with bio-inspired optimization techniques such as genetic algorithms and particle swarm optimization.

Such techniques have been proven to solve these issues and, in addition, their use in combination with variable selection techniques makes them very powerful tools.

More specifically, in this thesis, *kernel*-based approaches have been evaluated to select variables such as FSMKL (*filter* approach), GA and PSO with SVM (*wrapper* approach) and SVM-RFE (*embedded* approach). The results obtained by the *filter* approach have proven to be significantly better than the others, taking into account the comparison made with different statistical tests. Moreover, as expected, the time required to conduct an experiment with this approach has proven to be the shortest.

This thesis proposes new *kernels* for texture analysis to improve the results obtained with conventional *kernels* like RBF and polynomial *kernels*, and with linear combinations of *kernels* such as those obtained with MKL of these *kernels*. Out of those proposed, one of the *kernels* proved to be significantly better than the others taking into account the comparison made with different non-parametric tests. This thesis has studied the feasibility of using a parsimonious Mahalanobis *kernel* and has determined that it is not appropriate for carrying out a texture analysis.

This thesis proposes, for the selection algorithm of studied *filter* variables, the modification and use of different distance measures to establish the ranking of the variables in the pre-classification process. Four of the proposed measures proved to be significantly better than the rest, taking into account the comparison made with

different non-parametric tests. This work proposes the Bhattacharyya distance as the most appropriate to be used for texture analysis.

The technique proposed herein shows how a texture analysis can be performed on two-dimensional electrophoresis gel images to classify regions of interest corresponding to proteins and non-proteins. This is a very difficult task because of the high inter-and intra-variability among different clinicians as they have to mark manually the areas to be studied. With this type of data it can be concluded that from the entire space of input variables, the texture variables, which notably allow distinguishing between proteins and non-proteins, are *the inverse differential moment and the second angular momentum*, which are measures of the homogeneity of the image. They are second-order statistical operators which are calculated from the co-occurrence matrix of gray levels.

As mentioned previously, there is a dispute between the second-order operators and higher order ones, in the sense that some authors consider the latter as good as or better than the former, while others consider them less efficient. This thesis supports the approach of second-order operators as better distinguishing between proteins in a texture analysis on two-dimensional electrophoresis gel images; it further establishes the variable of *Inverse Differential Moment*, derived from the co-occurrence matrix of gray levels, which Haralick proposed as the most appropriate to use.

Finally, the proposed experimental design has proven to be a vital contribution to compare different techniques on the same dataset, under the same conditions. This is also a proposal that can be generalized and kept as good working practice for any type of experimentation being performed, since the included phases are not mandatory and can be adapted to the specific problem domain.

8. FUTUROS DESARROLLOS

Los buenos resultados obtenidos con las técnicas basadas en *kernel* propuestas en esta Tesis en el caso de estudio, hacen pensar que los futuros desarrollos vayan encaminados, por un lado, a la integración de la información obtenida con técnicas de procesado de imagen. Mientras que, por otro, se podría avanzar en el diseño de nuevas técnicas basadas en *kernel* para fusión de datos biomédicos. La propuesta presentada es una aproximación genérica que puede ser utilizada en diversos dominios para integración de datos sobre conjuntos heterogéneos de datos.

Un ejemplo de futuro desarrollo de la propuesta presentada en esta Tesis Doctoral tiene en cuenta que las técnicas basadas en *kernel* actuales sólo pueden usar datos que estén expresados en términos de relaciones directas con la enfermedad. Con la aparición de estudios genéticos a escala genómica, se puede realizar una nueva clasificación e incluso subclasificaciones de enfermedad. Por ejemplo, recientemente se ha propuesto la clasificación de enfermedades autoinmunes pasando de fenotipos a una reclasificación basada en fenotipos biológicos y genéticos [253]. Si eso se pudiese hacer para todo tipo de enfermedad, se llegaría a una mejor forma para establecer los posibles alineamientos entre enfermedades, lo que conduciría a una selección más acertada de terapias.

Existen clasificaciones bien conocidas y establecidas de enfermedad y ontologías de las enfermedades humanas como *Disease Ontology (DO)* [254], que integra nomenclatura de enfermedades con la inclusión de un mapeado cruzado con términos e identificadores de la *Medical Subject Headings (MeSH)* [255], *World Health Foundation (WHO)*, *International Classification of Diseases (ICD)* [256], *Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)* [257], *National Cancer Institute (NCI) Thesaurus* [258] y *Online Mendelian Inheritance in Man (OMIM)*[259] que clasifica enfermedades humanas basándose en análisis patológico y síntomas clínicos. Sin embargo, hay un número creciente de fuentes de datos heterogéneas como genómicos,

proteómicos, transcriptómicos y metabolómicos que no son utilizados para ayudar en esta clasificación. Surgen trabajos que enfatizan la importancia de redefinir la enfermedad humana como parte de una mejora en la prognosis y en la respuesta a terapias [260].

La integración de diferentes conjuntos de información heterogéneos se conoce como fusión de datos [8]. Las aproximaciones más utilizadas, normalmente siguen una aproximación de integración temprana o tardía, en función de si combinan entradas [261] o predicciones [262], respectivamente. Existe una tercera posibilidad que preserva la estructura de los datos infiriendo un único modelo, esta aproximación es conocida como intermedia [86,87,90,121,263]. El objetivo de un futuro desarrollo sería el diseño de una aproximación computacional que permita la fusión de diferentes conjuntos de datos con información molecular para tratar de encontrar nuevas asociaciones entre enfermedades.

El máximo exponente de esta aproximación intermedia para fusión de datos es conocido como *Multiple Kernel Learning* que es, junto con SVM, el núcleo de la técnica aquí presentada. La fusión de datos ha sido utilizada con éxito para tareas como la priorización de genes [8,89,264] y la reconstrucción de redes de genes y predicción de función [261,265].

Las ventajas que tiene seguir una aproximación intermedia son, entre otras, su acierto a la hora de inferir modelos de predicción (como de bien un modelo puede aprender a predecir asociaciones enfermedad-enfermedad) y, sobre todo, la posibilidad de medir explícitamente la contribución de cada conjunto de datos heterogéneo en el conocimiento generado [86,87].

Los objetivos específicos y científicos que se plantean alcanzar en los futuros desarrollos a seguir a partir del trabajo comenzado en esta Tesis Doctoral serían los de desarrollar nuevas técnicas para clasificación/regresión de datos biomédicos heterogéneos para la integración de conjuntos de datos multinivel y estudio de la correlación entre los diferentes niveles de expresión en conjuntos de datos “-ómicos”.

REFERENCIAS

1. FAN J, LV J (2010) A SELECTIVE OVERVIEW OF VARIABLE SELECTION IN HIGH DIMENSIONAL FEATURE SPACE. *STAT SIN* 20: 101-148.
2. CHAPELLE O, HAFFNER P, VAPNIK VN (1999) SUPPORT VECTOR MACHINES FOR HISTOGRAM-BASED IMAGE CLASSIFICATION. *IEEE TRANSACTIONS ON NEURAL NETWORKS* 10: 1055-1064.
3. MOULIN LS, DA SILVA APA, EL-SHARKAWI MA, MARKS RJ, II (2004) SUPPORT VECTOR MACHINES FOR TRANSIENT STABILITY ANALYSIS OF LARGE-SCALE POWER SYSTEMS. *POWER SYSTEMS, IEEE TRANSACTIONS ON* 19: 818-825.
4. FERNANDEZ-LOZANO C, SEOANE JA, GESTAL POSE M, RIVERO D, DORADO J, ET AL. A TEXTURE-BASED CLASSIFICATION METHOD FOR PROTEINS IN TWO-DIMENSIONAL ELECTROPHORESIS GEL IMAGES; 2013; BARCELONA. SCITEPRESS. PP. 401-404.
5. FERNANDEZ-LOZANO C, SEOANE JA, GESTAL POSE M, GAUNT TR, CAMPBELL C. TEXTURE CLASSIFICATION USING KERNEL-BASED TECHNIQUES. IN: IGNACIO ROJAS GJ, JOAN CABESTANY, EDITOR; 2013; PUERTO DE LA CRUZ. TENERIFE (SPAIN). SPRINGER HEIDELBERG. PP. 427-434.
6. CAMPBELL C, YING Y (2011) LEARNING WITH SUPPORT VECTOR MACHINES. *SYNTHESIS LECTURES ON ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING* 5: 1-95.
7. SONNENBURG S, RÄTSCH G, SCHÄFER C, SCHÖLKOPF B (2006) LARGE SCALE MULTIPLE KERNEL LEARNING. *J MACH LEARN RES* 7: 1531-1565.
8. YU S, FALCK T, DAEMEN A, TRANCHEVENT L-C, SUYKENS JAK, ET AL. (2010) L2-NORM MULTIPLE KERNEL LEARNING AND ITS APPLICATION TO BIOMEDICAL DATA FUSION. *BMC BIOINFORMATICS* 11: 309.
9. TRETYAKOV K (2006) METHODS OF GENOMIC DATA FUSION: AN OVERVIEW.
10. UBELYI ED (2009) ADAPTIVE NEURO-FUZZY INFERENCE SYSTEMS FOR AUTOMATIC DETECTION OF BREAST CANCER. *J MED SYST* 33: 353-358.
11. SMOLA AJ (1998) LEARNING WITH KERNELS. BERLIN, GERMANY: TECHICAL UNIVERSITY.
12. GUNN SR (1998) SUPPORT VECTOR MACHINES FOR CLASSIFICATION AND REGRESSION. DEPT. OF ELECTRONICS AND COMPUTER SCIENCE, UNIVERSITY OF SOUTHAMPTON.
13. TUCERYAN M, JAIN A (1999) TEXTURE ANALYSIS. *HANDBOOK OF PATTERN RECOGNITION AND COMPUTER VISION*: WORLD SCIENTIFIC PUBLISHING COMPANY, INCORPORATED.
14. COGGINS JM (1983) A FRAMEWORK FOR TEXTURE ANALYSIS BASED ON SPATIAL FILTERING: MICHIGAN STATE UNIVERSITY. 182 P.
15. HARALICK RM, SHANMUGAM K, DINSTEN I (1973) TEXTURAL FEATURES FOR IMAGE CLASSIFICATION. *IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS SMC* 3: 610-621.
16. BRODATZ P (1999) TEXTURES: A PHOTOGRAPHIC ALBUM FOR ARTISTS AND DESIGNERS: DOVER PUBLICATIONS, INCORPORATED.
17. HARALICK RM, SHAPIRO LG (1992) *COMPUTER AND ROBOT VISION*: ADDISON-WESLEY LONGMAN PUBLISHING Co., INC. 630 P.
18. MATERKA A, STRZELECKI M (1998) TEXTURE ANALYSIS METHODS-A REVIEW. TECHNICAL UNIVERSITY OF LODZ, INSTITUTE OF ELECTRONICS COST B11 REPORT.

