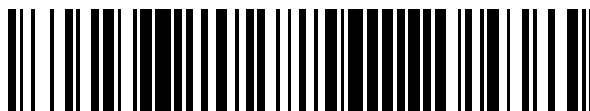


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 611 207**

51 Int. Cl.:

A61B 5/00 (2006.01)

G06K 9/62 (2006.01)

G06K 9/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.03.2010 E 14194588 (1)**

97 Fecha y número de publicación de la concesión europea: **23.11.2016 EP 2862505**

54 Título: **Técnicas de clasificación para diagnóstico médico mediante espectroscopía óptica**

30 Prioridad:

25.03.2009 US 163423 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

05.05.2017

73 Titular/es:

**TRUSTEES OF BOSTON UNIVERSITY (100.0%)
One Silber Way
Boston, MA 02215, US**

72 Inventor/es:

**BIGIO, IRVING;
RODRIGUEZ-DIAZ, ELADIO y
CASTANON, DAVID**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 611 207 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Técnicas de clasificación para diagnóstico médico mediante espectroscopía óptica

Antecedentes

1. Campo

5 La presente invención se refiere a sistemas y métodos para clasificar datos de imágenes espectroscópicas ópticas de tejidos biológicos tales como cáncer y otras patologías para diagnósticos médicos.

2. Técnica relacionada

10 El cribado actual para la detección temprana del cáncer de mama, por ejemplo, se basa en anomalías visibles en una mamografía o grumos detectados por los pacientes o médicos que usan palpación. Antes de iniciar cualquier tratamiento, el diagnóstico debe ser confirmado. Esto se logra generalmente mediante la realización de una biopsia, que es un procedimiento invasivo, y luego determinar la histología del tumor. Una alternativa menos invasiva es el uso de la citología por aspiración con aguja fina (FNA). Sin embargo, la FNA se utiliza poco frecuentemente, debido a una tasa significativa de falsos negativos. Aproximadamente 50.000 diagnósticos de lumpectomías se realizan anualmente en los EE.UU. de aquellos, sólo unos 12.000 resultan malignos cuando la histología es realizada por un patólogo. Si se hubiera conocido de antemano que las restantes 38.000 lesiones eran benignas, la cirugía potencialmente desfigurante podría haber sido evitada, ya que muchas lesiones benignas se resuelven espontáneamente en el tiempo, sin intervención.

20 El cáncer de cuello uterino es el tercer cáncer más común en mujeres en todo el mundo y es la principal causa de mortalidad por cáncer en mujeres en países en desarrollo. Cuando las lesiones precancerosas se detectan temprano, se pueden tratar fácilmente mediante ablación o escisión. En etapas más avanzadas, el cáncer de cuello uterino a menudo requiere histerectomía, quimioterapia, radioterapia o terapia de quimiorradiación combinada. El cribado actual para este tipo de cáncer se realiza primero mediante un frotis Papanicolaou (Pap), con sensibilidades y valores de especificidad que oscilan entre el 11% y el 99% y el 14% al 97%, respectivamente, y los resultados suelen estar disponibles en dos semanas. La segunda etapa del proceso de selección, después de un frotis de Pap anormal, es una coloscopia. Esta prueba tiene una excelente sensibilidad (>90%) pero pobre especificidad (<50%), incluso en manos de un profesional experimentado. Debido a la escasa especificidad, se requiere una biopsia para confirmar el diagnóstico. Actualmente, las mujeres a menudo esperan hasta ocho semanas para ser tratadas como parte de la atención estándar en el diagnóstico y tratamiento del cáncer cervical después de un frotis de Pap anormal.

30 El esófago de Barrett es una condición precancerosa que es un factor de riesgo importante en el desarrollo del adenocarcinoma esofágico, la forma más común de cáncer de esófago. Se asocia con la enfermedad crónica gastrointestinal del reflujo y está aumentando en incidencia en países occidentales. Se cree que el desarrollo de la malignidad es una progresión desde la mucosa de Barrett no diseminada, a través de la displasia de bajo grado (LGD) hasta la displasia de alto grado (HGD), hasta el carcinoma. En consecuencia, es fundamental identificar a los pacientes con esófago de Barrett que están más en riesgo de desarrollar cáncer. Los parches de displasia dentro de una sección de la mucosa de Barrett no pueden ser detectados visualmente usando la endoscopia convencional de luz blanca. El diagnóstico requiere múltiples biopsias al azar y posterior examen histológico. Pueden tomarse hasta 20-30 biopsias "al azar" en una sesión. Este es un procedimiento que consume mucho tiempo (y es caro), lo que conlleva cierto grado de riesgo para el paciente. Para cada biopsia convencional, la herramienta de biopsia debe ser retirada del endoscopio y la muestra retirada antes de que la herramienta pueda ser reinsertada para la siguiente biopsia. Debido a que las biopsias se toman al azar dentro de una sección del esófago de Barrett, la detección de cambios precancerosos es relativamente pobre.

45 En los últimos años, se han propuesto varias técnicas de espectroscopía como métodos potenciales para distinguir entre diferentes patologías tisulares. La motivación de estas técnicas es reducir o eliminar la necesidad de extracción quirúrgica de muestras de tejido de biopsia. En su lugar, se aplica alguna forma de análisis espectral del tejido a mediciones obtenidas con una sonda óptica colocada sobre o cerca de la superficie del tejido en cuestión. Luego se intenta un diagnóstico del tejido sobre la base de estas mediciones, *in situ*, no invasiva y en tiempo real. Además, existe la posibilidad de reducir el coste de la atención médica y la angustia del paciente como consecuencia de la necesidad reducida de histología y la necesidad del entorno quirúrgico necesario para tomar las muestras de biopsia. Algunas de estas técnicas espectroscópicas propuestas incluyen espectroscopía Raman, espectroscopía de autofluorescencia, espectroscopía de fluorescencia, espectroscopía de reflectancia y espectroscopía de dispersión elástica.

50 El cribado y/o la detección de cáncer en una etapa temprana son de importancia significativa ya que muchas incidencias de la enfermedad se pueden tratar con éxito en etapas tempranas. En los últimos años, estos métodos

de espectroscopía óptica han recibido mayor atención para este propósito, debido a que poseen algunas propiedades deseables - son no invasivas, *in situ*, y los resultados pueden obtenerse casi en tiempo real. Estos métodos proporcionan datos sensibles a los cambios del tejido subyacente (por ejemplo, estructurales, bioquímicos), que pueden ser explotados para el desarrollo de algoritmos de diagnóstico. Se han utilizado varios métodos de reconocimiento de patrones estadísticos y métodos de aprendizaje automático para desarrollar estos algoritmos de diagnóstico.

Por ejemplo, se utilizó un clasificador MAP A (Maximum A-Posteriori) para distinguir entre lesiones intraepiteliales escamosas (SIL) y epitelios escamosos normales, y para distinguir entre lesiones intraepiteliales escamosas de alto grado (HGSIL) y escamosas intraepiteliales de bajo grado (LGSILs) utilizando espectroscopía de fluorescencia aplicada al tejido cervical. Las probabilidades posteriores se calcularon después de ajustar los datos de entrenamiento a una función gamma. Se informó una sensibilidad y especificidad de 82% y 68%, respectivamente, para el primer caso y de 79% y 78% para el segundo caso.

También se ha utilizado un análisis discriminante lineal. Se informó una precisión de clasificación del 72% para distinguir melanoma maligno de nevus benignos en la piel usando espectros de reflectancia. Se utilizó espectroscopía de dispersión elástica para detectar displasia en el esófago. Se obtuvieron sensibilidad del 77% y especificidad del 77% en la detección de biopsias de "alto riesgo". La misma técnica de espectroscopía se ha empleado para detectar el cáncer en el ganglio linfático centinela para el cáncer de mama, con una sensibilidad resultante del 75% y especificidad del 89%.

También se ha utilizado el discriminante lineal de Fisher. Este método obtiene la función lineal que produce la razón máxima entre la dispersión entre clases y la dispersión dentro de la clase. La espectroscopía Raman se utilizó para distinguir entre tejido normal, displasia de bajo grado y displasia/carcinoma de alto grado *in situ* usando modelos de rata. Se obtuvo una especificidad del 93% y una sensibilidad del 78% para detectar displasia de bajo grado, y se obtuvo una sensibilidad y especificidad del 100% para detectar displasia/carcinoma de alto grado *in situ*. La espectroscopía de fluorescencia se aplicó con el fin de detectar el cáncer en la cavidad oral. Los resultados fueron de una sensibilidad del 73% y una especificidad del 92%, después de seleccionar características utilizando la eliminación de características recursivas (RFE).

Se han utilizado espectroscopía de fluorescencia y reflectancia, respectivamente, para diferenciar tejido cervical normal y precanceroso (neoplásico) usando un clasificador de distancia Mahalanobis. Se observó una sensibilidad del 72% y una especificidad del 81% al discriminar entre tejido escamoso normal y lesiones intraepiteliales escamosas de alto grado, mientras que se obtuvo una sensibilidad del 72% y una especificidad del 83% al discriminar el tejido normal columnar de las lesiones intraepiteliales escamosas de alta graduación. Se obtuvo una sensibilidad y especificidad promedio de 78% y 81% respectivamente cuando se realizó el análisis pareado entre tejido normal escamoso, tejido normal columnar, lesiones intraepiteliales escamosas de bajo grado y lesiones intraepiteliales escamosas de alto grado. El documento US 6,135,965 A describe un aparato y métodos para la detección espectroscópica de anomalía tisular utilizando redes neurales para analizar medidas *in vivo* de espectros fluorescentes.

Otro método que se aplica a datos espectroscópicos es redes neuronales artificiales (ANN). Estos son típicamente conocidos por ser capaces de manejar problemas no lineales. Como ejemplo de su uso, se utilizó un clasificador de ANN para distinguir melanoma maligno de nevus benignos en la piel utilizando espectros de reflectancia, con una precisión de clasificación del 86,7%. ANN produjo sensibilidades de 69% y 58%, y especificidades de 85% y 93%, para el tejido mamario y ganglios centinela, respectivamente, utilizando datos de espectroscopía de dispersión elástica.

En los últimos años, las máquinas de vectores de soporte (SVM) han recibido mayor atención en este tipo de aplicaciones. Esto se debe en parte al hecho de que las SVM muestran una buena capacidad de generalización y son capaces de producir fronteras de decisión no lineales a través del mapeo implícito de los datos a un espacio de mayor dimensión mediante el uso de funciones de núcleo. Se utilizaron SVM lineales para clasificar el tejido no maligno y maligno de mama medido con espectroscopía de fluorescencia, obteniéndose una sensibilidad del 70% y una especificidad del 92%. También se han utilizado SVMs con núcleos de función base lineal y radial (RBF). Se obtuvieron sensibilidades del 94% y 95%, y especificidades del 97% y 99%, respectivamente, para distinguir el tejido normal del carcinoma nasofaríngeo usando espectros de autofluorescencia. La espectroscopía de fluorescencia se ha aplicado con el fin de detectar el cáncer en la cavidad oral. Los resultados fueron de una sensibilidad de 88%, 90%, 93% y especificidad de 94%, 95%, 97%, para SVMs lineales, polinomiales y RFB, respectivamente, después de seleccionar características usando la eliminación de rasgos recursivos (RFE).

Como máquinas vectoriales de soporte (SVMs) han obtenido mayor atención por problemas de clasificación, se han presentado varias reglas de rechazo de errores para este tipo de clasificador. Para los puntos cercanos al hiperplano óptimo, el clasificador puede no estar muy seguro de las etiquetas de clase asignadas. En estos enfoques anteriores, se propuso un esquema de rechazo en el que se rechazan las muestras cuya distancia al hiperplano de separación está por debajo de cierto umbral. Un enfoque similar se utilizó cuando la distancia de los puntos de datos

5 probados de la óptima separación hiperplano acotados con el fin de rechazar un porcentaje definido por el usuario de los patrones de errores clasificados, lo que permite la reducción del riesgo esperado. Una regla de rechazo basada en ROC ha sido propuesta para SVMs. La curva ROC para el clasificador SVM se obtiene variando el umbral de decisión, que se supone es cero, de $-\infty$ a ∞ . Las tasas positivas verdaderas y falsas positivas se obtienen a partir de las densidades condicionales de clase producidas por los resultados de las SVMs. También se ha presentado un umbral de rechazo de distancia para los clasificadores SVM. El resultado de SVM es la distancia de ese patrón de entrada particular al hiperplano de separación óptimo.

10 El fundamento de estas reglas de rechazo de errores puede remontarse al trabajo presentado por Chow. Estas reglas de rechazo de errores son desventajosas porque suponen que se conoce la distribución de probabilidad para cada clase. En la mayoría de las aplicaciones de reconocimiento de patrones, las formas paramétricas comunes rara vez se ajustan a las densidades encontradas en la práctica.

Chow exploró primero el rechazo de errores en el contexto de la teoría de la decisión de Bayes. Dentro de este marco, se dice que un vector de entidad \mathbf{x} pertenece a la clase w_k si

$$P(w_k/\mathbf{x}) = \max_i P(w_i/\mathbf{x}), \quad i = 1, \dots, N \quad (1)$$

15 donde $P(w_i/x)$ es la probabilidad a posteriori y N es el número total de clases. Esta regla divide el espacio de características en N regiones $D_1 \dots D_N$ y clasifica \mathbf{x} como w_k si se encuentra en la región D_k . Además, es óptimo en el sentido de que minimiza la probabilidad de error, también llamado error de Bayes,

$$P_E = \sum_{i=1}^N \int \sum_{\substack{j=1 \\ j \neq i}}^N p(\mathbf{x}/w_j) P(w_j) d\mathbf{x}. \quad (2)$$

20 Chow introduce la opción de rechazo para obtener una probabilidad de error menor que el error de Bayes. Esto se logra evitando clasificar patrones que probablemente se clasifiquen erróneamente. La regla de Chow establece que un vector característico \mathbf{x} se clasifica como perteneciente a la clase w_k si

$$\max_{i=1, \dots, N} P(w_i/\mathbf{x}) = P(w_k/\mathbf{x}) \geq t \quad (3)$$

y rechazado si

$$25 \quad \max_{i=1, \dots, N} P(w_i/\mathbf{x}) = P(w_k/\mathbf{x}) < t \quad (4)$$

donde t es el umbral de rechazo. Por lo tanto, la introducción de la opción de rechazo divide el espacio de características en $N+1$ regiones de decisión D_0, D_1, \dots, D_N y clasifica \mathbf{x} como w_k si se encuentra en la región D_k y la rechaza si se encuentra en D_0 . Es óptima puesto que $P(w_k/x)$ es la probabilidad condicional de clasificar correctamente el patrón \mathbf{x} . Obsérvese que tanto la probabilidad de error (2) como la probabilidad de rechazo

$$30 \quad P_R = \int \sum_{D_0}^N p(\mathbf{x}/w_i) P(w_i) d\mathbf{x} \quad (5)$$

son ahora funciones del umbral t . Chow indica que, puesto que (2) y (5) son funciones monótonas del umbral t , el rendimiento del sistema de reconocimiento es completamente descrito por la curva resultante de (2) frente a (5). En esta curva de compensación de rechazo de error, P_E disminuye y P_R aumenta a medida que aumenta el umbral t . En particular, P_E es igual al error de Bayes y $P_R=0$ para $t^*=0$, y $P_E=0$ para $t=1$. Se presentó una relación similar entre las tasas de aceptación falsa y las tasas de rechazo falso en función del umbral de rechazo en la aplicación de sistemas de verificación biométrica.

40 Se propuso otro esquema de rechazo para mejorar la confiabilidad en redes neuronales. Se definen dos reglas de clasificación y encuentra un umbral para cada una. Sea O_k el nodo de salida correspondiente a la clase k , si las muestras de entrada corresponden a la k th clase entonces el nodo de salida $O_k=1$ mientras que todos los demás son iguales a cero. La primera regla establece que un patrón de entrada pertenece a la clase k si

$$\max_{i=1, \dots, N} O_i = O_k \geq \sigma \quad (6)$$

donde σ es el umbral de rechazo, similar a la regla de Chow. La segunda regla establece que

$$O_k - O_j < \delta \quad (7)$$

5 donde O_j es el nodo de salida con el segundo valor más alto y δ es el umbral de rechazo. Por lo tanto, si la diferencia entre los dos valores de salida más altos es menor que un cierto umbral, el patrón de entrada no está clasificado. Los dos umbrales se obtienen entonces maximizando una función de rendimiento que depende de las tasas de error y rechazo, así como de sus respectivos costes.

10 También se ha investigado la optimización de la regla de Chow. Algunos dicen que la regla de Chow es óptima sólo si las probabilidades posteriores de las clases de datos son exactamente conocidas; Sin embargo, generalmente no es el caso y las probabilidades posteriores tienen que ser estimadas a partir de los datos de entrenamiento. Como resultado, se obtienen resultados subóptimos cuando esta regla se aplica a las probabilidades estimadas, puesto que las regiones de decisión se desplazan con respecto a donde estarían en el caso óptimo. El uso de múltiples umbrales dependientes de la clase se ha propuesto como una solución. En este enfoque la regla de Chow se modifica, un patrón x se clasifica como perteneciente a la clase w_k si

$$\max_{i=1,\dots,N} \hat{P}(w_i/x) = \hat{P}(w_k/x) \geq t_k \quad (8)$$

15 y rechazado si

$$\max_{i=1,\dots,N} \hat{P}(w_i/x) = \hat{P}(w_k/x) < t_k. \quad (9)$$

Aquí $\hat{P}(w_i/x)$ es la probabilidad posterior estimada. Los umbrales se determinan maximizando la probabilidad de exactitud sujeto a mantener la probabilidad de rechazo por debajo de un valor definido por el usuario. Tanto la precisión como las probabilidades de rechazo son función de los umbrales de clase.

20 Otra regla de rechazo se basó en el análisis de la curva de características de funcionamiento del receptor (ROC). Las dos clases se llaman Positivas (P) y Negativas (N) y la regla de decisión se define como:

asignar la muestra a N	si $x < t_N$
asignar la muestra a P	si $x > t_P$
rechazar la muestra	si $t_N \leq x \leq t_P$

25 donde t_N y t_P ($t_N \leq t_P$) son los umbrales de rechazo. Los umbrales óptimos maximizan una función de desempeño definida por las tasas de falsos negativos, verdaderos negativos, falsos positivos, verdaderos positivos y de rechazo, los costes respectivos y las probabilidades previas. La solución produce un conjunto de líneas rectas paralelas cuyas pendientes están determinadas por los costes y las probabilidades previas. Los valores óptimos para los umbrales se encuentran entonces buscando el punto de la curva ROC, construido mediante la representación gráfica de la tasa positiva real frente a la tasa de falsos positivos, que intersectan estas líneas y tienen un valor mínimo.

30 Otra regla de rechazo que se ha considerado se refiere al conocimiento incompleto sobre las clases. En este trabajo se definieron dos umbrales de rechazo. El primero, el umbral de rechazo de la ambigüedad, al igual que la regla de Chow, pretende rechazar muestras con alto riesgo de mala clasificación para disminuir el error de clasificación. El segundo umbral de rechazo de distancia denotado, tiene como objetivo disminuir la probabilidad de clasificar erróneamente un patrón de entrada x en una de las N clases cuando está "lejos" de las clases conocidas (es decir, valores atípicos). La suposición es que no todos los patrones provienen de una de las N clases previamente definidas. Esta misma regla de rechazo se aplicó a clasificadores de redes neuronales. Se presenta la regla óptima de rechazo selectivo de la clase. Este enfoque es otra extensión de la regla de rechazo, cuando un patrón de entrada no puede clasificarse de forma fiable como una de las N clases definidas, en lugar de ser rechazado, se asigna a un subconjunto de clases a las que probablemente pertenece el patrón. Así, el espacio de características se divide en regiones de decisión $2^N - 1$ en lugar de regiones $N+1$ como en la regla de Chow.

35 Aunque estos métodos abarcan una amplia gama de aplicaciones utilizando diferentes métodos de espectroscopía y varios tipos de clasificadores, las sensibilidades y especificidades obtenidas no varían mucho para cada uno de los casos presentados. Con pocas excepciones, la sensibilidad y especificidad promedio fluctúa entre el 70% y el 85%.

45 Por lo tanto, lo que se necesita es un método mejorado para clasificar datos de espectroscopía óptica. Estos

métodos mejorados pueden usarse para mejorar el diagnóstico y, por tanto, el tratamiento del cáncer y otras patologías.

Resumen

- 5 La presente invención se define en y por las reivindicaciones adjuntas. El siguiente resumen de la invención se incluye con el fin de proporcionar una comprensión básica de algunos aspectos y características de la invención. Este resumen no es una visión general extensa de la invención y como tal no se pretende identificar particularmente elementos clave o críticos de la invención o delinear el alcance de la invención. Su único propósito es presentar algunos conceptos de la invención en una forma simplificada como un preámbulo a la descripción más detallada que se presenta a continuación.
- 10 Se describe aquí un método para clasificar una muestra de tejido que incluye entrenar un clasificador para determinar una región de rechazo de datos de especímenes biomédicos correspondientes a muestras de tejido que se pueden clasificar erróneamente; clasificar la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido no están en la región de rechazo; y detener la clasificación de la muestra de tejido si los datos de espectros biomédicos correspondientes a la muestra de tejido están en la región de rechazo.
- 15 El método también puede incluir la realización de una biopsia de la muestra de tejido si se mantiene la clasificación.
- El método puede incluir también la realización de mediciones ópticas adicionales de la muestra de tejido si se mantiene la clasificación.
- El entrenamiento del clasificador puede incluir la formulación de un problema de entrenamiento como un algoritmo de entrenamiento dual de problema de optimización convexa.
- 20 Los datos de espectros biomédicos pueden ser datos de ESS.
- 25 De acuerdo con un aspecto de la invención, se describe un método para clasificar una muestra de tejido que incluye clasificar una primera región de datos de espectros biomédicos utilizando un primer clasificador; clasificar una segunda región de datos de espectros biomédicos utilizando un segundo clasificador, siendo la primera región diferente de la segunda región; y combinar la clasificación del primer clasificador con la clasificación del segundo clasificador para determinar una clasificación de la muestra de tejido.
- Los datos de espectros biomédicos pueden ser datos de ESS.
- La clasificación del primer clasificador y la clasificación del segundo clasificador se seleccionan del grupo que consiste en positivo, negativo y rechazado.
- La primera región puede solapar la segunda región.
- 30 El método también puede incluir la clasificación de una tercera región de los datos de espectros biomédicos utilizando un tercer clasificador y la clasificación del tercer clasificador puede combinarse con la clasificación del primer clasificador y la clasificación del segundo clasificador para determinar la clasificación de la muestra de tejido.
- Clasificar la primera región puede incluir la detención de la clasificación de la primera región. Clasificar la segunda región puede incluir la detención de la clasificación de la segunda región.
- 35 El primer clasificador y el segundo clasificador pueden ser máquinas de soporte de vectores. El primer clasificador y el segundo clasificador pueden ser máquinas vectoriales de soporte incrustadas con rechazo de errores.
- La combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador puede utilizar una regla de votación mayoritaria. La combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador puede utilizar una regla ingenua de Bayes.
- 40 Según un aspecto adicional de la descripción, se describe un medio de almacenamiento legible por máquina que comprende instrucciones ejecutables por una máquina de procesamiento de datos para realizar los métodos descritos en la presente memoria. De acuerdo con aspectos adicionales de la descripción, se describe un sistema informático que incluye memoria y un procesador configurado para realizar las etapas de método descritas en la presente memoria y se describe un sistema informático que incluye medios para llevar a cabo las etapas de método descritas en la presente memoria.
- 45 Según otro aspecto de la descripción, se describe un sistema configurado para clasificar una muestra de tejido que

incluye una sonda óptica configurada para medir los datos de espectros biomédicos; y un sistema de clasificación acoplado a la sonda óptica que comprende una máquina de vector de soporte con rechazo de error incorporado para clasificar los datos de espectros biomédicos.

5 La sonda óptica puede ser una herramienta integrada de pinzas de biopsia. La herramienta de pinzas de biopsia integrada puede incluir una fibra óptica de iluminación para dirigir la luz en el tejido que se va a fotografiar y una fibra óptica de recogida para recoger la luz dispersada en el tejido. Los datos de los espectros biomédicos pueden ser datos de ESS.

10 El sistema de clasificación puede incluir una pluralidad de máquinas de vector de soporte con rechazo de errores incorporado para clasificar los datos de espectros biomédicos. Cada una de la pluralidad de máquinas vectoriales de soporte con rechazo de error incorporado está configurada para clasificar una región diferente de los datos de los espectros. El sistema de clasificación puede combinar el resultado de cada una de la pluralidad de máquinas de vectores de soporte con rechazo de errores incorporados para realizar una clasificación final de los datos de espectros biomédicos.

Breve descripción de los dibujos

15 Los dibujos adjuntos, que se incorporan y constituyen una parte de esta memoria descriptiva, ejemplifican las realizaciones de la presente invención y, junto con la descripción, sirven para explicar e ilustrar principios de la invención. Los dibujos están destinados a ilustrar las características principales de las formas de realización de ejemplo de una forma diagramática 14. Los dibujos no pretenden representar todas las características de realizaciones reales ni dimensiones relativas de los elementos representados, y no están dibujados a escala.

20 La figura 1A ilustra una geometría óptica de una sonda de fibra utilizada para ESS.

La Figura 1B ilustra un sistema de detección de sondas de fibra.

La figura 2A ilustra una pinza de biopsia integrada.

La Figura 2B ilustra una pinza de biopsia integrada.

La Figura 3 ilustra espectros de ESS medidos de ejemplo para tejido normal y canceroso.

25 La Figura 4 ilustra un algoritmo SFFS.

La Figura 5 ilustra un algoritmo SVDSS.

Las Figuras 6A y 6B ilustran los espectros medios y el histograma de las características seleccionadas con SFFS.

La figura 7A ilustra el comportamiento de la función $h(\xi_j, \epsilon)$.

La figura 7B ilustra el comportamiento de la función $g(\alpha_i, \epsilon)$.

30 La figura 8 ilustra parábolas.

La Figura 9 ilustra un proceso para optimizar analíticamente dos multiplicadores de Lagrange.

La Figura 10 ilustra las violaciones de KKT.

La Figura 11 ilustra regiones de rechazo con diferentes núcleos.

35 La Figura 12 ilustra la ponderación dependiente de la clase en un clasificador SVM con SMOER y con un núcleo RBF.

La Figura 13 ilustra un sistema clasificador múltiple.

La Figura 14 ilustra un algoritmo de entrenamiento de AdaBoost.

La figura 15 es un diagrama de bloques de un sistema informático de ejemplo.

Descripción detallada

5 Diversos tipos de espectroscopía óptica, en general, incluyendo la espectroscopía de dispersión elástica (ESS), proporcionan una aproximación mínimamente invasiva para obtener mediciones espectrales de las propiedades de dispersión del tejido. Estas técnicas permiten la adquisición de firmas espectrales del tejido subyacente, con el fin de detectar patologías que incluyen, por ejemplo, cáncer. El método ESS es sensible a la dependencia de la longitud de onda de las propiedades de dispersión óptica del tejido, que varían debido a cambios arquitectónicos a nivel celular y subcelular como resultado de las diferentes patologías. Estos cambios influyen en las firmas espectrales medidas, con variaciones espectrales que dependen de la patología.