19. JAIN AK (1989) FUNDAMENTALS OF DIGITAL IMAGE PROCESSING: PRENTICE-HALL, INC. 569 P.
20. JULESZ B (1975) EXPERIMENTS IN THE VISUAL PERCEPTION OF TEXTURE. SCI AM 232: 34-43.
21. GANESHAN B, PANAYIOTOU E, BURNAND K, DIZDAREVIC S, MILES K (2011) TUMOUR HETEROGENEITY IN NON-SMALL CELL LUNG CARCINOMA ASSESSED BY CT TEXTURE ANALYSIS: A POTENTIAL MARKER OF SURVIVAL. EUROPEAN RADIOLOGY 22: 796-802.
22. LEVINE MD (1985) VISION IN MAN AND MACHINE: MCGRAW-HILL.
23. ROSENFELD A, WESZKA JS (1980) PICTURE RECOGNITION. IN: FU K, EDITOR. DIGITAL PATTERN RECOGNITION: SPRINGER BERLIN HEIDELBERG. PP. 135-166.
24. MALLAT SG (1989) MULTIFREQUENCY CHANNEL DECOMPOSITIONS OF IMAGES AND WAVELET MODELS. ACOUSTICS, SPEECH, AND SIGNAL PROCESSING [SEE ALSO IEEE TRANSACTIONS ON SIGNAL PROCESSING], IEEE TRANSACTIONS ON 37: 2091-2110.
25. LU CS, CHUNG PC, CHEN CF (1997) UNSUPERVISED TEXTURE SEGMENTATION VIA WAVELET TRANSFORM. PATTERN RECOGNITION 30: 729-742.
26. DAUGMAN J (1985) UNCERTAINTY RELATION FOR RESOLUTION IN SPACE, SPATIAL FREQUENCY AND ORIENTATION OPTIMISED BY TWO-DIMENSIONAL VISUAL CORTICAL FILTERS. JOURNAL OF THE OPTICAL SOCIETY OF AMERICA 2: 1160-1169.
27. SUPER BJ, BOVIK AC (1991) LOCALIZED MEASUREMENT OF IMAGE FRACTAL DIMENSION USING GABOR FILTERS. JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION 2: 114-128.
28. LAINE A, FAN J (1993) TEXTURE CLASSIFICATION BY WAVELET PACKET SIGNATURES. PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON 15: 1186-1191.
29. WESZKA JS, DYER CR, ROSENFELD A (1976) A COMPARATIVE STUDY OF TEXTURE MEASURES FOR TERRAIN CLASSIFICATION. SYSTEMS, MAN AND CYBERNETICS, IEEE TRANSACTIONS ON SMC-6: 269-285.
30. NIEMANN H (1981) PATTERN ANALYSIS: SPRINGER.
31. HARALICK RM (1979) STATISTICAL AND STRUCTURAL APPROACHES TO TEXTURE. PROCEEDINGS OF THE IEEE 67: 786-804.
32. LEVINA E (2002) STATISTICAL ISSUES IN TEXTURE ANALYSIS: UNIVERSITY OF CALIFORNIA, BERKELEY.
33. LAWS KI (1980) TEXTURED IMAGE SEGMENTATION: UNIVERSITY OF SOUTHERN CALIFORNIA.
34. TOMITA F, TSUJI S (1990) COMPUTER ANALYSIS OF VISUAL TEXTURES: KLUWER ACADEMIC PUBLISHERS. 173 P.
35. TSATSANIS MK (1992) OBJECT AND TEXTURE CLASSIFICATION USING HIGHER ORDER STATISTICS. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 14: 733-750.
36. PRATT WK (1978) DIGITAL IMAGE PROCESSING: WILEY.
37. PAPOULIS A (1984) PROBABILITY, RANDOM VARIABLES, AND STOCHASTIC PROCESSES: MCGRAW-HILL.
38. LERSKI RA, STRAUGHAN K, SCHAD LR, BOYCE D, BLÜML S, ET AL. (1993) MRIMAGE TEXTURE ANALYSIS--AN APPROACH TO TISSUE CHARACTERIZATION. MAGN RESON IMAGING 11: 873-887.

39. GALLOWAY MM (1975) TEXTURE ANALYSIS USING GRAY LEVEL RUN LENGTHS. COMPUTER GRAPHICS AND IMAGE PROCESSING 4: 172-179.
40. CHU A, SEHGAL CM, GREENLEAF JF (1990) USE OF GRAY VALUE DISTRIBUTION OF RUN LENGTHS FOR TEXTURE ANALYSIS. PATTERN RECOGNITION LETTERS 11: 415-419.
41. NIKIAS CL, MENDEL JM (1993) SIGNAL PROCESSING WITH HIGHER-ORDER SPECTRA. SIGNAL PROCESSING MAGAZINE, IEEE 10: 10-37.
42. HALL TE, GIANNAKIS GB. TEXTURE MODEL VALIDATION USING HIGHER-ORDER STATISTICS; 1991 14-17 APR 1991. PP. 2673-2676 VOL.2674.
43. HINICH MJ (1982) TESTING FOR GAUSSIANTY AND LINEARITY OF A STATIONARY TIME SERIES. JOURNAL OF TIME SERIES ANALYSIS 3: 169-176.
44. GIANNAKIS GB, TSATSANIS MK (1990) SIGNAL DETECTION AND CLASSIFICATION USING MATCHED FILTERING AND HIGHER ORDER STATISTICS. ACOUSTICS, SPEECH AND SIGNAL PROCESSING, IEEE TRANSACTIONS ON 38: 1284-1296.
45. RAMPONI G, CARRATO S. MODIFIED FOURTH-ORDER MOMENTS IN TEXTURE RECOGNITION; 1991; CHAMROUSSE, FRANCE. PP. 139-142.
46. MAKOVEC A, RAMPONI G. HIGHER ORDER MOMENTS FOR TEXTURE SEGMENTATION IN PRESENCE OF NOISE; 1994; EDINBURGH.
47. CONNERS RW, HARLOW CA (1980) A THEORETICAL COMPARISON OF TEXTURE ALGORITHMS. IEEE TRANS PATTERN ANAL MACH INTELL 2: 204-222.
48. TANG X (1998) TEXTURE INFORMATION IN RUN-LENGTH MATRICES. IEEE TRANS IMAGE PROCESS 7: 1602-1609.
49. BAGCI U, YAO J, MILLER-JASTER K, CHEN X, MOLLURA DJ (2013) PREDICTING FUTURE MORPHOLOGICAL CHANGES OF LESIONS FROM RADIOTRACER UPTAKE IN 18F-FDG-PET IMAGES. PLOS ONE 8: e57105.
50. BLUM C, ROLI A (2003) METAHEURISTICS IN COMBINATORIAL OPTIMIZATION: OVERVIEW AND CONCEPTUAL COMPARISON. ACM COMPUT SURV 35: 268-308.
51. DRÄGER A, KRONFELD M, ZILLER MJ, SUPPER J, PLANATSCHER H, ET AL. (2009) MODELING METABOLIC NETWORKS IN C. GLUTAMICUM: A COMPARISON OF RATE LAWS IN COMBINATION WITH VARIOUS PARAMETER OPTIMIZATION STRATEGIES. BMC SYSTEMS BIOLOGY 3: 5.
52. DARWIN C (1859) ON THE ORIGIN OF THE SPECIES BY MEANS OF NATURAL SELECTION: JOHN MURRAY.
53. HOLLAND JH (1975) ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS: AN INTRODUCTORY ANALYSIS WITH APPLICATIONS TO BIOLOGY, CONTROL, AND ARTIFICIAL INTELLIGENCE: UNIVERSITY OF MICHIGAN PRESS.
54. GOLDBERG D (1989) GENETIC ALGORITHMS IN SEARCH, OPTIMIZATION, AND MACHINE LEARNING: ADDISON-WESLEY PROFESSIONAL.
55. REEVES CR (1993) MODERN HEURISTIC TECHNIQUES FOR COMBINATORIAL PROBLEMS: JOHN WILEY AND SONS, INC. 320 P.
56. MICHALEWICZ Z, FOGEL D (2004) HOW TO SOLVE IT: MODERN HEURISTICS: SPRINGER.
57. WRIGHT A (1991) GENETIC ALGORITHMS FOR REAL PARAMETER OPTIMIZATION. IN: RAWLINS G, EDITOR. FOUNDATIONS OF GENETIC ALGORITHMS: MORGAN KAUFMANN. PP. 205-218.
58. KOZA JR (1992) GENETIC PROGRAMMING: ON THE PROGRAMMING OF COMPUTERS BY MEANS OF NATURAL SELECTION: MIT PRESS. 680 P.

59. NOWAK M, SCHUSTER P (1989) ERROR THRESHOLDS OF REPLICATION IN FINITE POPULATIONS MUTATION FREQUENCIES AND THE ONSET OF MULLER'S RATCHET. JOURNAL OF THEORETICAL BIOLOGY 137: 375-395.
60. SCHUSTER P (1985) EFFECT OF FINITE POPULATION SIZE AND OTHER STOCHASTIC PHENOMENON IN MOLECULAR EVOLUTION; SPRINGER, EDITOR. BERLIN, NEW YORK, HEIDELBERG: SPRINGER.
61. JONG KAD (1975) AN ANALYSIS OF THE BEHAVIOR OF A CLASS OF GENETIC ADAPTIVE SYSTEMS: UNIVERSITY OF MICHIGAN. 266 P.
62. SCHAFFER JD, CARUANA RA, ESHELMAN LJ, DAS R (1989) A STUDY OF CONTROL PARAMETERS AFFECTING ONLINE PERFORMANCE OF GENETIC ALGORITHMS FOR FUNCTION OPTIMIZATION. PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON GENETIC ALGORITHMS. GEORGE MASON UNIVERSITY, UNITED STATES: MORGAN KAUFMANN PUBLISHERS INC. PP. 51-60.
63. FOGARTY TC (1989) VARYING THE PROBABILITY OF MUTATION IN THE GENETIC ALGORITHM. PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON GENETIC ALGORITHMS. GEORGE MASON UNIVERSITY, USA: MORGAN KAUFMANN PUBLISHERS INC. PP. 104-109.
64. DAVIS L (1985) APPLYING ADAPTIVE ALGORITHMS TO EPISTATIC DOMAINS. PROCEEDINGS OF THE 9TH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE - VOLUME 1. LOS ANGELES, CALIFORNIA: MORGAN KAUFMANN PUBLISHERS INC. PP. 162-164.
65. KENNEDY J, EBERHART R. PARTICLE SWARM OPTIMIZATION; 1995 NOV/DEC 1995. PP. 1942-1948 VOL.1944.
66. BANKS A, VINCENT J, ANYAKOHA C (2007) A REVIEW OF PARTICLE SWARM OPTIMIZATION. PART I: BACKGROUND AND DEVELOPMENT. 6: 467-484.
67. SHI Y, EBERHART R (1998) A MODIFIED PARTICLE SWARM OPTIMIZER. 69-73.
68. LIN S-W, YING K-C, CHEN S-C, LEE Z-J (2008) PARTICLE SWARM OPTIMIZATION FOR PARAMETER DETERMINATION AND FEATURE SELECTION OF SUPPORT VECTOR MACHINES. EXPERT SYSTEMS WITH APPLICATIONS 35: 1817-1824.
69. ALPAYDIN E (2010) INTRODUCTION TO MACHINE LEARNING: THE MIT PRESS. 584 P.
70. VAPNIK VN (1995) THE NATURE OF STATISTICAL LEARNING THEORY: SPRINGER-VERLAG NEW YORK, INC. 188 P.
71. BURGESS CJC (1998) A TUTORIAL ON SUPPORT VECTOR MACHINES FOR PATTERN RECOGNITION. DATA MINING AND KNOWLEDGE DISCOVERY 2: 121-167.
72. SCHÖLKOPF BS, SMOLA AJ (2002) LEARNING WITH KERNELS: SUPPORT VECTOR MACHINES, REGULARIZATION, OPTIMIZATION AND BEYOND: UNIVERSITY PRESS GROUP LIMITED.
73. SHAWE-TAYLOR J, CRISTIANINI N (2004) KERNEL METHODS FOR PATTERN ANALYSIS: CAMBRIDGE UNIVERSITY PRESS.
74. CRISTIANINI N, SHAWE-TAYLOR J (2000) AN INTRODUCTION TO SUPPORT VECTOR MACHINES: AND OTHER KERNEL-BASED LEARNING METHODS: CAMBRIDGE UNIVERSITY PRESS. 189 P.
75. VAPNIK VN (1979) ESTIMATION OF DEPENDENCES BASED ON EMPIRICAL DATA [IN RUSSIAN]. NAUKA: ENGLISH TRANSLATION SPRINGER VERLANG, 1982.
76. SCHÖLKOPF B, TSUDA K, VERT JP (2004) KERNEL METHODS IN COMPUTATIONAL BIOLOGY: BRADFORD BKS.