10 Las realizaciones de la presente invención hacen uso de sistemas de reconocimiento de patrones matemáticos/estadísticos, con nuevos enfoques en la implementación, para distinguir entre diferentes patologías y condiciones benignas (por ejemplo, tejido normal o canceroso) dado los espectros medidos usando espectroscopía óptica, tales como EES. En realizaciones de la presente invención, un espectro dado puede especificarse como normal o canceroso, pero también puede no ser clasificado (es decir, rechazado). Las muestras que no están clasificadas (es decir, rechazadas) incluyen muestras con un alto riesgo de clasificación errónea (es decir, muestras situadas cerca del límite de decisión dentro de ciertos umbrales). El algoritmo de entrenamiento se reformula para cada uno de los clasificadores para determinar de forma óptima el límite de decisión y estos umbrales sin cambiar significativamente la estructura original del clasificador.

20 Las realizaciones de la presente invención están dirigidas a sistemas clasificadores múltiples que toman una decisión diagnóstica final basada en la clasificación de cada uno de los clasificadores individuales. Cada uno de los clasificadores individuales puede estar diseñado para utilizar características extraídas de una región limitada del espectro de ESS. Cada uno de los clasificadores individuales también puede ser clasificadores SVM incorporados con rechazo de errores. En particular, el algoritmo de entrenamiento de estos clasificadores SVM puede ser formulado para determinar de manera óptima el límite de decisión.

25 Las Figuras 1A-1B ilustran una sonda EES de ejemplo 100. La sonda EES 100 se utiliza para recoger los datos de los espectros EES para su análisis. La sonda EES 100 está configurada para ser posicionada en contacto óptico con el tejido bajo examen 104. Como se muestra en la Figura 1A, la sonda EES 100 incluye una fibra iluminadora 108 y una fibra colectora 112. La luz 116, proporcionada por una fuente de luz 120, está dirigida al tejido 104 por la fibra iluminadora 108. La luz 116 experimenta una dispersión múltiple 124 a través de un pequeño volumen del tejido que se está examinando. La luz 116 es entonces recogida y transmitida por la fibra de recogida 112 al espectrómetro de análisis (el detector 128 que está acoplado a un ordenador e interfaz 132 como se muestra en la figura 1B. En una realización, la fuente de luz 120 es una fuente de luz de banda ancha de lámpara de arco de Xenón pulsada y el ordenador y la interfaz 132 es un ordenador portátil con software ESS personalizado y espectrómetro incorporado (por ejemplo, S2000 de Ocean Optics, Inc.)

35 La longitud de trayectoria efectiva resultante de los fotones recogidos es generalmente varias veces mayor que la separación real de las puntas de fibra. Se apreciará que la sonda de fibra examina solamente el sitio que está en contacto óptico con la misma y no representa la superficie del tejido.

Las Figuras 2A y 2B ilustran un ejemplo de pinza 200 de biopsia integrada. La pinza 200 de biopsia integrada incorpora una sonda 100 de ESS en pinzas de biopsia 204. Las pinzas 204 incluyen un hipotubo 208 metálico que incluye las fibras 108, 112 ópticas. El hipotubo 208 está situado en una cámara central que se extiende a lo largo de la pinza de biopsia integrada y se extiende dentro del espacio entre las mordazas 212, 216.

40 Las fibras 108, 112 ópticas en la pinza 200 de biopsia integrada son más pequeñas que las fibras ópticas convencionales y más cercanas entre sí que las fibras ópticas convencionales. En una realización, la separación entre las fibras 108, 112 de iluminación y de recogida es de aproximadamente 250 μm de centro a centro y cada fibra 108, 112 tiene un diámetro de aproximadamente 200 μm y una apertura numérica de aproximadamente 0.22 en aire. El diámetro del hipotubo 208 era de aproximadamente 0.470 mm. Se apreciará, sin embargo, que la geometría de la sonda puede tener una configuración diferente a la descrita en la presente memoria.

50 La sonda 100 genera un espectro que caracteriza las dependencias de la longitud de onda tanto de la dispersión como de la absorción del tejido subyacente, sin separar estas contribuciones. La Figura 3 ilustra un espectro de ejemplo para pólipos de colon. En el ejemplo mostrado en la Figura 3, se midieron medidas de ESS de 280 pólipos de 115 pacientes (98 correspondieron a pólipos hiperplásticos, 85 a crecimientos normales, 10 a pólipos inflamatorios, 83 a adenomas tubulares, 1 a adenoma de vellosidad tubular, 2 a adenocarcinoma y 1 a displasia de alto grado). Los datos se agruparon en polipéptidos no hipertróficos (pólipos hiperplásticos, pólipos normales e inflamatorios) y pólipos difusos (adenoma tubular, adenoma vellosidad tubular, adenocarcinoma y displasia de alto grado). En la Figura 3, el espectro medio para las mediciones no displásicas y displásicas se muestran con su desviación estándar.

55 Los algoritmos de diagnóstico que se basan en métodos de reconocimiento de patrones incluyen (2) pretratamiento y extracción/selección de características y (2) clasificación. Estos algoritmos luego se someten a (3) evaluación del rendimiento para calibrar su exactitud.

5 En el pretratamiento, los datos se preparan para el análisis, e incluyen típicamente normalización del conjunto de datos y/o operaciones de reducción de ruido (por ejemplo, suavizado de los espectros). En la extracción/selección de características, la información más relevante del conjunto de datos se obtiene de tal manera que la separabilidad de clase se conserva o mejora tanto como sea posible. Esto puede implicar transformar los datos en un subespacio de menor dimensión o seleccionar un subconjunto de características del conjunto de datos basado en ciertos criterios. Los métodos de extracción de características crean nuevas características basadas en transformaciones de combinaciones del conjunto de características original, por ejemplo Análisis de Componentes Principales (PCA). La selección de funciones selecciona el mejor subconjunto del conjunto de características de entrada, generalmente basado en alguna medida predefinida de separación de clases.

10 En la clasificación, se toma una decisión (por ejemplo, normal frente a cáncer, progresiones de la enfermedad, etc.) sobre la naturaleza del patrón de entrada (por ejemplo, los espectros de ESS). La importancia de evitar muestras mal clasificadas es clara, lo que en situaciones reales es difícil de lograr. Con este fin, se describe aquí una metodología que es capaz de identificar muestras que están en alto riesgo de ser clasificadas erróneamente. Al identificar estas muestras "rechazadas", pueden ser examinadas y diagnosticadas con mayor precisión por otros métodos.

15 En la evaluación del desempeño, se determina la exactitud de las decisiones tomadas por el clasificador. Los criterios utilizados para la evaluación son las sensibilidades, especificidades y errores de clasificación alcanzados con un clasificador particular. En algunos casos, el valor predictivo positivo y/o negativo es de interés y también se utiliza como criterio de evaluación. Cuando se emplea el rechazo de errores como criterio adicional, el porcentaje de muestras no clasificadas se considera en la evaluación de desempeño.

20 Las realizaciones de la invención se dirigen a sistemas y métodos que realizan la selección de características usando la Selección Secuencial Flotante de Avance (SFFS) en el campo de la espectroscopía biomédica para procesar previamente los datos de los espectros. SFFS es ventajoso porque el significado físico de las características no se pierde en el proceso de transformación. El método SFFS también es ventajoso porque actúa a la par con algoritmos que obtienen soluciones óptimas, es decir, métodos que buscan todos los subconjuntos posibles, pero que exigen recursos computacionales más bajos.

25 La Selección Secuencial Flotante de Avance (SFFS), mostrada en la Figura 4, incluye secuencialmente la característica más significativa en el paso directo y, después de este paso, excluye el menos significativo en un número de pasos hacia atrás, siempre y cuando los subconjuntos resultantes sean mejores que los previamente evaluados en ese nivel. Esto evita la "anidación" de los subconjuntos de características que surgen de las metodologías puras de Búsqueda Secuencial de Avance. Como se muestra en la Figura 4, la significación de una característica está determinada por la métrica $J(g)$, elegida *a priori*. Esto se elige generalmente para ser alguna medida de separabilidad de clase, por ejemplo la distancia de Bhattacharya o la tasa de error de clasificación.

30 El Análisis de Componentes Principales (PCA) es otro método para la extracción de características. El PCA reduce la dimensionalidad al restringir la atención a aquellas direcciones a lo largo de las cuales la varianza de los datos es mayor. Estas direcciones se obtienen seleccionando los autovectores de la matriz de covarianza de datos agrupados que corresponden a los valores propios más grandes. La reducción de la dimensionalidad se consigue entonces aplicando la transformación lineal de la forma:

$$y = A^T x \quad (0.1)$$

35 donde A es la matriz de transformación cuyas columnas contienen los autovectores d deseados y x es el vector de entrada. Cabe señalar que aunque PCA encuentra componentes que son útiles para representar los datos en dimensiones más bajas, no hay garantía de que estos componentes sean útiles para la discriminación entre clases. PCA es un método de extracción de características lineales sin supervisión, es decir, no se incorpora medida de separabilidad de clase mientras se encuentran los componentes principales. Además, dado que el PCA reduce la dimensionalidad por medio de una transformación lineal, los significados físicos de las características del vector de entrada se pierden efectivamente.

40 Un método alternativo para la selección de características es la selección de subconjunto de descomposición de valor singular (SVDSS), mostrada en la Figura 5. Aunque SFFS es un método de selección de entidad supervisada, es decir, se necesitan muestras marcadas en el algoritmo, SVDSS es un método de selección de características no supervisado. SVDSS selecciona las características d que mejor se aproximan a los primeros componentes d principales utilizando el algoritmo heurístico basado en la descomposición de valores singulares, como se muestra en la Figura 5. Al igual que PCA, mientras que no se utiliza criterio de separabilidad de clase, el algoritmo SVDSS exhibe buen rendimiento en aplicaciones prácticas.

Los histogramas de las características seleccionadas usando SFFS y SVDSS se muestran en las Figuras 6A y 6B,

respectivamente.

Tabla 1: Resultados de los métodos de extracción/selección de características.

Método	# de características	Sensibilidad	Especificidad	Error Tasa
PCA	15	.7417 ± .0757	.7538 ± .0528	.2499 ± .0364
SFFS	58	.7217 ± .0785	.7777 ± .0425	.2396 ± .0323
SVDSS	17	.7434 ± .0706	.7608 ± .0503	.2446 ± .0351

Si bien no hay una clara ganancia en el rendimiento, el uso de métodos de selección de características, como SFFS y SVDSS, puede ayudar a ganar una intuición en qué partes del espectro son más relevantes en términos de discriminación no displásico de pólipos displásicos. Como se muestra en las Figuras 6A y 6B, las longitudes de onda más cortas, 330-440 nm, son particularmente útiles para esta tarea. Estos experimentos condujeron al uso del 330-760 nm en lugar de 370-800 nm previamente utilizado.

Se seleccionaron máquinas de vectores de soporte (SVMs) para su clasificación. Las SVMs están en sus formas más simples clasificadoras lineales binarias donde el límite de decisión obtenido es un hiperplano que tiene el margen de separación máximo entre las clases. Dos propiedades importantes de este hiperplano son: (1) el hiperplano de separación con margen máximo es único, y (2) este hiperplano puede obtenerse resolviendo un problema de optimización cuadrática.

Sea $\mathbf{x}_i \in \mathcal{R}^d$, un vector de entrada $y_i \in \{1, -1\}$ su etiqueta correspondiente para $i = 1, \dots, l$, donde l es el número total de muestras de entrenamiento. Queremos aprender el clasificador lineal de la forma:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} + b) \quad (10)$$

que proporciona el margen máximo entre las clases. Además, suponemos que las clases son linealmente separables, es decir, se puede encontrar un hiperplano que separa las clases sin error. Si denotamos \mathbf{w} como la normal al hiperplano y $|b|/\|\mathbf{w}\|$ la distancia perpendicular desde el hiperplano al origen, donde $\|\cdot\|$ es la norma-2, entonces los puntos de entrenamiento deben satisfacer las siguientes restricciones:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq +1 & \text{for } y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b &\leq -1 & \text{for } y_i = -1 \end{aligned} \quad (11)$$

Los puntos de entrenamiento que satisfacen las igualdades en las limitaciones en (11) están en dos hiperplanos que definen el margen entre clases, y se puede ver que este margen es $2/\|\mathbf{w}\|$. Por lo tanto, encontrar el hiperplano con separación de margen máximo se puede hacer mediante la minimización de $\|\mathbf{w}\|^2$ sujeto a las restricciones en (11).

Se presentó una extensión de este problema para el caso en que los datos no son separables. Las restricciones se relajan para permitir errores, y se introdujo un plazo de penalización en la función de coste. Esto se logró introduciendo las variables de holgura positiva ξ_i , $i = 1, \dots, l$. Las restricciones se modifican a continuación:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq +1 - \xi_i & \text{for } y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b &\leq -1 + \xi_i & \text{for } y_i = -1. \\ \xi_i &\geq 0 & \forall i \end{aligned} \quad (12)$$

La nueva función de coste es entonces

$$\|\mathbf{w}\|^2 + C \sum_i \xi_i \quad (13)$$

donde C es una constante predefinida que controla la cantidad de penalidad que se da a los errores y que normalmente se selecciona usando métodos de validación cruzada. El sumando en (13) puede ser visto como un

límite superior en el número de errores de entrenamiento, observando que un error de entrenamiento aquí se define como patrones donde las desigualdades en (12) se cumplen con $\xi_i > 0$. Minimizar (13) con restricciones (12) construye un hiperplano de separación que minimiza la suma de las desviaciones, ξ_i , de los errores de entrenamiento y maximiza el margen para las muestras correctamente clasificadas.

5 Puesto que se trata de un problema de optimización convexa con limitaciones lineales, la solución se puede conseguir resolviendo la formulación de problema dual equivalente. Mediante el establecimiento de las derivadas de la función de Lagrange con respecto a las variables w , b y ξ_i , a cero y sustituyendo de nuevo se obtiene el siguiente problema equivalente:

maximizar

10
$$\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (14)$$

sujeto a

$$\begin{aligned} \sum_i \alpha_i y_i &= 0 \\ 0 &\leq \alpha_i \leq C \end{aligned} \quad (15)$$

De la derivada del Lagrangiano con respecto a w se obtiene una expresión para calcular w en términos de los multiplicadores de Lagrange α_i :

15
$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i, \quad \text{for } \alpha_i > 0. \quad (16)$$

20 A partir de este problema se obtiene un resultado interesante para el caso en el que los datos son separables por un límite de decisión no lineal. Observando que la única forma en que los datos aparecen en esta formulación está en forma de los productos de punto $\mathbf{x}_i^T \mathbf{x}_j$, los datos podrían ser asignados a un espacio superior, y posiblemente infinito, usando una función de núcleo $K(\mathbf{x}_i, \mathbf{x}_j)$. Esta función del núcleo en el espacio de entrada es equivalente al producto de punto en ese espacio dimensional alto. La única restricción es que esta función del núcleo debe satisfacer las condiciones de Mercer. Al sustituir el producto de punto $\mathbf{x}_i^T \mathbf{x}_j$ por el núcleo $K(\mathbf{x}_i, \mathbf{x}_j)$ en todas partes del algoritmo de entrenamiento, el algoritmo encontrará un límite de decisión lineal en el espacio de alta dimensión que cuando se asigna nuevamente al espacio de entrada será no lineal dependiendo del núcleo. Más importante aún, esto se logra resolviendo el mismo problema de entrenamiento convexo. La función de decisión se obtiene entonces mediante:

25
$$f(\mathbf{x}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad \text{for } \alpha_i > 0. \quad (17)$$

30 Se apreciará que la SVM puede usar cualquier tipo de núcleo. Los núcleos de ejemplo incluyen núcleos lineales, núcleos polinómicos de 2º grado, núcleos polinómicos de 3º grado, núcleos polinómicos de 4º grado, núcleos RBF y similares. La aplicación de diferentes núcleos, y por lo tanto los límites de decisión, al problema de clasificar los espectros ESS de pólipos no displásicos y no displásicos es un problema de clasificación que no sólo es linealmente no separable sino también no linealmente no separable. De este modo, el núcleo de la función de base radial (RBF) puede ser ventajoso sobre el núcleo lineal y los núcleos polinomiales.

Máquinas de vector de apoyo con error-rechazo incorporado

35 Las realizaciones de la invención se dirigen a un problema de formación de SVM con rechazo de errores incorporado que siguen siendo un problema de optimización convexa. Esto permite retener la solución global presente en la formulación SVM original. Además, un problema de optimización convexa facilita el desarrollo de un algoritmo dual en el que no hay diferencia de dualidad.

El problema de entrenamiento para SVM con rechazo de errores incorporado implica encontrar una región de decisión definida por un par de hiperplanos paralelos

$$\mathbf{w}^T \mathbf{x} - b \pm \varepsilon = 0 \quad (18)$$

En el que se rechazan las muestras situadas entre ellas y, de acuerdo con el problema original de formación de SVM, las muestras correctamente clasificadas deben separarse con un margen máximo. En otras palabras, el clasificador tiene la siguiente función de decisión

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \mathbf{w}^T \mathbf{x} - b \geq \varepsilon \\ -1, & \text{if } \mathbf{w}^T \mathbf{x} - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \mathbf{w}^T \mathbf{x} - b < \varepsilon \end{cases} \quad (19)$$

- 5 donde w es la normal al hiperplano, $|b|/||w||$ es la distancia perpendicular del hiperplano al origen, y $2\varepsilon/||w||$ es la distancia entre los hiperplanos paralelos que definen la región de rechazo. Al restringir $0 \leq \varepsilon \leq 1$, la región de rechazo se encuentra dentro del margen $2/||w||$.

Se introduce un nuevo funcional, $h_c(\xi_i, \varepsilon)$, que se define como:

$$h_c(\xi_i, \varepsilon) = \max \begin{pmatrix} w_c \xi_i, \\ w_r \xi_i + (w_r - w_c) \varepsilon + w_c - w_r, \\ w_e \xi_i + (2w_r - w_c - w_e) \varepsilon + w_c - w_e \end{pmatrix} \quad (20)$$

- 10 con $0 < w_c < w_r < w_e$, siendo w_c el coste asignado a las muestras correctamente clasificadas pero con $0 < \xi_i \leq 1 - \varepsilon$, es decir, las muestras correctamente clasificadas situadas dentro del margen, w_r es el coste de rechazo asignado a las muestras que satisfacen $1 - \varepsilon \leq \xi_i \leq 1 + \varepsilon$, las muestras rechazadas que se encuentran dentro de la región de rechazo, y w_e es el coste de las muestras mal clasificadas que satisfacen $1 + \varepsilon < \xi_i \leq w_e$. $\xi_i \geq 0$ son las desviaciones de los errores de entrenamiento. La figura 7A muestra el comportamiento de la función $h_c(\xi_i, \varepsilon)$.

- 15 Para encontrar la región de decisión óptima que maximiza el margen mientras se contabiliza la compensación entre muestras de rechazo y clasificación errónea, se plantea el siguiente problema de formación para un SVM con rechazo de errores incorporado:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i, \varepsilon} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_i h_c(\xi_i, \varepsilon) \\ \text{s.t. } & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \\ & 0 \leq \varepsilon \leq 1 \end{aligned} \quad (21)$$

- 20 donde $y_i = \{-1, 1\}$ es la etiqueta de la muestra de entrenamiento x_i , y $C > 0$ es el coste asignado a los errores. Así, la solución de este problema construye dos hiperplanos de separación de anchura $2\varepsilon/||w||$ entre ellos, que minimiza la suma de las desviaciones, ξ_i , de los errores de entrenamiento ponderados por w_c , w_r , y w_e mientras que maximizamos el margen, $2/||w||$, para las muestras correctamente clasificadas. Además, la solución es global ya que se resuelve un problema de optimización convexa con restricciones lineales dado que la función $h_c(\xi_i, \varepsilon)$ es una función convexa.

- 25 El hecho de que $h_c(\xi_i, \varepsilon)$ sea una función convexa por partes permite la formulación del siguiente problema equivalente del presentado en (21):

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i, \varepsilon, z_i} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_i z_i \\ \text{s.t. } & y_i (\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i \\ & z_i \geq C w_c \xi_i \\ & z_i \geq C \left[w_r \xi_i + (w_r - w_c) \varepsilon + w_c - w_r \right] \\ & z_i \geq C \left[w_e \xi_i + (2w_r - w_c - w_e) \varepsilon + w_c - w_e \right] \\ & \xi_i \geq 0 \\ & 0 \leq \varepsilon \leq 1. \end{aligned} \quad (22)$$

5 El objetivo ahora es obtener la formulación doble del problema en (22) (formulación primal), permitiendo la construcción de límites de decisión no lineales usando la función del núcleo $K(x_i, x_i)$. Las regiones de rechazo no lineales se obtienen resolviendo el doble del problema presentado en (22). Primero, se derivaron las condiciones de Karush-Kuhn-Tucker (KKT) del problema (22). Éstas consisten en la factibilidad, optimización y condiciones de holgura complementarias. Las condiciones de factibilidad son las limitaciones establecidas en (22). El Lagrangiano está dado por:

$$\begin{aligned} L = & 1/2 \mathbf{w}^T \mathbf{w} + \sum_i z_i - \sum_i \alpha_i y_i \mathbf{w}^T \mathbf{x}_i + \sum_i \alpha_i y_i b + \sum_i \alpha_i - \sum_i \alpha_i \xi_i \\ & - \sum_i \beta c_i z_i + \sum_i \beta c_i C w_c \xi_i - \sum_i \beta r_i z_i + \sum_i \beta r_i C w_r \xi_i + \sum_i \beta r_i C (w_r - w_c) \varepsilon \\ & + \sum_i \beta r_i C (w_c - w_r) - \sum_i \beta e_i z_i + \sum_i \beta e_i C w_e \xi_i + \sum_i \beta e_i C (2w_r - w_c - w_e) \varepsilon \\ & + \sum_i \beta e_i C (w_c - w_e) - \sum_i \mu_i \xi_i. \end{aligned} \quad (23)$$

10 Diferenciando el Lagrangiano con respecto a las variables primitivas, w , b , ξ_i , ε , z_i , se obtienen las siguientes condiciones de optimización:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \quad (24)$$

$$\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0 \quad (25)$$

$$\frac{\partial L}{\partial \xi_i} = -\alpha_i + C w_c \beta c_i + C w_r \beta r_i + C w_e \beta e_i - \mu_i = 0 \quad (26)$$

15
$$\frac{\partial L}{\partial \varepsilon} = C (w_r - w_c) \sum_i \beta r_i + C (2w_r - w_c - w_e) \sum_i \beta e_i = 0 \quad (27)$$

$$\frac{\partial L}{\partial z_i} = 1 - \beta c_i - \beta r_i - \beta e_i = 0. \quad (28)$$

Finalmente, las condiciones de holgura complementarias son dadas por:

$$\alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i - b) - 1 + \xi_i] = 0 \quad (29)$$

$$\beta c_i [z_i - C w_c \xi_i] = 0 \quad (30)$$

20
$$\beta r_i [z_i - C (w_r \xi_i + (w_r - w_c) \varepsilon + w_c - w_r)] = 0 \quad (31)$$

$$\beta e_i [z_i - C (w_e \xi_i + (2w_r - w_c - w_e) \varepsilon + w_c - w_e)] = 0 \quad (32)$$

$$\mu_i \xi_i = 0. \quad (33)$$

25 Para formular el problema dual, el Lagrangiano (23) se maximiza con respecto a las variables duales, α_i , βc_i , βr_i , βe_i , μ_i , sujeto a las restricciones que el gradiente del Lagrangiano con respecto a las variables primitivas, w , b , ξ_i , ε , z_i , desaparece (Eq. (24) - (28)). Después de algunas manipulaciones se obtiene el siguiente problema dual:

$$\begin{aligned}
 & \max_{\alpha_i, \beta_c, \beta_r, \beta_e} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + C(w_c - w_r) \sum_i \beta_r + C(w_c - w_e) \sum_i \beta_e \\
 & \text{s.t.} \quad \sum_i \alpha_i y_i = 0 \\
 & \quad \beta_c + \beta_r + \beta_e = 1 \\
 & \quad C(w_r - w_c) \sum_i \beta_r + C(2w_r - w_c - w_e) \sum_i \beta_e = 0 \\
 & \quad 0 \leq \alpha_i \leq Cw_c \beta_c + Cw_r \beta_r + Cw_e \beta_e \\
 & \quad \beta_c, \beta_r, \beta_e \geq 0.
 \end{aligned} \tag{34}$$

5 El problema anterior es un problema de optimización cóncava y dado que el primal es convexo también, no existe brecha de dualidad para las soluciones óptimas primal y dual. Los datos sólo aparecen en el problema dual (34) en forma de producto de punto de punto $\mathbf{x}_i^T \mathbf{x}_j$, por lo que se obtendrán regiones de decisión no lineales si se utiliza un núcleo $K(\mathbf{x}_i, \mathbf{x}_j)$ en su lugar. La anchura de la región de rechazo está determinada por los pesos $0 < w_c < w_r < w_e$.

El valor de las variables primitivas w , b , ε de las variables duales óptimas se calcula para obtener la función de decisión

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases} . \tag{35}$$

El valor de w se obtiene de la condición de optimización en la ecuación (24) como sigue:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i, \quad \alpha_i > 0. \tag{36}$$

10

Observando que para una muestra \mathbf{x}_i cuya variable dual correspondiente α_i^* tiene valores $0 < \alpha_i^* < Cw_c$ la variable primal $\xi_i^* = 0$ ya que $\mu_i = Cw_c - \alpha_i > 0$, la condición de holgura complementaria en la ecuación (29) se utiliza para obtener el valor de b como sigue:

$$b = \sum_j \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i) - y_i, \quad 0 < \alpha_i < Cw_c \tag{37}$$

15 Es mejor numéricamente calcular b como la media de todos los puntos i que satisfacen la condición anterior. La anchura de la región de decisión, ε , se puede obtener de una manera similar observando que para $Cw_c < \alpha_i^* < Cw_r$, $\xi_i^* = 1 - \varepsilon$ y para $Cw_r < \alpha_i^* < Cw_e$, $\xi_i^* = 1 + \varepsilon$. otra vez usando la ecuación (29), ε se calcula como:

$$\varepsilon = y_i \left(\sum_j \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i) - b \right), \quad Cw_c < \alpha_i < Cw_r \tag{38}$$

$$\varepsilon = -y_i \left(\sum_j \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i) - b \right), \quad Cw_r < \alpha_i < Cw_e \tag{39}$$

20

Como con el cálculo de b , se obtiene ε con la media de todos los puntos i que satisfacen estas dos condiciones.

Una ventaja de formular el SVM con un algoritmo de entrenamiento de rechazo de errores incorporado como un

5 problema de optimización convexo es el hecho de que no existe un espacio de dualidad entre la solución primal y la solución dual. Uno de estos algoritmos de entrenamiento dual es la Optimización Mínima Secuencial (SMO). Este algoritmo resuelve el gran problema de optimización de la programación cuadrática requerido por el entrenamiento de SVM dividiéndolo en una serie de problemas más pequeños. El problema más pequeño posible implica dos multiplicadores de Lagrange. En cada paso, se eligen dos multiplicadores de Lagrange y se optimizan conjuntamente. La optimización se determina satisfaciendo las condiciones KKT. El SVM se actualiza para reflejar los valores óptimos. La principal ventaja de este algoritmo reside en que estos problemas más pequeños pueden resolverse analíticamente, evitando la optimización numérica de la programación cuadrática como un bucle interno.