77. MOULIN LS, ALVES DA SILVA AP, EL-SHARKAWI MA, MARKS II RJ (2004) SUPPORT VECTOR MACHINES FOR TRANSIENT STABILITY ANALYSIS OF LARGE-SCALE POWER SYSTEMS. *IEEE TRANSACTIONS ON POWER SYSTEMS* 19: 818-825.
78. DUDA RO, HART PE, STORK DG (2001) *PATTERN CLASSIFICATION: WILEY.*
79. CORTES C, VAPNIK V. *SUPPORT-VECTOR NETWORKS*; 1995. pp. 273-297.
80. SCHÖLKOPF B, MIKA S, BURGESS CJC, KNIRSCH P, MULLER KR, ET AL. (1999) INPUT SPACE VERSUS FEATURE SPACE IN KERNEL-BASED METHODS. *IEEE TRANSACTIONS ON NEURAL NETWORKS* 10: 1000-1017.
81. MERCER J (1909) FUNCTIONS OF POSITIVE AND NEGATIVE TYPE, AND THEIR CONNECTION WITH THE THEORY OF INTEGRAL EQUATIONS. *PROCEEDINGS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES* 83: 69-70.
82. ARONSZAJN N (1950) THEORY OF REPRODUCING KERNELS. *TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY* 68.
83. GIROSI F (1997) AN EQUIVALENCE BETWEEN SPARSE APPROXIMATION AND SUPPORT VECTOR MACHINES. MASSACHUSETTS INSTITUTE OF TECHNOLOGY.
84. BERG C, CHRISTENSEN JPR, RESSEL P (1984) *HARMONIC ANALYSIS ON SEMIGROUPS: THEORY OF POSITIVE DEFINITE AND RELATED FUNCTIONS: SPRINGER-VERLAG.*
85. ABE S (2010) *SUPPORT VECTOR MACHINES FOR PATTERN CLASSIFICATION: SPRINGER LONDON, LIMITED.*
86. LANCKRIET GRG, DE BIE T, CRISTIANINI N, JORDAN MI, NOBLE WS (2004) A STATISTICAL FRAMEWORK FOR GENOMIC DATA FUSION. *BIOINFORMATICS* 20: 2626-2635.
87. GEVAERT O, SMET FD, TIMMERMAN D, MOREAU Y, MOOR BD (2006) PREDICTING THE PROGNOSIS OF BREAST CANCER BY INTEGRATING CLINICAL AND MICROARRAY DATA WITH BAYESIAN NETWORKS. *BIOINFORMATICS* 22: e184-e190.
88. TAN P, VAN VLIET MH, HORLINGS HM, VAN DE VIJVER MJ, REINDERS MJT, ET AL. (2012) INTEGRATION OF CLINICAL AND GENE EXPRESSION DATA HAS A SYNERGETIC EFFECT ON PREDICTING BREAST CANCER OUTCOME. *PLOS ONE* 7: e40358.
89. DE BIE T, TRANCHEVENT LC, VAN OEFFELEN LMM, MOREAU Y (2007) KERNEL-BASED DATA FUSION FOR GENE PRIORITIZATION. *BIOINFORMATICS* 23: i125-i132.
90. LANCKRIET GRG, CRISTIANINI N, BARTLETT P, GHAOUI LE, JORDAN MI (2004) LEARNING THE KERNEL MATRIX WITH SEMIDEFINITE PROGRAMMING. *J MACH LEARN RES* 5: 27-72.
91. THERRIEN CW (1989) *DECISION, ESTIMATION, AND CLASSIFICATION: AN INTRODUCTION TO PATTERN RECOGNITION AND RELATED TOPICS: WILEY.*
92. ZONGKER D, JAIN A (1996) ALGORITHMS FOR FEATURE SELECTION: AN EVALUATION. *PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION - VOLUME 2: IEEE COMPUTER SOCIETY.* pp. 18.
93. STEARNS SD. *ON SELECTING FEATURES FOR PATTERN CLASSIFIERS*; 1976. pp. 71-75.
94. JAIN AK, CHANDRASEKARAN B (1982) DIMENSIONALITY AND SAMPLE SIZE CONSIDERATIONS. IN: KRISHNAIAH PR, KANAL LN, EDITORS. *PATTERN RECOGNITION IN PRACTICE.* pp. 835--855.
95. ALONSO-ATIENZA F, ROJO-ÁLVAREZ JL, ROSADO-MUÑOZ A, VINAGRE JJ, GARCÍA-ALBEROLA A, ET AL. (2012) FEATURE SELECTION USING SUPPORT VECTOR

- MACHINES AND BOOTSTRAP METHODS FOR VENTRICULAR FIBRILLATION DETECTION. EXPERT SYSTEMS WITH APPLICATIONS 39: 1956-1967.
96. SAEYS Y, INZA I, LARRAÑAGA P (2007) A REVIEW OF FEATURE SELECTION TECHNIQUES IN BIOINFORMATICS. BIOINFORMATICS 23: 2507-2517.
 97. KOHAVI R, JOHN GH (1997) WRAPPERS FOR FEATURE SUBSET SELECTION. ARTIF INTELL 97: 273-324.
 98. GUYON I, ELISSEFF A (2003) AN INTRODUCTION TO VARIABLE AND FEATURE SELECTION. J MACH LEARN RES 3: 1157-1182.
 99. BLUM AL, LANGLEY P (1997) SELECTION OF RELEVANT FEATURES AND EXAMPLES IN MACHINE LEARNING. ARTIF INTELL 97: 245-271.
 100. SALCEDO-SANZ S, CAMPS-VALLS G, PREZ-CRUZ F, SEPLVEDA-SANCHIS J, BOUSOOCALZN C (2004) ENHANCING GENETIC FEATURE SELECTION THROUGH RESTRICTED SEARCH AND {W}ALSH ANALYSIS. {IEEE} TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, PART C 34: 398-406.
 101. RIVERO D, GUO L, SEOANE JA, DORADO J (2012) USING GENETIC ALGORITHMS AND K-NEAREST NEIGHBOUR FOR AUTOMATIC FREQUENCY BAND SELECTION FOR SIGNAL CLASSIFICATION. IET SIGNAL PROCESSING 6: 186.
 102. HOLLAND J (1992) ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS: MIT PRESS.
 103. KUDO M, SKLANSKY J (1998) A COMPARATIVE EVALUATION OF MEDIUM- AND LARGE-SCALE FEATURE SELECTORS FOR PATTERN CLASSIFIERS. KYBERNETIKA 34: 429-434.
 104. LI S, KWOK JT, ZHU H, WANG Y (2003) TEXTURE CLASSIFICATION USING THE SUPPORT VECTOR MACHINES. PATTERN RECOGNITION 36: 2883-2893.
 105. KIM KI, JUNG K, PARK SH, KIM HJ (2002) SUPPORT VECTOR MACHINES FOR TEXTURE CLASSIFICATION. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 24: 1542-1550.
 106. BUCIU I, KOTROPOULOS C, PITAS I (2006) DEMONSTRATING THE STABILITY OF SUPPORT VECTOR MACHINES FOR CLASSIFICATION. SIGNAL PROCESSING 86: 2364-2380.
 107. HUANG CL, WANG CJ (2006) A GA-BASED FEATURE SELECTION AND PARAMETERS OPTIMIZATIONFOR SUPPORT VECTOR MACHINES. EXPERT SYSTEMS WITH APPLICATIONS 31: 231-240.
 108. MANIMALA K, SELVI K, AHILA R (2011) HYBRID SOFT COMPUTING TECHNIQUES FOR FEATURE SELECTION AND PARAMETER OPTIMIZATION IN POWER QUALITY DATA MINING. APPLIED SOFT COMPUTING JOURNAL 11: 5485-5497.
 109. FROHLICH H, CHAPELLE O, SCHOLKOPF B. FEATURE SELECTION FOR SUPPORT VECTOR MACHINES BY MEANS OF GENETIC ALGORITHM; 2003 3-5 Nov. 2003. PP. 142-148.
 110. KEERTHI SS, LIN C-J (2003) ASYMPTOTIC BEHAVIORS OF SUPPORT VECTOR MACHINES WITH GAUSSIAN KERNEL. NEURAL COMPUTATION 15: 1667-1689.
 111. MIN S-H, LEE J, HAN I (2006) HYBRID GENETIC ALGORITHMS AND SUPPORT VECTOR MACHINES FOR BANKRUPTCY PREDICTION. EXPERT SYSTEMS WITH APPLICATIONS 31: 652-660.
 112. DONG SEONG K, HA-NAM N, JONG SOU P. GENETIC ALGORITHM TO IMPROVE SVM BASED NETWORK INTRUSION DETECTION SYSTEM; 2005 28-30 MARCH 2005. PP. 155-158 VOL.152.
 113. HUANG C-L, DUN J-F (2008) A DISTRIBUTED PSO-SVM HYBRID SYSTEM WITH FEATURE SELECTION AND PARAMETER OPTIMIZATION. APPLIED SOFT COMPUTING 8: 1381-1391.

114. CANTÚ-PAZ E, NEWSAM S, KAMATH C (2004) FEATURE SELECTION IN SCIENTIFIC APPLICATIONS. PROCEEDINGS OF THE TENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. SEATTLE, WA, USA: ACM. PP. 788-793.
115. GUYON I, WESTON J, BARNHILL S, VAPNIK V (2002) GENE SELECTION FOR CANCER CLASSIFICATION USING SUPPORT VECTOR MACHINES. MACH LEARN 46: 389-422.
116. RAKOTOMAMONJY A, BACH F, CANU S, GRANDVALET Y (2008) SIMPLEMKL. JOURNAL OF MACHINE LEARNING RESEARCH 9: 2491-2521.
117. WESTON J, ANDR, #233, ELISSEEFF, SCH B, ET AL. (2003) USE OF THE ZERO NORM WITH LINEAR MODELS AND KERNEL METHODS. J MACH LEARN RES 3: 1439-1461.
118. RAKOTOMAMONJY A (2003) VARIABLE SELECTION USING SVM BASED CRITERIA. J MACH LEARN RES 3: 1357-1370.
119. WESTON J, ELISSEEFF A, TIPPING M, SCHÖLKOPF B (2003) USE OF THE ZERO NORM WITH LINEAR MODELS AND KERNEL METHODS. J MACH LEARN RES 3: 1439-1461.
120. FURLANELLO C, SERAFINI M, MERLER S, JURMAN G (2003) BMC BIOINFORMATICS 4: 54.
121. SEOANE JA, DAY INM, GAUNT TR, CAMPBELL C (2013) A PATHWAY-BASED DATA INTEGRATION FRAMEWORK FOR PREDICTION OF DISEASE PROGRESSION. BIOINFORMATICS.
122. LANCKRIET GRG, CRISTIANINI N, JORDAN MI, NOBEL WS (2004) KERNEL-BASED INTEGRATION OF GENOMIC DATA USING SEMIDEFINITE PROGRAMMING. IN: VERT KTAJ-P, EDITOR. KERNEL METHODS IN COMPUTATIONAL BIOLOGY: MIT PRESS.
123. JANSEN R (2003) A BAYESIAN NETWORKS APPROACH FOR PREDICTING PROTEIN-PROTEIN INTERACTIONS FROM GENOMIC DATA. SCIENCE 302: 449-453.
124. GILCHRIST MA, SALTER LA, WAGNER A (2004) A STATISTICAL FRAMEWORK FOR COMBINING AND INTERPRETING PROTEOMIC DATASETS. BIOINFORMATICS 20: 689-700.
125. ZHANG LV, WONG SL, KING OD, ROTH FP (2004) PREDICTING CO-COMPLEXED PROTEIN PAIRS USING GENOMIC AND PROTEOMIC DATA INTEGRATION. BMC BIOINFORMATICS 5: 38.
126. GE H, LIU Z, CHURCH GM, VIDAL M (2001) CORRELATION BETWEEN TRANSCRIPTOME AND INTERACTOME MAPPING DATA FROM SACCHAROMYCES CEREVISIAE. NAT GENET 29: 482-486.
127. TANAY A, SHARAN R, SHAMIR R (2002) DISCOVERING STATISTICALLY SIGNIFICANT BICLUSTERS IN GENE EXPRESSION DATA. BIOINFORMATICS 18: S136-S144.
128. VON MERING C, KRAUSE R, SNEL B, CORNELL M, OLIVER SG, ET AL. (2002) COMPARATIVE ASSESSMENT OF LARGE-SCALE DATA SETS OF PROTEIN-PROTEIN INTERACTIONS. NATURE 417: 399-403.
129. DE BIE T, MONSIEURS P, ENGELEN K, DE MOOR B, CRISTIANINI N, ET AL. (2005) DISCOVERING TRANSCRIPTIONAL MODULES FROM MOTIF, CHIP-CHIP AND MICROARRAY DATA. PAC SYMP BIOCOMPUT: 483-494.
130. TSAKANIKAS P, MANOLAKOS ES (2009) IMPROVING 2-DE GEL IMAGE DENOISING USING CONTOURLETS. PROTEOMICS 9: 3877-3888.
131. EFRAT A, HOFFMANN F, KRIEGEL K, SCHULTZ C, WENK C (2002) GEOMETRIC ALGORITHMS FOR THE ANALYSIS OF 2D-ELECTROPHORESIS GELS. JOURNAL OF COMPUTATIONAL BIOLOGY 9: 299-315.

132. WÖRZ S, WINZ ML, ROHR K (2009) GEOMETRIC ALIGNMENT OF 2D GEL ELECTROPHORESIS IMAGES. *METHODS OF INFORMATION IN MEDICINE* 48: 320-323.
133. CANNISTRACI CV, MONTEVECCHI FM, ALESSIO M (2009) MEDIAN-MODIFIED WIENER FILTER PROVIDES EFFICIENT DENOISING, PRESERVING SPOT EDGE AND MORPHOLOGY IN 2-DE IMAGE PROCESSING. *PROTEOMICS* 9: 4908-4919.
134. TSAKANIKAS P, MANOLAKOS ES (2009) IMPROVING 2-DE GEL IMAGE DENOISING USING CONTOURLETS. *PROTEOMICS* 9: 3877-3888.
135. XIN H, ZHAO F. EFFECTIVE DENOISING METHODS FOR TWO-DIMENSIONAL GEL ELECTROPHORESIS IMAGES; 2011. pp. 1571-1574.
136. KACZMAREK K, WALCZAK B, DE JONG S, VANDEGINSTE BG (2004) PREPROCESSING OF TWO-DIMENSIONAL GEL ELECTROPHORESIS IMAGES. *PROTEOMICS* 4: 2377-2389.
137. DASZYKOWSKI M, STANIMIROVA I, BODZON-KULAKOWSKA A, SILBERRING J, LUBEC G, ET AL. (2007) START-TO-END PROCESSING OF TWO-DIMENSIONAL GEL ELECTROPHORETIC IMAGES. *JOURNAL OF CHROMATOGRAPHY A* 1158: 306-317.
138. MURINO V, OTTONELLO C, PAGNAN S (1998) NOISY TEXTURE CLASSIFICATION: A HIGHER-ORDER STATISTICS APPROACH. *PATTERN RECOGNITION* 31: 383-393.
139. RABILLOUD T, CHEVALLET M, LUCHE S, LELONG C (2010) TWO-DIMENSIONAL GEL ELECTROPHORESIS IN PROTEOMICS: PAST, PRESENT AND FUTURE. *JOURNAL OF PROTEOMICS* 73: 2064-2077.
140. LU P, VOGEL C, WANG R, YAO X, MARCOTTE EM (2007) ABSOLUTE PROTEIN EXPRESSION PROFILING ESTIMATES THE RELATIVE CONTRIBUTIONS OF TRANSCRIPTIONAL AND TRANSLATIONAL REGULATION. *NAT BIOTECH* 25: 117-124.
141. LEMKIN PF THE LECB 2D PAGE GEL IMAGE DATA SET CENTER FOR CANCER RESEARCH NANOBIOLOGY PROGRAM (CCRNP).
142. ZHANG J, TAN T (2002) BRIEF REVIEW OF INVARIANT TEXTURE ANALYSIS METHODS. *PATTERN RECOGNITION* 35: 735-747.
143. LEE TS, MUMFORD D, ROMERO R, LAMME VA (1998) THE ROLE OF THE PRIMARY VISUAL CORTEX IN HIGHER LEVEL VISION. *VISION RES* 38: 2429-2454.
144. PRATER JS, RICHARD WD (1992) SEGMENTING ULTRASOUND IMAGES OF THE PROSTATE USING NEURAL NETWORKS. *ULTRASONIC IMAGING* 14: 159-185.
145. RICHARD WD, KEEN CG (1996) AUTOMATED TEXTURE-BASED SEGMENTATION OF ULTRASOUND IMAGES OF THE PROSTATE. *COMPUT MED IMAGING GRAPH* 20: 131-140.
146. MIGNOTTE M, MEUNIER J, TARDIF J-C (2001) ENDOCARDIAL BOUNDARY ESTIMATION AND TRACKING IN ECHOCARDIOGRAPHIC IMAGES USING DEFORMABLE TEMPLATE AND MARKOV RANDOM FIELDS. *PATTERN ANALYSIS & APPLICATIONS* 4: 256-271.
147. PITIOT A, TOGA AW, THOMPSON PM (2002) ADAPTIVE ELASTIC SEGMENTATION OF BRAIN MRI VIA SHAPE-MODEL-GUIDED EVOLUTIONARY PROGRAMMING. *IEEE TRANS MED IMAGING* 21: 910-923.
148. KAVZOGLU T (2009) INCREASING THE ACCURACY OF NEURAL NETWORK CLASSIFICATION USING REFINED TRAINING DATA. *ENVIRONMENTAL MODELLING & SOFTWARE* 24: 850-858.
149. EDEN E, WAISMAN D, RUDZSKY M, BITTERMAN H, BROD V, ET AL. (2005) AN AUTOMATED METHOD FOR ANALYSIS OF FLOW CHARACTERISTICS OF CIRCULATING PARTICLES FROM IN VIVO VIDEO MICROSCOPY. *MEDICAL IMAGING, IEEE TRANSACTIONS ON* 24: 1011-1024.