10 Se describe un algoritmo de entrenamiento para SVMs con rechazo de errores basado en extensiones de SMO (SMOER), observando que este problema se plantea ahora como un problema de optimización convexa mediante la introducción de la funcional en (20). Esto es ventajoso porque las condiciones de KKT son suficientes y necesarias para la óptimización, además de una solución global y sin brecha de dualidad.

Se supone que ε se mantiene constante en (21). Entonces, el dual de (21) para un ε fijo se convierte en:

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \min_{\xi_i \geq 0} C \sum_i h_c(\xi_i, \varepsilon) - \alpha_i \xi_i \\ \text{s.t.} \quad & \sum_i \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C w_e. \end{aligned} \quad (40)$$

15 El término $g(\alpha_i, \varepsilon)$, donde

$$g(\alpha_i, \varepsilon) = \min_{\xi_i \geq 0} C \sum_i h_c(\xi_i, \varepsilon) - \alpha_i \xi_i, \quad (41)$$

es el conjugado cóncavo de $h_c(\xi_i, \varepsilon)$. Además, minimizar $g(\alpha_i, \varepsilon)$ con respecto a ξ_i da como resultado la expresión:

$$g(\alpha_i, \varepsilon) = \begin{cases} 0, & 0 \leq \alpha_i \leq C w_c \\ C w_c (1 - \varepsilon) - \alpha_i (1 - \varepsilon), & C w_c < \alpha_i \leq C w_r \\ C w_r (1 + \varepsilon) + C (w_c - w_r) (1 - \varepsilon) - \alpha_i (1 + \varepsilon), & C w_r < \alpha_i \leq C w_e \end{cases} \quad (42)$$

o similarmente

$$g(\alpha_i, \varepsilon) = \min \begin{bmatrix} 0, \\ C w_c (1 - \varepsilon) - \alpha_i (1 - \varepsilon), \\ C w_r (1 + \varepsilon) + C (w_c - w_r) (1 - \varepsilon) - \alpha_i (1 + \varepsilon), \end{bmatrix}. \quad (43)$$

20 La Figura 7B ilustra el comportamiento de $g(\alpha_i, \varepsilon)$. Finalmente, sustituyendo $g(\alpha_i, \varepsilon)$ en (40) se obtiene el siguiente problema dual:

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i g(\alpha_i, \varepsilon) \\ \text{s.t.} \quad & \sum_i \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C w_e \end{aligned} \quad (44)$$

25 El problema dual como se indica en (44) es el punto de partida para el desarrollo del algoritmo SMO para formar SVM con rechazo de errores incorporado. En cada paso, el SMO elige dos multiplicadores de Lagrange para optimizar conjuntamente, encuentra los valores óptimos para estos multiplicadores y actualiza el SVM para reflejar los nuevos valores óptimos. Esto es ventajoso porque la resolución de dos multiplicadores de Lagrange se puede hacer analíticamente, evitando cualquier optimización numérica QP. La optimización de las variables duales para el SVM con rechazo de errores incorporado incluye tres características: resolución analítica de los dos multiplicadores

30

de Lagrange, una heurística para elegir qué multiplicadores optimizar y un método para calcular el umbral b.

5 Como se ha descrito anteriormente, el SMO selecciona dos multiplicadores de Lagrange para optimizar analíticamente. Por simplicidad, todas las variables que se refieren al primer y segundo multiplicador tendrán el subíndice 1 y 2 respectivamente. La optimización de estos dos multiplicadores se realiza dentro de las restricciones de igualdad y limitación en el problema (44). La restricción enlazada hace que los multiplicadores se encuentren dentro de una caja, mientras que la restricción lineal hace que los multiplicadores se encuentren a lo largo de la línea diagonal,

$$\alpha_1 + s\alpha_2 = \gamma, \quad (45)$$

10 donde $s = y_1 y_2$ y $\gamma = \alpha_1^{old} + s\alpha_2^{old}$. Los extremos del segmento de línea diagonal cuando las etiquetas de α_1 y α_2 no son iguales ($y_1 \neq y_2$) se expresan como:

$$\begin{aligned} L &= \max(0, \alpha_2^{old} - \alpha_1^{old}) \\ H &= \min(Cw_e, Cw_e + \alpha_2^{old} - \alpha_1^{old}), \end{aligned} \quad (46)$$

y para el caso en el que las etiquetas de α_1 y α_2 son iguales ($y_1 = y_2$) como:

$$\begin{aligned} L &= \max(0, \alpha_1^{old} + \alpha_2^{old} - Cw_e) \\ H &= \min(Cw_e, \alpha_1^{old} + \alpha_2^{old}). \end{aligned} \quad (47)$$

15

20 El problema en (44) tiene que ser formulado en términos de α_1 y α_2 para resolver analíticamente estas dos variables duales. Utilizando la expresión en (45) de la restricción lineal y después de algunas manipulaciones se obtiene la siguiente función objetiva cóncava sin compresión:

$$\frac{1}{2}\eta\alpha_2^2 + \left(y_2(E_1^{old} - E_2^{old}) - \eta\alpha_2^{old}\right)\alpha_2 + g(\alpha_2, \epsilon) + g(\gamma - s\alpha_2, \epsilon) + const. \quad (48)$$

donde $E_i = f^{old}(x_i) - y_i$ es el error en la i -ésima muestra de entrenamiento. Diferenciarse con respecto a α_2 e igualar a cero resulta en

$$\alpha_2 = \alpha_2^{old} - \frac{1}{\eta} \left(y_2(E_1 - E_2) + \frac{\partial}{\partial \alpha_2} g(\alpha_2, \epsilon) - s \frac{\partial}{\partial \alpha_1} g(\alpha_1, \epsilon) \right), \quad (49)$$

25 usando el hecho de que

$$\frac{\partial}{\partial \alpha_2} g(\gamma - s\alpha_2, \epsilon) = -s \frac{\partial}{\partial \alpha_1} g(\alpha_1, \epsilon), \quad (50)$$

donde $\eta = 2K(x_1, x_2) - K(x_1, x_1) - K(x_2, x_2)$ y notando que la función $\frac{\partial}{\partial \alpha_i} g(\alpha_i, \epsilon)$ es la subdiferencial de $g(\alpha_i, \epsilon)$ y se define como:

$$\frac{\partial}{\partial \alpha_i} g(\alpha_i, \varepsilon) = \begin{cases} 0, & 0 \leq \alpha_i \leq Cw_c \\ -(1 - \varepsilon), & Cw_c < \alpha_i \leq Cw_r \\ -(1 + \varepsilon), & Cw_r < \alpha_i \leq Cw_e \end{cases} \quad (51)$$

La expresión en (49) no proporciona el máximo de (48); es el máximo de una de las nueve posibles parábolas correspondientes a todas las diferentes combinaciones entre los intervalos de $g(\alpha_2, \varepsilon_2)$ y $g(\gamma - s\alpha_2, \varepsilon_1)$ en (42). La optimización para α_1 y α_2 en cualquier paso dado implica encontrar el máximo de (48), que es una función cóncava por partes formada por el mínimo del conjunto compuesto de las parábolas para un α_2 dado, como se puede ver en la Figura 8.

Se describe un método para optimizar analíticamente las variables duales evitando evaluar la función $g(\alpha_i, \varepsilon)$. Como se ha descrito anteriormente, la optimización de un par de variables duales implica encontrar el máximo de la función cóncava por trozos en (48). Aprovechando la estructura del problema, se describe un procedimiento para obtener este máximo, evitando cualquier evaluación de la función y minimizando así los cálculos necesarios.

Las etapas de encontrar el máximo de (48) se muestran en la Figura 9. El objetivo es evaluar (48) sobre la secuencia de pares de intervalos de α_1 y α_2 , recortando el máximo en el límite correspondiente, hasta un máximo es

alcanzado. Cuando el proceso comienza los valores de $\frac{\partial}{\partial \alpha_2} g(\alpha_2, \varepsilon)$ y $\frac{\partial}{\partial \alpha_1} g(\alpha_1, \varepsilon)$ se determinan a partir de los valores iniciales de α_1 y α_2 (Etapa 1). Estos valores pueden ser 0, $-(1 - \varepsilon)$ y $-(1 + \varepsilon)$ para valores de α_i en los intervalos $0 \leq \alpha_i \leq Cw_c$, $Cw_c < \alpha_i \leq Cw_r$ y $Cw_r < \alpha_i \leq Cw_e$, respectivamente. Entonces, el máximo de (48), α_2^{new} , se

obtiene usando (49) para $\frac{\partial}{\partial \alpha_2} g(\alpha_2, \varepsilon)$ y $\frac{\partial}{\partial \alpha_1} g(\alpha_1, \varepsilon)$ como se determinó en la Etapa 1 y se recorta para satisfacer la restricción lineal de la ecuación 51 (Etapa 2 y Etapa 3). El valor correspondiente de α_1^{new} entonces se puede calcular usando $\alpha_1^{new} = \alpha_1 + s(\alpha_2 - \alpha_2^{new})$ (Etapa 4). Una vez α_2^{new} es calculado, la dirección de ascenso se obtiene mediante $\text{sgn}(\alpha_2^{new} - \alpha_2)$ (Etapa 5). Esto determinará el límite superior en el intervalo, por ejemplo para

el caso $Cw_c < \alpha_2^{new} \leq Cw_r$ donde el límite superior para α_2^{new} una dirección de ascenso positiva será Cw_r mientras que para una dirección de ascenso negativa el límite sería Cw_c . El siguiente paso (Etapa 6) es verificar que los valores de α_2^{new} y α_1^{new} están dentro del intervalo respectivo usado para determinar sus valores. El calculado

α_2^{new} es el máximo de la parábola producida por los valores de $\frac{\partial}{\partial \alpha_2} g(\alpha_2, \varepsilon)$ y $\frac{\partial}{\partial \alpha_1} g(\alpha_1, \varepsilon)$ para esos intervalos particulares, y puede o no puede estar dentro de ellos. En el caso de que estas condiciones se cumplan entonces

α_2^{new} es el máximo (48) y el proceso se detiene. De lo contrario, si están α_2^{new} y α_1^{new} fuera de su respectivo intervalo, estos valores se recortan al límite superior determinado por la dirección de ascenso. Si estas últimas circunstancias ocurren el valor de α_2^{new} y/o α_1^{new} correspondería a una esquina de (48). A continuación, se realiza la verificación de si esta esquina es el máximo; de lo contrario, se continúa el proceso de optimización (Etapa 7).

Para lograr esto, α_2^{next} , se calcula el máximo de la parábola producida por el siguiente intervalo, se recorta $L \leq \alpha_2^{next} \leq H$ de manera que se obtiene la nueva dirección de ascenso. Si la nueva dirección de ascenso cambia entonces el máximo de (48) está en esa esquina y el proceso se detiene. De lo contrario, el proceso continúa con el paso 6 hasta que se encuentre un máximo.

Como se ha descrito anteriormente, el algoritmo SMO optimiza dos multiplicadores de Lagrange en cada paso. Uno de estos multiplicadores ha violado anteriormente las condiciones de KKT antes del paso, es decir, SMO siempre alterará dos multiplicadores de Lagrange para moverse cuesta arriba en la función objetivo en (48), manteniendo al mismo tiempo un vector multiplicador de Lagrange viable. Así, en cada paso la función objetivo global aumentará hasta que el algoritmo converge asintóticamente.

Se usan dos heurísticas separadas para elegir el multiplicador primero y segundo. La elección de la primera heurística proporciona el bucle externo del algoritmo. El bucle exterior primero itera sobre todas las muestras en el conjunto de entrenamiento, determinando si una muestra viola las condiciones KKT. Una vez que se encuentra una muestra violada, se selecciona un segundo multiplicador usando la segunda heurística de elección y ambos se optimizan conjuntamente mientras se mantiene la factibilidad. El SVM se actualiza utilizando estos dos nuevos valores, y el bucle externo se reanuda buscando infractores KKT. Para acelerar el entrenamiento, el bucle externo

no siempre recorre todo el conjunto de entrenamiento. Después de un paso a través del conjunto de entrenamiento, el bucle exterior itera sobre muestras no unidas solamente. Estos se definen como multiplicadores cuyo valor no es ni 0 ni C. El bucle exterior itera sobre estas muestras hasta que ninguno de ellos viola las condiciones KKT. A continuación, el bucle exterior sobre todo el conjunto de entrenamiento de nuevo. Para el resto del algoritmo el bucle externo alterna entre estos dos hasta que el conjunto entero obedece las condiciones KKT, y en este punto el algoritmo termina.

La heurística de primera elección también se utiliza en el algoritmo SMOER. En SMO, el bucle exterior se concentra la mayor parte del tiempo en las muestras no enlazadas, que son las que más probablemente violan las condiciones KKT. A medida que progresa el algoritmo, los multiplicadores que están en los límites probablemente permanecerán allí, mientras que los multiplicadores no enlazados cambiarán a medida que otros se optimicen. En SMOER, el bucle externo inicialmente itera sobre todas las muestras, como en SMO, y luego itera sobre las muestras *no de esquina*. Las muestras que son no de esquina se definen como multiplicadores cuyos valores no son ni 0, Cw_c , Cw_r , Cw_e , y son las muestras más probables de cambiar a medida que avanza el algoritmo. Al igual que SMO, el bucle externo alterna entre iterar sobre todas las muestras y las muestras no de esquina hasta que se cumplen las condiciones de KKT para todas las muestras, concentrándose la mayor parte del tiempo en las muestras no de esquina.

Para seleccionar el segundo multiplicador para optimizar, SMOER utiliza la misma heurística que SMO. El objetivo de esta etapa es maximizar el tamaño de la etapa tomada durante la optimización de la articulación, que se aproxima por el valor absoluto de la diferencia de error en (49), $|E1 - E2|$. En la circunstancia inusual que no se puede hacer un progreso positivo usando la heurística descrita anteriormente, se utiliza una jerarquía de heurística de segunda elección hasta que se logre un progreso positivo cuando se optimiza conjuntamente un par de multiplicadores. En ese caso, entonces SMOER itera a través de las muestras no de esquina, comenzando en un punto aleatorio, buscando un segundo multiplicador que puede hacer un progreso positivo. Si ninguna de las muestras no de esquina realiza esto, SMOER itera a través de todo el conjunto de entrenamiento, comenzando de nuevo en un punto aleatorio, hasta que se encuentra una muestra que hace un progreso positivo. En el caso extremo en el que ninguna de las muestras hace progresos positivos cuando se emparejan con el primer multiplicador, se omite el primer multiplicador y SMOER continúa con otro primer multiplicador elegido.

Se describió anteriormente una serie de heurísticas para seleccionar un par de multiplicadores para optimizar a una iteración dada usando condiciones KKT. Las condiciones KKT también desempeñan un papel en la terminación del algoritmo, ya que todas las muestras deben satisfacer estas condiciones para que SMO y SMOER terminen.

La condición KKT se deriva basándose en los valores de los multiplicadores de Lagrange. Las condiciones siguientes son el resultado de la condición de optimización en (26) y la condición de holgura complementaria en (29) y (33) que resultan del problema de optimización en (22).

i. Si $\alpha_i = 0$, $\mu_i = Cw_c > 0$, $\xi_i = 0$;

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 \geq 0$$

ii. Si $0 < \alpha_i < Cw_c$, $\mu_i = Cw_c - \alpha_i > 0$, $\xi_i = 0$;

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 = 0$$

iii. Si $\alpha_i = Cw_c$, $\mu_i = 0$, $0 < \xi_i < 1 - \epsilon$;

$$\begin{aligned} y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 &= -\xi_i \\ -(1 - \epsilon) < y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 < 0. \end{aligned}$$

iv. Si $Cw_c < \alpha_i < Cw_r$, $\mu_i = 0$, $\xi_i = 1 - \epsilon$;

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 = -(1 - \epsilon).$$

v. Si $\alpha_i = Cw_r$, $\mu_i = 0$, $1 - \epsilon < \xi_i < 1 + \epsilon$;

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 = -\xi_i$$

$$-(1 + \varepsilon) < y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 < -(1 - \varepsilon).$$

vi. Si $Cw_r < \alpha_i < Cw_e$, $\mu_i = 0$, $\xi_i = 1 + \varepsilon$;

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 = -(1 + \varepsilon).$$

vii. Si $\alpha_i = Cw_e$, $\mu_i = 0$, $\xi_i > 1 + \varepsilon$;

5

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 = -\xi_i$$

$$y_i (\mathbf{w}^t \mathbf{x}_i - b) - 1 < -(1 + \varepsilon).$$

10

A medida que el algoritmo SMOER itera en el bucle exterior, la heurística de primera elección selecciona una muestra que ha violado las condiciones KKT como la primera en el posible par de multiplicadores que se optimizarán conjuntamente. El conjunto de infracciones KKT verificadas en este paso, y resultantes de las condiciones descritas anteriormente, se resumen en la Figura 10.

Resolver para los multiplicadores de Lagrange no determina el umbral b de la SVM y debe calcularse por separado. En SMO, b se recalcula después de cada paso, de modo que se cumplan las condiciones KKT para ambos ejemplos optimizados. El umbral se obtiene mediante:

$$b_i = E_2 + y_1 (a_1^{new} - a_1) k(\mathbf{x}_1, \mathbf{x}_i) + y_2 (a_2^{new} - a_2) k(\mathbf{x}_i, \mathbf{x}_2) + b^{old}, \quad (52)$$

15

El umbral b_i es válido cuando el nuevo α_i no está en los límites, es decir, $0 < \alpha_i < C$, porque obliga a la salida del SVM a ser y_i cuando la entrada es \mathbf{x}_i . Cuando ambos multiplicadores de Lagrange nuevos no están unidos, b_1 y b_2 son iguales. En el caso en que ambos nuevos multiplicadores están unidos, entonces el intervalo entre b_1 y b_2 son todos los umbrales que son consistentes con las condiciones KKT y SMO elige el umbral para estar a medio camino entre b_1 y b_2 .

20

El algoritmo SMOER calcula y actualiza el umbral b de una manera similar. La única modificación necesaria para adaptar la forma de calcular el umbral en SMO a SMOER es calculando b_i , usando (52), para multiplicadores cuyo valor está dentro de 0 y Cw_c . En (37) se vio que este intervalo permitía calcular el umbral al resolver el problema dual. Por lo tanto, en SMOER el umbral se actualiza sólo cuando los valores de los nuevos multiplicadores Lagrange están dentro de este rango. Como en SMO, cuando ambos nuevos multiplicadores tienen valores de 0 y/o Cw_c , el umbral actualizado es el valor a medio camino entre b_1 y b_2 .

25

El uso de la formulación convexa de SVM con rechazo de error incorporado resolviendo un subproblema con una muestra que viola las condiciones KKT mejora la función objetivo global y mantiene la viabilidad. SMO y SMOER convergen asintóticamente resolviendo una serie de estos problemas.

30

En cualquier iteración dada, el valor de b se actualiza basándose en el par actual de variables dobles que se optimizan. En el caso de SMOER esto ocurre es α_1 y/o α_2 están entre $0 \leq \alpha_i \leq Cw_c$, de lo contrario b no se actualiza. En SMO y SMOER, es posible que un α_i cuyos valores satisfagan las condiciones de KKT pueda considerarse erróneamente como violándolas, y viceversa. Este es el resultado de usar valores subóptimos de b para comprobar si un α_i viola o no las condiciones KKT. Por lo tanto, es posible seleccionar un par de variables duales para optimizar que no violen las condiciones KKT y el progreso no se realiza. En este caso, como resultado de la heurística de elección, SMO y SMOER continúan buscando un par de variables que hacen progresos.

35

El entrenamiento del SVM con rechazo de error embebido en (21) implica la optimización de las variables primitivas w , b , ξ_i , ε . Hasta este punto han sido capaces de desarrollar métodos para optimizar w , b , ξ_i para un valor fijo de ε . La variable w se optimiza cuando la solución a las variables duales, α_i , se encuentra con SMOER, mientras que el umbral b se obtiene durante ese proceso. La variable ξ_i se optimiza analíticamente al obtener la expresión en (42). La última pieza en el desarrollo de un algoritmo de entrenamiento para SVM con el rechazo de error incorporado es optimizar ε .

40

Recordando el problema indicado en (40), utilizando la expresión en (42), y suponiendo que α_i es constante, ε se obtienen resolviendo:

$$\min_{0 \leq \varepsilon \leq 1} \sum_i g(\alpha_i, \varepsilon). \quad (53)$$

Para resolver este problema, se utiliza un método de subgradiente. Este método utiliza la iteración

$$\varepsilon_{k+1} = \varepsilon_k - \lambda_k d_k \quad (54)$$

5 donde λ_k es el tamaño del paso en la k^{th} iteración y d_k es cualquier subgradiente de (53) en ε_k . El subdiferencial de la función objetivo en (53) es

$$\sum_i \frac{\partial g(\alpha_i, \varepsilon)}{\partial \varepsilon} \quad (55)$$

donde

$$\frac{\partial}{\partial \varepsilon} g(\alpha_i, \varepsilon) = \begin{cases} 0, & 0 \leq \alpha_i \leq Cw_c \\ -Cw_c + \alpha_i, & Cw_c < \alpha_i \leq Cw_r \\ C(2w_r - w_c) - \alpha_i, & Cw_r < \alpha_i \leq Cw_e \end{cases} \quad (56)$$

Se utiliza la siguiente regla de tamaño de paso que garantiza la convergencia del método:

$$\lambda_k = \frac{a |\varepsilon_k - \varepsilon_{k-1}|}{|d_k|}. \quad (57)$$

10

Para valores de $0 < a < 1$, esto puede ser visto como una regla de tamaño de paso decreciente. Además, al seleccionar $\varepsilon_{-1} = 0$ si $d_0/|d_0| > 0$ o $\varepsilon_{-1} = 1$ si $d_0/|d_0| < 0$ en el paso inicial, se aplica la restricción $0 \leq \varepsilon \leq 1$.

15

El SVM con algoritmo de rechazo de errores incorporado consiste entonces en una serie de iteraciones en las que se optimizan las variables duales α_i y, por lo tanto, las variables primitivas w , b , ξ_i , seguido de la optimización de las variables primitivas ε . El algoritmo comienza con una conjetura inicial de ε , proporcionada por el usuario. La rutina SMOER desarrollada se llama entonces para optimizar las variables duales con los valores iniciales de ε . A continuación, los valores de ε se actualizan utilizando el método de subgradiente y los valores previamente optimizados de α_i . El algoritmo mantiene la alternancia entre la optimización de α_i y ε , hasta la convergencia de ε .

20

Los criterios de parada considerados para la convergencia son:

$$|\varepsilon_k - \varepsilon_{k+1}| \leq tol, \quad (58)$$

es decir cuando el cambio en ε es menor que una tolerancia predeterminada.

25

Hay casos en los que puede ser de interés asignar a diferentes muestras de formación diferentes costes, no sólo a errores de clasificación sino también a rechazos, en función de la clase a la que pertenecen. Esto permite manejar aplicaciones en las que rechazar muestras de una clase en particular es más costosa que la otra. Por ejemplo, en el cribado del cáncer colorrectal, sería deseable tomar una decisión sobre los pólipos no displásicos con base en las mediciones del ESS, evitando al mismo tiempo rechazarlos tanto como sea posible (y, por tanto, tomar biopsias de los mismos). Por lo tanto, un mayor coste de rechazo podría ser asignado a estos casos en el proceso de formación. Este tipo de ponderación dependiente de la clase se puede incorporar en el problema de optimización convexa para el entrenamiento de máquinas de vectores de soporte con rechazo de errores incorporado mediante la reformulación del problema como:

30

$$\begin{aligned}
 \min_{\mathbf{w}, b, \xi_i, \varepsilon^+, \varepsilon^-} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C^+ \sum_{i^+} h_c(\xi_i, \varepsilon^+) + C^- \sum_{i^-} h_c(\xi_i, \varepsilon^-) \\
 \text{s.t.} & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \\
 & \xi_i \geq 0 \\
 & 0 \leq \varepsilon^+, \varepsilon^- \leq 1
 \end{aligned} \tag{59}$$

5 donde los superíndices +, - denotan muestras para las cuales las etiquetas $y_i = 1$ y $y_i = -1$ respectivamente. Así, la solución de este problema construye dos hiperplanos de separación de anchura $(\varepsilon^+ + \varepsilon^-)/\|\mathbf{w}\|$ entre ellos, que minimiza la suma de las desviaciones, ξ_i , de los errores de entrenamiento ponderados por w_c^+ , w_r^+ , w_e^+ y w_c^- , w_r^- , w_e^- en consecuencia. Al obtener el problema dual correspondiente

$$\begin{aligned}
 \max_{\alpha_i} & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \\
 \min_{\substack{\xi_i, \varepsilon^+, \varepsilon^- \\ \xi_i \geq 0 \\ 0 \leq \varepsilon^+, \varepsilon^- \leq 1}} & C^+ \sum_{i^+} h_c(\xi_i, \varepsilon^+) - \alpha_i \xi_i + C^- \sum_{i^-} h_c(\xi_i, \varepsilon^-) - \alpha_i \xi_i \\
 \text{s.t.} & \sum_i \alpha_i y_i = 0 \\
 & 0 \leq \alpha_i^{+,-} \leq C^{+,-} w_e^{+,-}
 \end{aligned} \tag{60}$$

10 También se pueden obtener regiones de decisión no lineales, así como el algoritmo de entrenamiento SMOER descrito anteriormente, para el entrenamiento.

Ejemplos

15 Se ensayaron las máquinas de vectores de soporte con rechazo de errores incorporado descrito anteriormente. El primer experimento aplicó el clasificador desarrollado a un dato artificial consistente en dos clases, y se ilustró visualmente cómo la orientación y el ancho de la región de rechazo varían para diferentes costes de rechazo, w_r . Cada clase consta de 100 muestras generadas a partir de una distribución normal bidimensional con media en [-1,0] para una clase y en [1,0] para la otra. Ambas clases tenían una covarianza igual a la matriz de identidad. Los resultados se muestran en la Figura 11 para los diferentes núcleos. En la Figura 11, los resultados para un núcleo lineal se muestran en la columna izquierda, un polinomio de tercer grado en la columna central y un núcleo RBF en la columna derecha. En la figura 11, el peso de rechazo w_r se reduce de la primera fila a la tercera fila.

20 En el segundo experimento, se usó el SVM con rechazo de errores incorporado para clasificar el conjunto de datos de pólipos de colon. La dimensionalidad de los datos se redujo a 15 características utilizando PCA. Los SVMs fueron entrenados con rechazo de errores incorporados con diferentes valores del coste de rechazo w_r y se generó un gráfico de sensibilidad/especificidad frente a tasa de rechazo. Para la comparación, los resultados obtenidos por el umbral de las salidas de un estándar SVM con núcleo lineal también se generaron en forma de la sensibilidad/especificidad frente a la tasa de rechazo parcela. Se utilizó la validación cruzada de K-fold, utilizando los mismos 50 muestreos aleatorios de los datos.

30 El experimento final ilustra los resultados obtenidos usando pesos de rechazo dependientes de clase w_r^-, w_r^+ . Para este ejemplo, de nuevo, la dimensionalidad de los datos se redujo a 15 características usando PCA. SVM con rechazo de error incorporado con núcleo RBF se utilizó. Los pesos de rechazo w_r^-, w_r^+ se variaron independientemente para diseñar clasificadores con regiones de rechazo de diferentes orientaciones y anchuras. La Figura 12 ilustra las estadísticas obtenidas cuando se modificaron los pesos de rechazo.