150. ROJAS-DOMÍNGUEZ A, NANDI AK (2009) DEVELOPMENT OF TOLERANT FEATURES FOR CHARACTERIZATION OF MASSES IN MAMMOGRAMS. *COMPUTERS IN BIOLOGY AND MEDICINE* 39: 678-688.
151. BYUNG-WOO H, BONG-SOO S (2010) SEGMENTATION OF REGIONS OF INTEREST IN MAMMOGRAMS IN A TOPOGRAPHIC APPROACH. *INFORMATION TECHNOLOGY IN BIOMEDICINE, IEEE TRANSACTIONS ON* 14: 129-139.
152. MAHAPATRA D, SCHUEFFLER P, TIELBEEK JW, BUHMANN J, VOS F (2013) A SUPERVISED LEARNING APPROACH FOR CROHN'S DISEASE DETECTION USING HIGHER-ORDER IMAGE STATISTICS AND A NOVEL SHAPE ASYMMETRY MEASURE. *JOURNAL OF DIGITAL IMAGING*: 1-12.
153. MARKEL D, CALDWELL C, ALASTI H, SOLIMAN H, UNG Y, ET AL. (2013) AUTOMATIC SEGMENTATION OF LUNG CARCINOMA USING 3D TEXTURE FEATURES IN 18-FDG PET/CT. 13 p.
154. WEAVER JB, WANG J, FAN Z, VANDENBORNE K, WALTER G, ET AL. (2013) STATISTICAL TEXTURE ANALYSIS BASED MRI QUANTIFICATION OF DUCHENNE MUSCULAR DYSTROPHY IN A CANINE MODEL. 8672: 86720F-86720F-86728.
155. ABBEY CK, NOSRATIEH A, SOHL-DICKSTEIN J, YANG K, BOONE JM (2012) NON-GAUSSIAN STATISTICAL PROPERTIES OF BREAST IMAGES. *MEDICAL PHYSICS* 39: 7121.
156. HOLLI K, LÄÄPERI A-L, HARRISON L, LUUKKAALA T, TOIVONEN T, ET AL. (2010) CHARACTERIZATION OF BREAST CANCER TYPES BY TEXTURE ANALYSIS OF MAGNETIC RESONANCE IMAGES. *ACADEMIC RADIOLOGY* 17: 135-141.
157. GANESHAN B, SKOGEN K, PRESSNEY I, COUTROUBIS D, MILES K (2012) TUMOUR HETEROGENEITY IN OESOPHAGEAL CANCER ASSESSED BY CT TEXTURE ANALYSIS: PRELIMINARY EVIDENCE OF AN ASSOCIATION WITH TUMOUR METABOLISM, STAGE, AND SURVIVAL. *CLINICAL RADIOLOGY* 67: 157-164.
158. GANESHAN B, GOH V, MANDEVILLE HC, NG QS, HOSKIN PJ, ET AL. (2012) NON-SMALL CELL LUNG CANCER: HISTOPATHOLOGIC CORRELATES FOR TEXTURE PARAMETERS AT CT. *RADIOLOGY* 266: 326-336.
159. GANESHAN B, BURNAND K, YOUNG R, CHATWIN C, MILES K (2011) DYNAMIC CONTRAST-ENHANCED TEXTURE ANALYSIS OF THE LIVER. *INVESTIGATIVE RADIOLOGY* 46: 160-168.
160. AL-KADI OS (2010) ASSESSMENT OF TEXTURE MEASURES SUSCEPTIBILITY TO NOISE IN CONVENTIONAL AND CONTRAST ENHANCED COMPUTED TOMOGRAPHY LUNG TUMOUR IMAGES. *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS* 34: 494-503.
161. GOH V, GANESHAN B, NATHAN P, JUTTILA JK, VINAYAN A, ET AL. (2011) ASSESSMENT OF RESPONSE TO TYROSINE KINASE INHIBITORS IN METASTATIC RENAL CELL CANCER: CT TEXTURE AS A PREDICTIVE BIOMARKER. *RADIOLOGY* 261: 165-171.
162. ELIAT P-A, OLIVIÉ D, SAÏKALI S, CARSIN B, SAINT-JALMES H, ET AL. (2012) CAN DYNAMIC CONTRAST-ENHANCED MAGNETIC RESONANCE IMAGING COMBINED WITH TEXTURE ANALYSIS DIFFERENTIATE MALIGNANT GLIONEURONAL TUMORS FROM OTHER GLIOBLASTOMA? *NEUROLOGY RESEARCH INTERNATIONAL* 2012: 1-7.
163. HAYKIN S (1998) *NEURAL NETWORKS: A COMPREHENSIVE FOUNDATION*: PRENTICE HALL PTR. 842 p.
164. BISHOP CM (1995) *NEURAL NETWORKS FOR PATTERN RECOGNITION*: OXFORD UNIVERSITY PRESS, INC. 482 p.
165. WOODS BJ, CLYMER BD, KURC T, HEVERHAGEN JT, STEVENS R, ET AL. (2007) MALIGNANT-LESION SEGMENTATION USING 4D CO-OCCURRENCE TEXTURE

- ANALYSIS APPLIED TO DYNAMIC CONTRAST-ENHANCED MAGNETIC RESONANCE BREAST IMAGE DATA. *JOURNAL OF MAGNETIC RESONANCE IMAGING* 25: 495-501.
166. CABAN JJ, JIANHUA Y, BAGCI U, MOLLURA DJ (2011) MONITORING PULMONARY FIBROSIS BY FUSING CLINICAL, PHYSIOLOGICAL, AND COMPUTED TOMOGRAPHY FEATURES. 6216-6219.
 167. BAGCI U, BRAY M, CABAN J, YAO J, MOLLURA DJ (2012) COMPUTER-ASSISTED DETECTION OF INFECTIOUS LUNG DISEASES: A REVIEW. *COMPUT MED IMAGING GRAPH* 36: 72-84.
 168. ZACHARAKI EI, WANG S, CHAWLA S, SOO YOO D, WOLF R, ET AL. (2009) CLASSIFICATION OF BRAIN TUMOR TYPE AND GRADE USING MRI TEXTURE AND SHAPE IN A MACHINE LEARNING SCHEME. *MAGN RESON MED* 62: 1609-1618.
 169. LOPES R, AYACHE A, MAKNI N, PUECH P, VILLERS A, ET AL. (2011) PROSTATE CANCER CHARACTERIZATION ON MR IMAGES USING FRACTAL FEATURES. *MED PHYS* 38: 83-95.
 170. REYES-ALDASORO C, BHALERAO A (2003) VOLUMETRIC TEXTURE DESCRIPTION AND DISCRIMINANT FEATURE SELECTION FOR MRI. IN: TAYLOR C, NOBLE JA, EDITORS. *INFORMATION PROCESSING IN MEDICAL IMAGING: SPRINGER BERLIN HEIDELBERG*. PP. 282-293.
 171. SAHINER B (1996) IMAGE FEATURE SELECTION BY A GENETIC ALGORITHM: APPLICATION TO CLASSIFICATION OF MASS AND NORMAL BREAST TISSUE. *MEDICAL PHYSICS* 23: 1671.
 172. ZHENG B, CHANG Y-H, WANG X-H, GOOD WF, GUR D (1999) FEATURE SELECTION FOR COMPUTERIZED MASS DETECTION IN DIGITIZED MAMMOGRAMS BY USING A GENETIC ALGORITHM. *ACADEMIC RADIOLOGY* 6: 327-332.
 173. VERMA B, ZHANG P (2007) A NOVEL NEURAL-GENETIC ALGORITHM TO FIND THE MOST SIGNIFICANT COMBINATION OF FEATURES IN DIGITAL MAMMOGRAMS. *APPLIED SOFT COMPUTING* 7: 612-625.
 174. PUSHPA RATHI VPG (2012) BRAIN TUMOR MRI IMAGE CLASSIFICATION WITH FEATURE SELECTION AND EXTRACTION USING LINEAR DISCRIMINANT ANALYSIS. *INTERNATIONAL JOURNAL OF INFORMATION SCIENCES AND TECHNIQUES* 2: 131-146.
 175. CHANG C-Y, LAI C-C, LAI C-T, CHEN S-J (2013) INTEGRATING PSONN AND BOLTZMANN FUNCTION FOR FEATURE SELECTION AND CLASSIFICATION OF LYMPH NODES IN ULTRASOUND IMAGES. *JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION* 24: 23-30.
 176. LOIZOU CP, PANTZIARIS M, PATTICHIS CS, SEIMENIS I (2012) BRAIN MR IMAGE NORMALIZATION IN TEXTURE ANALYSIS OF MULTIPLE SCLEROSIS. *JOURNAL OF BIOMEDICAL GRAPHICS AND COMPUTING* 3.
 177. LI F, ZHAO C, XIA Z, WANG Y, ZHOU X, ET AL. (2012) COMPUTER-ASSISTED LIP DIAGNOSIS ON TRADITIONAL CHINESE MEDICINE USING MULTI-CLASS SUPPORT VECTOR MACHINES. *BMC COMPLEMENTARY AND ALTERNATIVE MEDICINE* 12: 127.
 178. THEODORIDIS S, KOUTROUMBAS K (2009) *PATTERN RECOGNITION: ELSEVIER SCIENCE*.
 179. BACHE K, LICHMAN M (2013) UCI MACHINE LEARNING REPOSITORY. IN: UNIVERSITY OF CALIFORNIA SOIACS, EDITOR. IRVINE, CA.
 180. SZCZYPISKI PM, STRZELECKI M, MATERKA A, KLEPACZKO A (2009) MAZDA-A SOFTWARE PACKAGE FOR IMAGE TEXTURE ANALYSIS. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE* 94: 66-76.

181. SZCZYPINSKI PM, STRZELECKI M, MATERKA A. MAZDA - A SOFTWARE FOR TEXTURE ANALYSIS; 2007. PP. 245-249.
182. TECHNICAL UNIVERSITY OF LODZ IOE MAZDA WEB SITE.
183. BONILHA L, KOBAYASHI E, CASTELLANO G, COELHO G, TINOIS E, ET AL. (2003) TEXTURE ANALYSIS OF HIPPOCAMPAL SCLEROSIS. *EPILEPSIA* 44: 1546-1550.
184. MAYERHOEFER ME, BREITENSEHER MJ, KRAMER J, AIGNER N, HOFMANN S, ET AL. (2005) TEXTURE ANALYSIS FOR TISSUE DISCRIMINATION ON T1-WEIGHTED MR IMAGES OF THE KNEE JOINT IN A MULTICENTER STUDY: TRANSFERABILITY OF TEXTURE FEATURES AND COMPARISON OF FEATURE SELECTION METHODS AND CLASSIFIERS. *JOURNAL OF MAGNETIC RESONANCE IMAGING* 22: 674-680.
185. SZYMANSKI JJ, JAMISON JT, DEGRACIA DJ (2012) TEXTURE ANALYSIS OF POLY-ADENYLATED mRNA STAINING FOLLOWING GLOBAL BRAIN ISCHEMIA AND REPERFUSION. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE* 105: 81-94.
186. MARCHAND-LIBOUBAN H, GUILLAUME B, BELLAICHE N, CHAPPARD D (2013) TEXTURE ANALYSIS OF COMPUTED TOMOGRAPHIC IMAGES IN OSTEOPOROTIC PATIENTS WITH SINUS LIFT BONE GRAFT RECONSTRUCTION. *CLIN ORAL INVESTIG* 17: 1267-1272.
187. SIKIÖ M, HARRISON LCV, HOLLI KK, RUOTTINEN H, ELOVAARA I, ET AL. (2012) TEXTURE ANALYSIS AS A TOOL IN DIAGNOSIS OF PARKINSON'S DISEASE. IN: JOBBÁGY Á, EDITOR. 5TH EUROPEAN CONFERENCE OF THE INTERNATIONAL FEDERATION FOR MEDICAL AND BIOLOGICAL ENGINEERING: SPRINGER BERLIN HEIDELBERG. PP. 571-574.
188. FERNANDEZ-LOZANO C, SEOANE JA, MESEJO P, NASHED YSG, CAGNONI S, ET AL. 2D-PAGE TEXTURE CLASSIFICATION USING SUPPORT VECTOR MACHINES AND GENETIC ALGORITHMS; 2013. SCITEPRESS. PP. 5-14.
189. DASU T, JOHNSON T (2003) EXPLORATORY DATA MINING AND DATA CLEANING: JOHN WILEY & SONS, INC. 203 P.
190. GORMLEY M, SUN T, ZHANG R, WANG J, LI X, ET AL. (2013) COMPUTER-AIDED DIAGNOSIS FOR EARLY-STAGE LUNG CANCER BASED ON LONGITUDINAL AND BALANCED DATA. *PLOS ONE* 8: e63559.
191. FERNÁNDEZ-NAVARRO F, HERVÁS-MARTÍNEZ C, ANTONIO GUTIÉRREZ P (2011) A DYNAMIC OVER-SAMPLING PROCEDURE BASED ON SENSITIVITY FOR MULTI-CLASS PROBLEMS. *PATTERN RECOGNITION* 44: 1821-1833.
192. WEISS G, PROVOST F (2001) THE EFFECT OF CLASS DISTRIBUTION ON CLASSIFIER LEARNING: AN EMPIRICAL STUDY.
193. LAURIKKALA J (2001) IMPROVING IDENTIFICATION OF DIFFICULT SMALL CLASSES BY BALANCING CLASS DISTRIBUTION. PROCEEDINGS OF THE 8TH CONFERENCE ON AI IN MEDICINE IN EUROPE: ARTIFICIAL INTELLIGENCE MEDICINE: SPRINGER-VERLAG. PP. 63-66.
194. ESTABROOKS A, JO T, JAPKOWICZ N (2004) A MULTIPLE RESAMPLING METHOD FOR LEARNING FROM IMBALANCED DATA SETS. *COMPUTATIONAL INTELLIGENCE* 20: 18-36.
195. CHAWLA NV, BOWYER KW, HALL LO, KEGELMEYER WP (2002) SMOTE: SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE. *J ARTIF INT RES* 16: 321-357.
196. BELLMAN R (1961) ADAPTIVE CONTROL PROCESSES - A GUIDED TOUR: PRINCETON UNIVERSITY PRESS. 255 P.

197. DONOHO DL (2000) HIGH-DIMENSIONAL DATA ANALYSIS: THE CURSES AND BLESSINGS OF DIMENSIONALITY. AMERICAN MATHEMATICAL SOCIETY CONF MATH CHALLENGES OF THE 21ST CENTURY.
198. FERRI C, HERNÁNDEZ-ORALLO J, MODROIU R (2009) AN EXPERIMENTAL COMPARISON OF PERFORMANCE MEASURES FOR CLASSIFICATION. PATTERN RECOGNITION LETTERS 30: 27-38.
199. HANCZAR B, HUA J, SIMA C, WEINSTEIN J, BITTNER M, ET AL. (2010) SMALL-SAMPLE PRECISION OF ROC-RELATED ESTIMATES. BIOINFORMATICS 26: 822-830.
200. HAND DJ (2009) MEASURING CLASSIFIER PERFORMANCE: A COHERENT ALTERNATIVE TO THE AREA UNDER THE ROC CURVE. MACHINE LEARNING 77: 103-123.
201. LOBO JM, JIMÉNEZ-VALVERDE A, REAL R (2008) AUC: A MISLEADING MEASURE OF THE PERFORMANCE OF PREDICTIVE DISTRIBUTION MODELS. GLOBAL ECOLOGY AND BIOGEOGRAPHY 17: 145-151.
202. SPACKMAN KA (1989) SIGNAL DETECTION THEORY: VALUABLE TOOLS FOR EVALUATING INDUCTIVE LEARNING. PROCEEDINGS OF THE SIXTH INTERNATIONAL WORKSHOP ON MACHINE LEARNING. ITHACA, NEW YORK, USA: MORGAN KAUFMANN PUBLISHERS INC. PP. 160-163.
203. PROVOST F, FAWCETT T (2001) ROBUST CLASSIFICATION FOR IMPRECISE ENVIRONMENTS. MACH LEARN 42: 203-231.
204. PROVOST FJ, FAWCETT T, KOHAVI R (1998) THE CASE AGAINST ACCURACY ESTIMATION FOR COMPARING INDUCTION ALGORITHMS. PROCEEDINGS OF THE FIFTEENTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING: MORGAN KAUFMANN PUBLISHERS INC. PP. 445-453.
205. HUANG J, LING CX (2005) USING AUC AND ACCURACY IN EVALUATING LEARNING ALGORITHMS. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING 17: 299-310.
206. FAWCETT T (2006) AN INTRODUCTION TO ROC ANALYSIS. PATTERN RECOGNITION LETTERS 27: 861-874.
207. GARCÍA S, FERNÁNDEZ A, LUENGO J, HERRERA F (2010) ADVANCED NONPARAMETRIC TESTS FOR MULTIPLE COMPARISONS IN THE DESIGN OF EXPERIMENTS IN COMPUTATIONAL INTELLIGENCE AND DATA MINING: EXPERIMENTAL ANALYSIS OF POWER. INFORMATION SCIENCES 180: 2044-2064.
208. SHESKIN DJ (2011) HANDBOOK OF PARAMETRIC AND NONPARAMETRIC STATISTICAL PROCEDURES: TAYLOR AND FRANCIS.
209. DEMSAR J (2006) STATISTICAL COMPARISONS OF CLASSIFIERS OVER MULTIPLE DATA SETS. JOURNAL OF MACHINE LEARNING RESEARCH 7: 1-30.
210. GARCÍA S, HERRERA F (2008) AN EXTENSION ON "STATISTICAL COMPARISONS OF CLASSIFIERS OVER MULTIPLE DATA SETS" FOR ALL PAIRWISE COMPARISONS. JOURNAL OF MACHINE LEARNING RESEARCH 9: 2677-2694.
211. ZAR JH (2010) BIostatistical ANALYSIS: PRENTICE-HALL/PEARSON.
212. GARCÍA S, MOLINA D, LOZANO M, HERRERA F (2008) A STUDY ON THE USE OF NON-PARAMETRIC TESTS FOR ANALYZING THE EVOLUTIONARY ALGORITHMS' BEHAVIOUR: A CASE STUDY ON THE CEC'2005 SPECIAL SESSION ON REAL PARAMETER OPTIMIZATION. JOURNAL OF HEURISTICS 15: 617-644.
213. SHAPIRO SS, WILK MB (1965) AN ANALYSIS OF VARIANCE TEST FOR NORMALITY (COMPLETE SAMPLES). BIOMETRIKA 52: 591-611.

214. GARCÍA S, FERNÁNDEZ A, LUENGO J, HERRERA F (2009) A STUDY OF STATISTICAL TECHNIQUES AND PERFORMANCE MEASURES FOR GENETICS-BASED MACHINE LEARNING: ACCURACY AND INTERPRETABILITY. *SOFT COMPUTING* 13: 959-977.
215. BARTLETT MS (1937) PROPERTIES OF SUFFICIENCY AND STATISTICAL TESTS. *PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A, MATHEMATICAL AND PHYSICAL SCIENCES* 160: 268-282.
216. FRIEDMAN M (1937) THE USE OF RANKS TO AVOID THE ASSUMPTION OF NORMALITY IMPLICIT IN THE ANALYSIS OF VARIANCE. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 32: 675-701.
217. TUKEY JW (1949) COMPARING INDIVIDUAL MEANS IN THE ANALYSIS OF VARIANCE. *BIOMETRICS* 5: 99-114.
218. DUNNETT CW (1955) A MULTIPLE COMPARISON PROCEDURE FOR COMPARING SEVERAL TREATMENTS WITH A CONTROL. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 50: 1096-1121.
219. HSU JC (1996) *MULTIPLE COMPARISONS: THEORY AND METHODS*: CHAPMAN AND HALL/CRC.
220. DUNN OJ (1961) MULTIPLE COMPARISONS AMONG MEANS. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 56: 52-64.
221. HOLM S (1979) A SIMPLE SEQUENTIALLY REJECTIVE MULTIPLE TEST PROCEDURE. *SCANDINAVIAN JOURNAL OF STATISTICS* 6: 65-70.
222. HOCHBERG Y (1988) A SHARPER BONFERRONI PROCEDURE FOR MULTIPLE TESTS OF SIGNIFICANCE. *BIOMETRIKA* 75: 800-802.
223. HOMMEL G (1988) A STAGEWISE REJECTIVE MULTIPLE TEST PROCEDURE BASED ON A MODIFIED BONFERRONI TEST. *BIOMETRIKA* 75: 383-386.
224. DE LAPLACE PS (1814) *THÉORIE ANALYTIQUE DES PROBABILITÉS*: VE. COURCIER.
225. QUADE D (1979) USING WEIGHTED RANKINGS IN THE ANALYSIS OF COMPLETE BLOCKS WITH ADDITIVE BLOCK EFFECTS. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 74: 680-683.
226. SMITS GF, JORDAAN EM (2002) IMPROVED SVM REGRESSION USING MIXTURES OF KERNELS. *INTERNATIONAL SYMPOSIUM ON NEURAL NETWORKS*. PP. 2785-2790.
227. BENGIO Y, DELALLEAU O, ROUX NL (2005) THE CURSE OF HIGHLY VARIABLE FUNCTIONS FOR LOCAL KERNEL MACHINES. *NEURAL INFORMATION PROCESSING SYSTEMS*.
228. STEINWART I (2002) ON THE INFLUENCE OF THE KERNEL ON THE CONSISTENCY OF SUPPORT VECTOR MACHINES. *J MACH LEARN RES* 2: 67-93.
229. LIU H, MOTODA H (1998) *FEATURE SELECTION FOR KNOWLEDGE DISCOVERY AND DATA MINING*: KLUWER ACADEMIC PUBLISHERS. 214 P.
230. MARDIA KV. MAHALANOBIS DISTANCES AND ANGLES. IN: KRISHNAIAH PR, EDITOR. *MULTIVARATE ANALYSIS*; 1977; WRIGHT STATE UNIVERSITY. DAYTON. OHIO. NORTH-HOLLAND. PP. 495-511.
231. SWAIN PH, KING RC. TWO EFFECTIVE FEATURE SELECTION CRITERIA FOR MULTISPECTRAL REMOTE SENSING; 1973 OCTOBER 30-NOVEMBER 1; WASHINGTON, D. C. INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. PP. 536-540.
232. RICHARDS JA, JIA X (2006) *REMOTE SENSING DIGITAL IMAGE ANALYSIS: AN INTRODUCTION*: SPRINGER.
233. DEVELOPMENT CORE T (2005) *R: A LANGUAGE AND ENVIRONMENT FOR STATISTICAL COMPUTING*. R FOUNDATION FOR STATISTICAL COMPUTING, VIENNA, AUSTRIA.