35 Los resultados obtenidos aplicando la formulación convexa desarrollada para SVM con rechazo de error incorporado muestran que la obtención de la orientación y anchura de la región de decisión durante el proceso de entrenamiento proporciona una mejora en el rendimiento sobre el umbral de las salidas de una SVM estándar. Los resultados obtenidos utilizando SVM con error de referencia incorporado con núcleo lineal son comparables o mejores que el umbral de las salidas de un SVM estándar con núcleo lineal. Adicionalmente, se puede obtener una mejora posterior usando funciones de núcleo para generar regiones de decisión no lineales en lugar de sólo un núcleo lineal. Mientras que el uso de los núcleos polinomio 2º y 3º grado no proporcionó mucha mejora, el uso del núcleo RBF había mejorado los resultados. Cuando se utiliza el núcleo RBF se obtiene una mayor sensibilidad y, en particular,

se obtienen especificidades para las mismas tasas de rechazo cuando se compara con los valores obtenidos mediante el umbral de las salidas de una SVM estándar. El uso de núcleos no lineales permite identificar mejor la región donde es más probable que se produzcan errores de clasificación, reduciendo así el número de muestras rechazadas que podrían clasificarse correctamente. Por último, el uso de pesos de rechazo dependientes de la clase permite diseñar clasificadores que asignen más coste a rechazar muestras de una clase. Puesto que los pesos de rechazo para cada clase se varían independientemente, las curvas de rechazo de errores o sensibilidad/especificidad son ahora superficies bidimensionales. Los clasificadores pueden diseñarse para diferentes sensibilidades y especificidades manteniendo la tasa de rechazo global constante variando el número de muestras rechazadas de cada clase.

10 Sistemas de clasificadores múltiples

Las realizaciones de la presente invención están dirigidas también a sistemas clasificadores múltiples. Estos sistemas de clasificación múltiple mejoran la sensibilidad y especificidad general del sistema. Un sistema 1300 clasificador múltiple de ejemplo es mostrado en la Figura 13. Como se muestra en la Figura 13, el sistema 1300 clasificador múltiple recibe datos de espectro ESS 1304 que incluyen diferentes regiones de datos 1308 de espectros. Un subconjunto 1312 de las características de patrón se identifica como se ha descrito anteriormente (por ejemplo, selección de características usando SFSS o SVDSS) para entrenamiento. El sistema 1300 incluye múltiples clasificadores 1316. Los múltiples clasificadores 1316 están diseñados para operar en paralelo. Cada clasificador 1316 es entrenado en una región del espectro ESS 1308. La clasificación o diagnóstico final 1324 se obtiene combinando las salidas de los clasificadores 1316 de base usando una regla de combinación 1320.

20 La regla de combinación 1320 fusiona las salidas de los clasificadores 1316. Dos reglas de combinación de ejemplo que pueden usarse son: el voto por mayoría y el combinador Bayes ingenuo. La regla de votación mayoritaria asigna una muestra a la etiqueta de clase en la que coinciden 50%+1 de los clasificadores base. La regla de combinación Bayes ingenua se expresa de la siguiente manera: sea x un vector de muestra perteneciente a una de las posibles clases ω_k , $k = 1, K, c$. También sea L el número de clasificadores base en el conjunto. Se dice que una muestra x pertenece a la clase ω_k si $\mu_k(x)$ es máxima, donde

$$\mu_k(\mathbf{x}) \propto P(\omega_k) \prod_{i=1}^L P(y_i | \omega_k), \quad (61)$$

y y_i son las etiquetas de salida de los clasificadores base. Para la aplicación práctica del problema de clasificación binaria, y suponiendo que si $y_i=1$ la muestra no es displásica y si $y_i=-1$ la muestra es displásica, las probabilidades condicionales pueden estimarse como:

$$\begin{aligned} P(y_i = -1 | \omega_k = -1) &= \text{Sensitivity}, \\ P(y_i = 1 | \omega_k = -1) &= 1 - \text{Sensitivity}, \\ P(y_i = 1 | \omega_k = 1) &= \text{Specificity}, \\ P(y_i = -1 | \omega_k = 1) &= 1 - \text{Specificity}. \end{aligned} \quad (62)$$

30 $P(\omega_k)$ son las probabilidades previas y se pueden estimar a partir de los datos de entrenamiento. Por lo tanto, esta regla de combinación utiliza el rendimiento de cada clasificador base, en forma de sensibilidad y especificidad, para ponderar su decisión.

35 En una realización, los clasificadores 1316 están diseñados con rechazo de errores también. Por ejemplo, los clasificadores 1316 pueden ser clasificadores SVM con rechazo de errores incorporados como se ha descrito anteriormente.

Ejemplo

40 La Tabla 2 muestra los resultados de clasificación de un clasificador diseñado usando todo el espectro, 330-760 nm, y un clasificador usando sólo la región 330-400 nm. Las tasas de error para ambos clasificadores fue similar. Sin embargo, los errores no se cometieron necesariamente en las mismas muestras, como se ilustra por la diferencia de sensibilidad y especificidad.

Tabla 2: Resultados de clasificación usando el espectro completo (330-760 nm) y usando la región 330-400 nm. PCA y SVM con Núcleo Lineal.

Región	# de características	Sensibilidad	Especificidad	Tasa de error
330-760 nm	15	.7417 ± .0757	.7538 ± .0528	.2499 ± .0364
330-400 nm	10	.7977 ± .0732	.7300 ± .0397	.2490 ± .0276

5 Las regiones espectrales se seleccionaron *a priori* de una manera heurística. Las regiones se eligieron en base a las áreas del espectro que se mostraron más informativas usando métodos de selección de características descritos anteriormente y basándose en donde había la mayor diferencia en el espectro promedio para las muestras displásicas y no displásicas. Algunas regiones fueron elegidas para asegurarse de que todas las áreas del espectro fueron consideradas. Las regiones seleccionadas abarcaban todo el espectro y se permitía la superposición entre ellas. La intención principal era promover la diversidad centrándose en diferentes áreas del espectro de ESS. En total, se consideraron veintidós regiones. Las características de cada región se obtuvieron utilizando PCA y el clasificador base fue un SVM. Se utilizaron dos tipos de núcleos: los núcleos lineales y RBF. Todos los resultados se obtuvieron mediante la validación cruzada de K-veces con los 50 conjuntos muestreados al azar descritos anteriormente.

15 La Tabla 3 resume el rendimiento de cada clasificador individual en cada una de las regiones y muestra el número de características extraídas para cada región. Aunque no se muestran, la desviación estándar para la sensibilidad, especificidad y tasa de error fue de alrededor de 0.08, 0.06 y 0.04 respectivamente.

Tabla 3: Rendimiento de los clasificadores SVM en cada una de las regiones espectrales consideradas.

Índice	Región	Núcleo lineal				Núcleo RBF			
		PC's	Se	Sp	Tasa de error	PC's	Se	Sp	Tasa de error
1	330-760 nm	15	.7417	.7538	.2499	15	.7463	.7731	.2365
2	330-800 nm	15	.7183	.7546	.2566	15	.7206	.7736	.2428
3	330-600 nm	12	.7697	.7392	.2513	16	.7629	.7628	.2372
4	330-380 nm	5	.7731	.6626	.3032	10	.8000	.6508	.3030
5	330-360 nm	8	.7903	.6469	.3087	8	.8000	.6423	.3088
6	330-400 nm	10	.7977	.7300	.2490	10	.7914	.7303	.2508
7	360-400 nm	6	.7966	.7156	.2593	6	.8046	.7085	.2618
8	530-600 nm	7	.6960	.6913	.3073	7	.6966	.6938	.3053
9	460-590 nm	14	.6994	.6826	.3122	14	.6966	.6841	.3120
10	360-600 nm	12	.7691	.7305	.2575	12	.7737	.7433	.2535
11	360-760 nm	14	.6989	.7621	.2575	16	.7206	.7685	.2464
12	360-800 nm	13	.6914	.7721	.2529	13	.6920	.7726	.2524
13	500-600 nm	9	.6794	.7003	.3062	9	.6863	.7105	.2970
14	590-760 nm	12	.6160	.7131	.3170	11	.6166	.7174	.3138
15	590-700nm	9	.6211	.7462	.2926	9	.6240	.7487	.2899
16	590-800 nm	15	.6463	.7072	.3117	14	.6423	.7172	.3060
17	700-760 nm	6	.5046	.6385	.4030	6	.4817	.6636	.3927
18	700-800 nm	7	.5903	.6690	.3554	7	.5909	.6708	.3540
19	460-700 nm	14	.6577	.7110	.3055	17	.6594	.7274	.2936
20	500-760 nm	7	.6274	.6946	.3262	15	.6377	.7162	.3081
21	500-700 nm	16	.6794	.7438	.2761	16	.6829	.7449	.2743
22	400-760 nm	11	.6954	.7321	.2793	12	.7000	.7323	.2777

20 En primer lugar, se seleccionaron todas las combinaciones de tamaño 3, 5, 7, 9 de los veintidós clasificadores posibles. Para cada combinación, se utilizó la regla de combinación de votación mayoritaria para obtener la decisión final. Para cada tamaño (3, 5, 7, 9), se seleccionó la combinación de regiones que resultaron en la mejor tasa de reconocimiento promedio (ARR, promedio entre sensibilidad y especificidad), tasa de error, sensibilidad y especificidad. Los resultados se muestran en la Tabla 4. Esto se repitió con el combinador Bayes ingenuo, obteniéndose los resultados en la Tabla 5.

25

Tabla 4: Resumen de resultados con la regla de combinación de votación mayoritaria.

L	Núcleo Lineal				Núcleo RBF			
	ARR	Error	Sen	Spe	ARR	Error	Sen	Spe

3	Reg	3,7, 21	3,7, 15	3,5,6	1,7,15	3,7,21	1,7,21	1,5,6	2,11,18
	Sen	.7966	.7806	.8234	.7623	.7994	.7794	.8251	.7091
	Spe	.7751	.7846	.7115	.7897	.7721	.7846	.7233	.7944
	Error	.2182	.2166	.2538	.2188	.2195	.2170	.2451	.2320
5	Reg	1,3,7, 15,21	1,3,7, 15,21	3,4,5, 6,7	3,7,12, 15,21	1,3,5, 6,21	1,3,7, 18,21	1,4,5, 6,10	1,2,7, 15,18
	Sen	.7760	.7760	.8183	.7480	.8091	.7606	.8251	.7417
	Spe	.7921	.7921	.7044	.7967	.7644	.7990	.7069	.8021
	Error	.2129	.2129	.2604	.2184	.2218	.2129	.2565	.2166
7	Reg	1,3,6, 7,15, 19,21	1,3,6, 7,15, 19,21	3,4,5, 6,7, 8,10	3,7,11, 12,15, 18,21	1,3,7, 10,13, 15,21	1,2,7, 11,15, 18,21	1,3,4, 5,6, 7,9	1,2,7, 11,15, 17,18
	Sen	.7806	.7806	.8171	.7349	.7869	.7480	.8223	.7200
	Spe	.7879	.7879	.7146	.7954	.7790	.8031	.7215	.8044
	Error	.2143	.2143	.2536	.2234	.2186	.2140	.2473	.2218
9	Reg	3,4,5,6,7, 10,15, 21,22	1,3,6, 7,10, 15,18, 19, 21	3,4,5, 6,7, 8,10, 11,21	1,2,3, 7,12, 15,17, 18,21	1,3,5, 6,10, 16,19, 21,22	1,2,6, 7,11, 13,15, 18,21	1,3,4, 5,6, 7,8, 9,10	1,2,3, 7,11, 15,17, 18,21
	Sen	.8046	.7703	.8114	.7217	.7851	.7606	.8103	.7314
	Spe	.7551	.7885	.7413	.7954	.7777	.7959	.7297	.8028
	Error	.2296	.2172	.2370	.2274	.2200	.2150	.2453	.2193

Tabla 5: Resumen de los resultados con la regla de combinación de Bayes ingenua.

L		Lineal Núcleo				Núcleo RBF			
		ARR	Error	Sen	Spe	ARR	Error	Sen	Spe
3	Reg	3,7,21	1,7,17	6,7,15	3,17,21	1,7,21	1,7,71	6,7,13	13,17,18
	Sen	.7480	.6720	.7686	.5920	.7417	.6823	.7737	.4903
	Spe	.8200	.8662	.7697	.8767	.8279	.8597	.7682	.8813
	Error	.2023	.1940	.2306	.2115	.1988	.1952	.2301	.2398
5	Reg	3,5,7, 21,22	3,7,12, 17,21	3,5,6, 7,13	1,7,17, 18,21	1,7,10, 17,21	1, 7,10 17,21	1,3,5, 6,10	1,7,17, 18,19
	Sen	.7577	.7166	.7771	.6869	.7526	.7526	.7783	.6863
	Spe	.8136	.8397	.7759	.8495	.8262	.8262	.7944	.8487
	Error	.2037	.1984	.2237	.2009	.1966	.1966	.2106	.2016
7	Reg	3,4,6,7,8, 21,22	3,6,7, 12,17, 18,21	3, 5,6, 7, 8, 10,21	3,7,9, 15,17, 18,21	1,3,5, 6,7, 21,22	1,3,7, 17,18, 19,21	1,2,4, 5,6, 7,17	2,3,7, 15,17, 18,21
	Sen	.7800	.7371	.7823	.6960	.7766	.7331	.7834	.7240
	Spe	.7936	.8290	.7859	.8323	.8013	.8315	.7674	.8341
	Error	.2106	.1995	.2152	.2099	.2064	.1989	.2276	.2000
9	Reg	1,3,5, 6,7,10, 15,19,21	3,6,7, 12,13, 15,17, 18, 21	3,4,5, 6,7, 8,10, 20,22	3,6,7, 12,15, 16,17, 18,21	1,2,3, 5,6, 7,15, 17,21	1,2,3, 7,13, 15,17, 18,21	1,3,4, 5,6, 7,10, 17,21	3,6,7, 12,15, 16,17, 18,21
	Sen	.7737	.7411	.7834	.7320	.7709	.7526	.7846	.7320
	Spe	.7951	.8215	.7718	.8228	.8059	.8215	.7818	.8264
	Error	.2115	.2034	.2246	.2053	.2050	.1998	.2173	.2028

5

10

Como se observa en la Tabla 3, se varió el rendimiento de los clasificadores diseñados en cada una de las regiones consideradas, con algunos como 3 (330-600 nm), 6 (330-400 nm), 7 (360-400 nm) y 10 (360-600 nm) dando resultados particularmente buenos. Esta eficacia variada fomenta su uso en clasificadores de conjuntos. A partir de los resultados, las combinaciones de diferentes regiones proporcionan un aumento en el rendimiento de clasificación cuando se compara con un solo clasificador. Ambas reglas de combinación evaluadas lograron mejores resultados, con el combinador ingenuo de Bayes aparentemente funcionando mejor.