234. KUHN M (2008) BUILDING PREDICTIVE MODELS IN R USING THE CARET PACKAGE. JOURNAL OF STATISTICAL SOFTWARE 28: 1-26.
235. ROBIN X, TURCK N, HAINARD A, TIBERTI N, LISACEK F, ET AL. (2011) PROC: AN OPEN-SOURCE PACKAGE FOR R AND S+ TO ANALYZE AND COMPARE ROC CURVES. BMC BIOINFORMATICS 12: 77.
236. MARTEN R MARTEN LAB PROTEOMICS PAGE.
237. MILLIONI R, SBRIGNADELLO S, TURA A, IORI E, MURPHY E, ET AL. (2010) THE INTER- AND INTRA-OPERATOR VARIABILITY IN MANUAL SPOT SEGMENTATION AND ITS EFFECT ON SPOT QUANTITATION IN TWO-DIMENSIONAL ELECTROPHORESIS ANALYSIS. ELECTROPHORESIS 31: 1739-1742.
238. HUNT SMN, THOMAS MR, SEBASTIAN LT, PEDERSEN SK, HARCOURT RL, ET AL. (2005) OPTIMAL REPLICATION AND THE IMPORTANCE OF EXPERIMENTAL DESIGN FOR GEL-BASED QUANTITATIVE PROTEOMICS. JOURNAL OF PROTEOME RESEARCH 4: 809-819.
239. DOWSEY AW, ENGLISH JA, LISACEK F, MORRIS JS, YANG G, ET AL. (2010) IMAGE ANALYSIS TOOLS AND EMERGING ALGORITHMS FOR EXPRESSION PROTEOMICS. PROTEOMICS 10: 4226-4257.
240. DOWSEY AW, DUNN MJ, YANG GZ (2008) AUTOMATED IMAGE ALIGNMENT FOR 2D GEL ELECTROPHORESIS IN A HIGH-THROUGHPUT PROTEOMICS PIPELINE. BIOINFORMATICS 24: 950-957.
241. ANDERSON NL, TAYLOR J, SCANDORA AE, COULTER BP, ANDERSON NG (1981) THE TYCHO SYSTEM FOR COMPUTER ANALYSIS OF TWO-DIMENSIONAL GEL ELECTROPHORESIS PATTERNS. CLIN CHEM 27: 1807-1820.
242. APPEL RD, VARGAS JR, PALAGI PM, WALTHER D, HOCHSTRASSER DF (1997) MELANIE II--A THIRD-GENERATION SOFTWARE PACKAGE FOR ANALYSIS OF TWO-DIMENSIONAL ELECTROPHORESIS IMAGES: II. ALGORITHMS. ELECTROPHORESIS 18: 2735-2748.
243. BETTENS E, SCHEUNDERS P, VAN DYCK D, MOENS L, VAN OSTA P (1997) COMPUTER ANALYSIS OF TWO-DIMENSIONAL ELECTROPHORESIS GELS: A NEW SEGMENTATION AND MODELING ALGORITHM. ELECTROPHORESIS 18: 792-798.
244. CUTLER P, HEALD G, WHITE IR, RUAN J (2003) A NOVEL APPROACH TO SPOT DETECTION FOR TWO-DIMENSIONAL GEL ELECTROPHORESIS IMAGES USING PIXEL VALUE COLLECTION. PROTEOMICS 3: 392-401.
245. MATERKA A, STRZELECKI M, SZCZYPINSKI P (2013) MANUAL DE USO DE PROGRAMA MAZDA.
246. FAUVEL M, CHANUSSOT J, BENEDIKTSSON JA, VILLA A (2013) PARSIMONIOUS MAHALANOBIS KERNEL FOR THE CLASSIFICATION OF HIGH DIMENSIONAL DATA. PATTERN RECOGNITION 46: 845-854.
247. CHAPPELLE O, VAPNIK V (2000) MODEL SELECTION FOR SUPPORT VECTOR MACHINES. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 12: PROCEEDINGS OF THE 1999 CONFERENCE: MIT PRESS. PP. 230-236.
248. AGGARWAL C, HINNEBURG A, KEIM D (2001) ON THE SURPRISING BEHAVIOR OF DISTANCE METRICS IN HIGH DIMENSIONAL SPACE. IN: BUSSCHE J, VIANU V, EDITORS. DATABASE THEORY — ICDT 2001: SPRINGER BERLIN HEIDELBERG. PP. 420-434.
249. VERLEYSEN M, FRANCOIS D, SIMON G, WERTZ V (2003) ON THE EFFECTS OF DIMENSIONALITY ON DATA ANALYSIS WITH NEURAL NETWORKS. PROCEEDINGS OF

- THE 7TH INTERNATIONAL WORK-CONFERENCE ON ARTIFICIAL AND NATURAL NEURAL NETWORKS: PART II: ARTIFICIAL NEURAL NETS PROBLEM SOLVING METHODS. MA \& \#243;, MENORCA, SPAIN: SPRINGER-VERLAG. PP. 105-112.
250. VERLEYSEN M (2003) LEARNING HIGH-DIMENSIONAL DATA. IN: (EDS) SAEA, EDITOR. LIMITATIONS AND FUTURE TRENDS IN NEURAL COMPUTATION: IOS PRESS. PP. 141-162.
 251. BOUVEYRON C, GIRARD S, SCHMID C (2007) HIGH-DIMENSIONAL DATA CLUSTERING. COMPUTATIONAL STATISTICS & DATA ANALYSIS 52: 502-519.
 252. BOUVEYRON C, GIRARD S, SCHMID C (2007) HIGH-DIMENSIONAL DISCRIMINANT ANALYSIS. COMMUNICATIONS IN STATISTICS - THEORY AND METHODS 36: 2607-2623.
 253. HAFLER DA (2012) PERSPECTIVE: DECONSTRUCTING A DISEASE. NATURE 484: S6.
 254. SCHRIML LM, ARZE C, NADENDLA S, CHANG YWW, MAZAITIS M, ET AL. (2011) DISEASE ONTOLOGY: A BACKBONE FOR DISEASE SEMANTIC INTEGRATION. NUCLEIC ACIDS RESEARCH 40: D940-D946.
 255. NELSON SJ, SCHOPEN M, SAVAGE AG, SCHULMAN JL, ARLUK N (2004) THE MESH TRANSLATION MAINTENANCE SYSTEM: STRUCTURE, INTERFACE DESIGN, AND IMPLEMENTATION. STUD HEALTH TECHNOL INFORM 107: 67-69.
 256. AYMÉ S, RATH A, BELLET B (2010) WHO INTERNATIONAL CLASSIFICATION OF DISEASES (ICD) REVISION PROCESS: INCORPORATING RARE DISEASES INTO THE CLASSIFICATION SCHEME: STATE OF ART. ORPHANET JOURNAL OF RARE DISEASES 5: P1.
 257. CORNET R, DE KEIZER N (2008) FORTY YEARS OF SNOMED: A LITERATURE REVIEW. BMC MEDICAL INFORMATICS AND DECISION MAKING 8: S2.
 258. SIOTOS N, CORONADO SD, HABER MW, HARTEL FW, SHAIU W-L, ET AL. (2007) NCI THESAURUS: A SEMANTIC MODEL INTEGRATING CANCER-RELATED CLINICAL AND MOLECULAR INFORMATION. J OF BIOMEDICAL INFORMATICS 40: 30-43.
 259. AMBERGER J, BOCCHINI C, HAMOSH A (2011) A NEW FACE AND NEW CHALLENGES FOR ONLINE MENDELIAN INHERITANCE IN MAN (OMIM®). HUMAN MUTATION 32: 564-567.
 260. LOSCALZO J, KOHANE I, BARABASI AL (2007) HUMAN DISEASE CLASSIFICATION IN THE POSTGENOMIC ERA: A COMPLEX SYSTEMS APPROACH TO HUMAN PATHOBIOLOGY. MOL SYST BIOL 3: 124.
 261. MOSTAFAVI S, MORRIS Q (2012) COMBINING MANY INTERACTION NETWORKS TO PREDICT GENE FUNCTION AND ANALYZE GENE LISTS. PROTEOMICS 12: 1687-1696.
 262. PANDEY G, ZHANG B, CHANG AN, MYERS CL, ZHU J, ET AL. (2010) AN INTEGRATIVE MULTI-NETWORK AND MULTI-CLASSIFIER APPROACH TO PREDICT GENETIC INTERACTIONS. PLOS COMPUT BIOL 6.
 263. VAN VLIET MH, HORLINGS HM, VAN DE VIJVER MJ, REINDERS MJT, WESSELS LFA (2012) INTEGRATION OF CLINICAL AND GENE EXPRESSION DATA HAS A SYNERGETIC EFFECT ON PREDICTING BREAST CANCER OUTCOME. PLOS ONE 7: E40358.
 264. AERTS S, LAMBRECHTS D, MAITY S, VAN LOO P, COESSENS B, ET AL. (2006) GENE PRIORITIZATION THROUGH GENOMIC DATA FUSION. NAT BIOTECHNOL 24: 537-544.
 265. CHEN Z, ZHANG W (2013) INTEGRATIVE ANALYSIS USING MODULE-GUIDED RANDOM FORESTS REVEALS CORRELATED GENETIC FACTORS RELATED TO MOUSE WEIGHT. PLOS COMPUT BIOL 9: E1002956.

266. BRADLEY AP (1997) THE USE OF THE AREA UNDER THE ROC CURVE IN THE EVALUATION OF MACHINE LEARNING ALGORITHMS. PATTERN RECOGNITION 30: 1145-1159.
267. HANLEY JA, MCNEIL BJ (1982) THE MEANING AND USE OF THE AREA UNDER A RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE. RADIOLOGY 143: 29-36.
268. GARCÍA S, FERNÁNDEZ A, HERRERA F (2009) ENHANCING THE EFFECTIVENESS AND INTERPRETABILITY OF DECISION TREE AND RULE INDUCTION CLASSIFIERS WITH EVOLUTIONARY TRAINING SET SELECTION OVER IMBALANCED PROBLEMS. APPLIED SOFT COMPUTING 9: 1304-1314.

ANEXOS

ANEXO A: VARIABLES DE TEXTURA BASADAS EN LA MATRIZ DE COOCURRENCIA

Sean (x_1, y_1) y (x_2, y_2) un par de píxeles en la imagen separados por una distancia d con un ángulo α . Si los niveles de gris de cada par de píxeles son i y j respectivamente, se puede definir la probabilidad de que los niveles de gris i y j estén presentes en la imagen a una distancia d y con un ángulo α como $P(i, j, d, \alpha)$.

- Momento angular segundo: es una medida de homogeneidad en la imagen. Cuanto más homogénea es la imagen, menor número de niveles de gris contiene aunque de valores altos.

$$MAS = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\}^2$$

- Contraste: esta medida favorece contribuciones de $P(i, j)$ fuera de la diagonal $i \neq j$.

$$CON = \sum_{i=0}^{G-1} n^2 \left\{ \sum_{i=1}^G \sum_{j=1}^G P(i, j) \right\}, |i - j| = n$$

- Correlación: es una medida de la dependencia lineal de los niveles de gris entre píxeles en una determinada posición con respecto a otra.

$$COR = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i \times j\} \times P(i, j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$$

- Suma de cuadrados: también conocida como varianza, esta variable destaca valores altos que difieren de la media de $P(i, j)$.

$$VAR = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 P(i, j)$$

- Momento diferencial inverso: valor influenciado por la homogeneidad de la imagen de tal manera que valores bajos de MDI indican que es una imagen poco homogénea y al revés.

$$MDI = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i-j)^2} P(i, j)$$

- Suma promedio

$$SUM = \sum_{i=0}^{2G-2} iP_{x+y}(i)$$

- Suma varianza

$$SVAR = \sum_{i=0}^{2G-2} (i - SENT)^2 P_{x+y}(i)$$

- Entropía: las imágenes poco homogéneas tienen un valor bajo de entropía.

$$ENT = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \times \log(P(i, j))$$

- Suma de entropías

$$SENT = - \sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i))$$

- Diferencia de entropías

$$DENT = - \sum_{i=0}^{G-1} P_{x-y}(i) \log(P_{x-y}(i))$$

ANEXO B: ANÁLISIS CON CURVAS ROC

A pesar de que las curvas ROC (*Receiver Operating Characteristics*) son conceptualmente sencillas y fáciles de entender, existen aspectos que no son tan obvios acerca de cómo se construyen que es necesario clarificar para un mejor entendimiento. En este anexo se introducirá inicialmente la matriz de confusión y medidas asociadas a ella para problemas de clasificación binaria. Posteriormente se mostrará como dibujar una curva ROC y se explicará el significado del valor del área bajo la curva AUROC. Se presentan en este anexo los conceptos básicos y definiciones suficientes para comprender cómo se obtienen las curvas ROC, para una explicación más exhaustiva acerca de cómo realizar un análisis ROC se recomienda consultar [206].

Un modelo de clasificación para problemas binarios identifica cada ejemplo con su clase de pertenencia. Con alguno de estos modelos de clasificación se obtienen salidas continuas sobre las que se pueden aplicar diferentes umbrales para predecir la clase de pertenencia del modelo. Por tanto, existen cuatro posibilidades cuando un clasificador responde acerca de la clase de pertenencia de un ejemplo, considerando que éste sea binario y que distinga entre casos positivos y negativos; si el ejemplo es positivo y se clasifica como positivo, se dice que es un verdadero positivo (VP), si es clasificado como negativo, se dice que es un falso negativo (FN). Si el ejemplo es negativo y se clasifica como negativo, se dice que es un verdadero negativo (VN), si es clasificado como positivo, se dice que es un falso positivo (FP).

A partir de estas observaciones puede construirse la matriz de confusión que, al ser un problema de clasificación binaria, tendrá el aspecto mostrado en la Tabla 38.

		Valor real	
		Clase positiva	Clase negativa
Resultado del clasificador	Clase positiva	VP	FP
	Clase negativa	FN	VN

TABLA 38. MATRIZ DE CONFUSIÓN PARA UN PROBLEMA DE CLASIFICACIÓN BINARIO

En el dominio de un trabajo de investigación, el error de tipo I o FP se comete cuando no se acepta la hipótesis nula siendo verdadera y el error de tipo II o FN, se comete cuando no se rechaza la hipótesis nula siendo ésta falsa. Se presentaron los test de contraste de hipótesis nula en el capítulo 4.1.4 de esta Tesis Doctoral.

A partir de los valores calculados con la matriz de confusión pueden obtenerse diferentes medidas de rendimiento que se expondrán a continuación:

- Precisión: es la tasa de acierto global del sistema

$$Precisión = \frac{VP + VN}{VP + VN + FP + FN}$$

- Sensibilidad: es la probabilidad de que se obtenga un resultado positivo para un caso positivo. También conocida como *recall*.

$$Sensibilidad = \frac{VP}{VP + FN}$$

- Especificidad: es la probabilidad de que se obtenga un resultado negativo para un caso negativo.

$$Especificidad = \frac{VN}{VN + VP}$$

- Valor predictivo positivo: es la proporción de casos positivos que han sido correctamente clasificados.

$$VPP = \frac{VP}{VP + FP}$$

- Valor predictivo negativo: es la proporción de casos negativos que han sido correctamente clasificados.

$$VPN = \frac{VN}{VN + FN}$$

- Tasa de falsos positivos: se calcula como $TFP = 1 - especificidad$.
- Tasa de falsos negativos: se calcula como $TFN = 1 - sensibilidad$.
- Tasa de verdaderos positivos: se calcula como

$$TVP = \frac{\text{Casos positivos correctamente clasificados}}{\text{Total de casos positivos}}$$

- Tasa de falsos positivos: se calcula como

$$TFP = \frac{\text{Casos negativos incorrectamente clasificados}}{\text{Total de casos negativos}}$$

A partir del umbral de discriminación del clasificador, se puede mostrar gráficamente la TVP y TFP mediante una curva ROC que es una representación gráfica en la que se dibuja en el eje de ordenadas la TVP y en el eje abscisas la TFP. En toda gráfica ROC, hay tres puntos que es necesario explicar y tener en cuenta como se puede ver en la Figura 62: con el punto (0,0) se representa la estrategia de no emitir clasificaciones positivas, de tal manera que no se comete ningún error de FP, pero tampoco aciertos con los VP. Con la estrategia contraria de emitir incondicionalmente clasificaciones positivas, se obtiene el punto (1,1). El punto (0,1) representa una clasificación perfecta. La línea diagonal que uno el punto (0,0) con (1,1) representa la estrategia de proponer la clase de pertenencia de manera aleatoria. Se representa en verde una curva ROC. De manera no formal se puede afirmar que un punto en una gráfica ROC es mejor que otro si su posición en la misma es más próxima al punto (0,1) que el del primero (mayor valor de tasa de verdaderos positivos y menor valor de tasa de falsos positivos, o ambos).

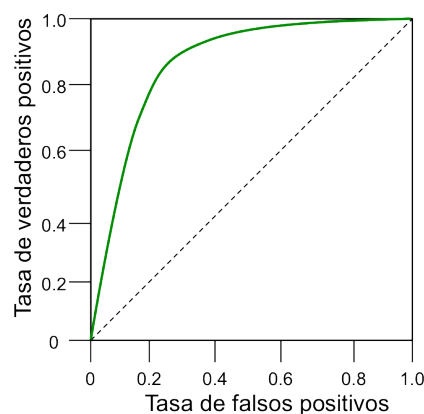


FIGURA 62. IDENTIFICACIÓN DE PUNTOS IMPORTANTES EN UNA CURVA ROC

Un clasificador discreto es aquel que como salida tiene una etiqueta correspondiente a la clase de pertenencia. Cada uno de estos clasificadores, produce un único punto en la gráfica como los mostrados en la Figura 63 como A, B y C.

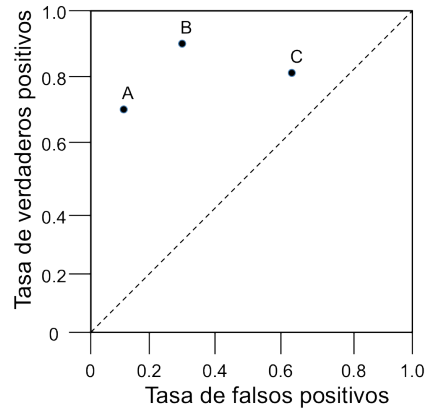


FIGURA 63. PUNTOS INDIVIDUALES CORRESPONDIENTES A CLASIFICADORES DISCRETOS

Para calcular un único valor que represente el rendimiento de un clasificador a partir de una curva ROC, se calcula el valor del área bajo la curva ROC o AUROC [266,267]. Es un valor comprendido entre el 0.0 y el 1.0 ya que se calcula como la proporción de área de una curva ROC contenida en un cuadrado de lado 1. Por debajo de la línea diagonal, el área es 0.5 y ningún clasificador realista debería obtener un valor inferior de AUROC de 0.5.

ANEXO C: SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)

En general, y para las técnicas basadas en kernel en particular, es recomendable siempre que sea posible, trabajar con un conjunto balanceado de datos [190,191] ya que su principal objetivo es optimizar su función objetivo obteniendo el mayor valor de acierto. Cuando se entrena con conjuntos desbalanceados de datos, tiende a producirse clasificadores que sobrepredicen la clase mayoritaria. Esto sucede porque el clasificador intenta reducir el error global, de manera que el error del clasificador no tiene en cuenta la distribución que puedan tener los datos, es decir, los algoritmos de clasificación tienen una tendencia hacia la clase mayoritaria. Es por este motivo que evaluar un clasificador únicamente por su valor de acierto no es adecuado, y debe de utilizarse algún tipo de medida combinada.

Existen múltiples trabajos en los que se puede apreciar que un clasificador entrenado con un conjunto balanceado de datos ofrece mejores resultados globales comparado con un conjunto desbalanceado [192-194]. Existen dos posibilidades, hacer un sobre muestreo (balanceando los datos de entrenamiento) o un sub muestreo (eliminando ejemplos ruidosos). En este caso, se propone la utilización de alguna técnica que se encargue de corregirlo como SMOTE [195] (*Synthetic Minority Over-Sampling Technique*).

Esta técnica realiza un sobre muestreo dirigido de los casos de la clase minoritaria, se puede ver una comparación de este algoritmo sobre problemas desbalanceados y ganancia en interpretabilidad en [268]. Se generan para la clase minoritaria ejemplos sintéticos que se interpolan de manera que se correspondan con los k vecinos más próximos. Por ejemplo, si se están usando 5 vecinos próximos y se necesita hacer un *over-sampling* del 200%, solo se seleccionan dos vecinos de los 5 vecinos próximos del ejemplo y cada uno se genera en la dirección del vecino seleccionado. El proceso sigue estos pasos:

- Se toma la diferencia entre el ejemplo que se está estudiando y su vecino más próximo

- Se multiplica este valor por un número aleatorio entre 0 y 1
- Se le suma al ejemplo que se estaba estudiando

Considerando la representación en el espacio de características de los ejemplos de un conjunto de datos cualquiera, los nuevos ejemplos sintéticos generados serán como se muestra en la Figura 64.

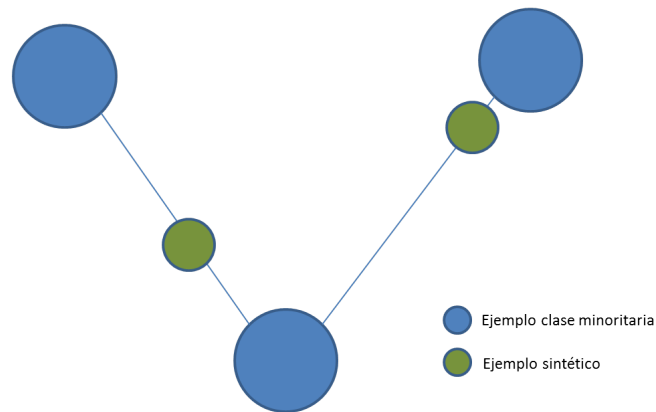


FIGURA 64. SMOTE: GENERACIÓN DE NUEVOS EJEMPLOS SINTÉTICOS SEGÚN LOS VECINOS PRÓXIMOS

Es frecuente encontrarse con problemas no balanceados en el dominio médico debido a que suelen tenerse pocos datos de pacientes enfermos y muchos de pacientes sanos, residiendo el conocimiento más novedoso en los casos raros.

ÍNDICE GENERAL DE PALABRAS SIGNIFICATIVAS

análisis de texturas, 13, 17, 19, 20, 22, 31, 40, 42, 45, 54, 55, 57, 58, 59, 60, 61, 62, 66, 70, 76, 77, 82, 91, 97, 107, 116, 121, 124, 147, 149, 150

contraste de hipótesis, 68, 71, 72, 89, 174

Friedman, 72, 74, 75, 109, 111, 112, 113, 121, 135, 139, 140, 141, 167

Quade, 72, 75, 123, 124, 167

FSMKL, 48, 76, 79, 88, 97, 105, 106, 107, 108, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 122, 123, 124, 126, 127, 136, 141, 143, 146, 147, 149, 151

kernel

gaussiano, 79

lineal, 46, 79

polinomial, 100, 106

RBF, 79, 98, 99, 100, 105, 106, 117, 121, 144, 146

métodos de selección

embedded, 42, 44, 88, 149, 151

filter, 42, 44, 48, 61

wrapper, 42, 43, 44, 45, 88, 149, 151

SVM, 30, 31, 42, 43, 44, 45, 46, 47, 48, 55, 56, 57, 76, 77, 78, 79, 87, 88, 97, 98, 99, 100, 104, 105, 106, 107, 108, 110, 111, 112, 113, 117, 149, 151, 154, 160, 167

test estadísticos

hipótesis nula, 71, 73, 74, 111, 122, 123, 139, 174

test de Barlett, 111, 123, 139

test de Shapiro, 72, 109, 111, 121, 122, 135