Método de diagnóstico óptico ESS con biopsia

5 Las realizaciones de la presente invención también se dirigen a un método para tomar decisiones o diagnósticos basados en mediciones de ESS. El método comienza con un paso de reducción de la dimensionalidad para manejar las mediciones dimensionales altas generadas por ESS. Este paso se centra en áreas diferentes, y a menudo más pequeñas, del espectro de ESS. El método continúa entrenando cada uno de los múltiples clasificadores en una de las áreas más pequeñas del espectro de ESS. En una realización, los múltiples clasificadores son entrenados con rechazo de errores incorporados. El método continúa, durante una fase de diagnóstico, clasificando datos de espectros para una muestra de tejido por cada uno de los múltiples clasificadores. Cada clasificador analiza los datos de los espectros para la muestra en la que se formó el clasificador. En las realizaciones en las que los múltiples clasificadores son entrenados con rechazo de errores incorporados, la decisión del clasificador puede ser positiva, rechazada o negativa $\{+1, 0, -1\}$; de lo contrario, la decisión será positiva o negativa $\{+1, 0, -1\}$. A continuación, el método continúa combinando la clasificación de cada uno de los clasificadores para determinar un diagnóstico final de la muestra de tejido.

15 Con la inclusión de la opción de rechazo en los clasificadores de base, ya que la salida de estos clasificadores es $\{-1, 0, 1\}$, se requiere una modificación de las reglas de combinación. La combinación de las decisiones de los clasificadores de base es ahora un proceso de dos pasos: primero, se decide si la muestra es rechazada; en segundo lugar, si la muestra no se rechaza, se toma una decisión basada en los clasificadores que asignaron una etiqueta, es decir, no rechazaron la muestra. Se puede utilizar una regla de mayoría para determinar si rechazar o no una muestra. Una muestra se rechaza si el 50%+1 de los clasificadores de base se omiten al asignar una etiqueta. Para las muestras que no se rechazan, puede aplicarse cualquier regla de combinación dada a aquellos clasificadores que no se negaron a tomar una decisión. Una vez más, se pueden usar las reglas de combinación Bayes de votación mayoritaria e ingenuas.

Ejemplo

25 Todas las regiones espectrales de ESS usadas anteriormente se usaron de nuevo en este ejemplo. PCA se utilizó para reducir la dimensionalidad y el número de características utilizadas es el mismo que en la sección anterior (resumido en la Tabla 3). SVMs con rechazo de errores incorporado, con núcleos lineales y RBF, fueron entrenados en cada región para diferentes tasas de rechazo al variar el coste de rechazo. Para construir el sistema clasificador múltiple, se seleccionaron clasificadores a una tasa de rechazo particular de cada una de las regiones (específicamente, se usaron tasas de rechazo de aproximadamente 0.33 y 0.50) para los clasificadores base. En las Tablas 6 y 7 se resume el desempeño para los clasificadores en cada región, para la tasa de rechazo de 0.33 y 0.50 respectivamente.

35 Se utilizó la validación cruzada de K-fold con 50 conjuntos muestreados al azar. La combinación de regiones que produjeron la mejor tasa de reconocimiento media (ARR, promedio entre sensibilidad y especificidad), tasa de error, la sensibilidad y especificidad de todos los posibles conjuntos de tamaño $L = \{3, 5, 7, 9\}$ que puede ser creado se utilizaron. La desviación estándar para las sensibilidades, especificidades, las tasas de error y las tasas de rechazo oscilaba alrededor de 0.08, 0.06, 0.04 y 0.05, respectivamente. Los resultados se resumen en las Tablas 8 y 9 para una tasa de rechazo base clasificador de aproximadamente 0.33 recurriendo al voto por mayoría y combinador ingenuo de Bayes, respectivamente. Para clasificadores base con la tasa de rechazo de alrededor de 0.50 Los resultados se resumen en la Tabla 10, la votación por mayoría, y en la Tabla 11, para combinador de Bayes ingenuo.

Tabla 6: Rendimiento de los clasificadores con una tasa de rechazo de 0.33 en cada una de las regiones espectrales consideradas.

Índice	Región	Núcleo lineal				Núcleo RBF			
		Sen	Spe	Error	Rej	Sen	Spe	Error	Rej
1	330-760 nm	.7891	.8282	.1823	.3046	.7990	.8526	.1624	.3248
2	330-800 nm	.7668	.8287	.1883	.3120	.7830	.8542	.1668	.3407
3	330-600 nm	.8277	.8222	.1747	.3193	.8317	.8401	.1618	.3207
4	330-380 nm	.8142	.7151	.2561	.2927	.8004	.7472	.2379	.3301

ES 2 611 207 T3

5	330-360 nm	.8292	.7265	.2442	.3531	.8142	.7333	.2438	.3335
6	330-400 nm	.8605	.7923	.1879	.2915	.8633	.8007	.1806	.3370
7	360-400 nm	.8633	.7735	.1990	.3312	.8587	.7773	.1976	.3177
8	530-600 nm	.7487	.7464	.2518	.3465	.7400	.7479	.2541	.3322
9	460-590 nm	.7381	.7435	.2582	.2975	.7388	.7534	.2508	.3248
10	360-600 nm	.8389	.8325	.1648	.3634	.8322	.8298	.1693	.3262
11	360-760 nm	.7712	.8148	.1977	.3358	.7971	.8365	.1752	.3372
12	360-800 nm	.7585	.8234	.1957	.3519	.7530	.8206	.1993	.3287
13	500-600 nm	.7427	.7519	.2500	.3283	.7380	.7566	.2482	.3352
14	590-760 nm	.6619	.7653	.2682	.3039	.6733	.7674	.2632	.3400
15	590-700 nm	.6983	.7937	.2364	.3205	.7061	.7974	.2313	.3416
16	590-800 nm	.7042	.7838	.2407	.3605	.7093	.7762	.2445	.3281
17	700-760 nm	.5530	.6470	.3828	.2888	.5709	.6389	.3810	.3299
18	700-800 nm	.6162	.6955	.3294	.3676	.6257	.6779	.3392	.3184
19	460-700 nm	.7135	.7580	.2559	.3122	.7035	.7882	.2386	.3150
20	500-760 nm	.6967	.7535	.2637	.3412	.7105	.7676	.2497	.3411
21	500-700 nm	.7176	.8153	.2156	.3644	.7156	.8107	.2196	.3432
22	400-760 nm	.7777	.7983	.2067	.3681	.7663	.7935	.2138	.3384

Tabla 7: Rendimiento de los clasificadores con una tasa de rechazo de 50 en cada una de las regiones espectrales consideradas.

Índice	Región	Núcleo lineal				Núcleo RBF			
		Sen	Spe	Error	Rej	Sen	Spe	Error	Rej
1	330-760 nm	.8157	.8440	.1618	.4772	.8150	.8862	.1334	.5131
2	330-800 nm	.7996	.8575	.1571	.5000	.8052	.8822	.1394	.5306
3	330-600 nm	.8261	.8498	.1541	.5002	.8449	.8666	.1381	.4940
4	330-380 nm	.8296	.7380	.2337	.4584	.8129	.7699	.2167	.5019
5	330-360 nm	.8600	.7455	.2210	.5370	.8473	.7580	.2157	.5041
6	330-400 nm	.8726	.8279	.1575	.4846	.8750	.8258	.1584	.4885
7	360-400 nm	.8803	.8111	.1675	.5159	.8899	.7890	.1790	.4919
8	530-600 nm	.7646	.7736	.2288	.5310	.7626	.7643	.2353	.4894
9	460-590 nm	.7609	.7721	.2305	.4669	.7692	.7805	.2227	.4837
10	360-600 nm	.8697	.8533	.1401	.4558	.8812	.8665	.1265	.5260
11	360-760 nm	.8080	.8477	.1632	.5333	.8248	.8625	.1472	.4749
12	360-800 nm	.7993	.8450	.1667	.5533	.7846	.8470	.1701	.5193
13	500-600 nm	.7587	.7926	.2184	.5030	.7624	.7863	.2209	.4927
14	590-760 nm	.7006	.7883	.2392	.4828	.7054	.8052	.2255	.5023
15	590-700 nm	.7372	.8266	.2019	.4634	.7587	.8437	.1830	.5152
16	590-800 nm	.7142	.7976	.2274	.4533	.7391	.8128	.2089	.5145

17	700-760 nm	.5326	.7092	.3485	.4805	.5494	.7073	.3435	.5023
18	700-800 nm	.6130	.7167	.3173	.4646	.6313	.7044	.3191	.4901
19	460-700 nm	.7445	.7833	.2268	.4779	.7537	.8356	.1895	.5129
20	500-760 nm	.7287	.8002	.2218	.5204	.7415	.8042	.2151	.5067
21	500-700 nm	.7488	.8531	.1791	.5453	.7415	.8447	.1884	.5150
22	400-760 nm	.7906	.8150	.1910	.4487	.7943	.8280	.1809	.5104

Tabla 8: Resumen de los resultados con la regla de combinación de votación mayoritaria y los clasificadores base con una tasa de rechazo de 0.33.

L		Lineal Núcleo				Núcleo RBF			
		ARR	Error	Sen	Spe	ARR	Error	Sen	Spe
3	Reg	3,7,10	6,10,21	6,7,10	1,10,12	3,7,10	3,6,21	6,7,10	1,2,11
	Sen	.8738	.8486	.8831	.8027	.8629	.8400	.8832	.7995
	Spe	.8345	.8459	.8007	.8527	.8360	.8507	.8052	.8593
	Error	.1528	.1526	.1750	.1615	.1560	.1520	.1716	.1576
	Rej	.3308	.2995	.3156	.3315	.3152	.2945	.3177	.3338
5	Reg	3,5,7, 10,21	3,7,10, 16,21	3,5,6, 7,10	3,7,10, 16,21	3,6,7, 10,19	2,3,7, 19,21	3,6,7, 10,19	2,6,11, 19,21
	Sen	.8520	.8301	.8645	.8301	.8722	.8268	.8722	.8004
	Spe	.8392	.8582	.8085	.8582	.8250	.8585	.8250	.8615
	Error	.1559	.1505	.1747	.1505	.1609	.1511	.1609	.1571
	Rej	.3081	.3041	.2954	.3041	.2894	.2828	.2894	.2952
7	Reg	3,5,6, 7,10, 16,21	1,3,5, 7,10, 16,21	3,4,5, 6,7, 10,18	3,5,7, 10,16, 20,21	1,2,3, 6,7, 10,22	1,2,3, 6,7, 10,22	1,3,5, 6,7, 10,19	1,2,6, 11,16, 19,21
	Sen	.8555	.8356	.8609	.8101	.8414	.8414	.8580	.8015
	Spe	.8387	.8493	.7960	.8565	.8523	.8523	.8327	.8600
	Error	.1562	.1546	.1845	.1575	.1503	.1503	.1595	.1576
	Rej	.2869	.2963	.2685	.3002	.3019	.3019	.2802	.2890
9	Reg	1,3,5, 6,7, 10,13, 16,21	1,3,5, 6,7, 10,13, 16,21	1,3,4, 5,6, 7,10, 13,16	2,3,5, 7,8, 10,16, 18,21	1,3,5, 6,7, 10,13, 16,19	1,3,5, 6,7, 10,13, 16,19	3,4,5, 6,7, 8,10, 11,16	3,4,7, 11,13, 15,16, 17,21
	Sen	.8440	.8440	.8512	.8172	.8467	.8467	.8530	.7852
	Spe	.8437	.8437	.8207	.8526	.8466	.8466	.8307	.8556
	Error	.1557	.1557	.1698	.1580	.1534	.1534	.1633	.1661
	Rej	.2841	.2841	.2738	.2841	.2687	.2687	.2775	.2628

5

Tabla 9: Resumen de los resultados con la regla de combinación de Bayes ingenua y clasificadores de base con una tasa de rechazo de 0.33.

L		Lineal Núcleo				Núcleo RBF			
		ARR	Error	Sen	Spe	ARR	Error	Sen	Spe
3	Reg	3,7,10	3,7,22	6,7,10	2,7,21	3,7,10	3,7,16	6,7,10	1,7,21
	Sen	.8644	.8308	.8710	.7724	.8569	.8267	.8706	.7903
	Spe	.8576	.8768	.8204	.8872	.8514	.8797	.8221	.8862
	Error	.1399	.1372	.1646	.1483	.1470	.1368	.1637	.1433
	Rej	.3308	.3181	.3165	.2959	.3152	.2823	.3177	.2938
5	Reg	3,6,7, 10,22	3,7,10, 16,21	3,6,7, 10,22	3,7,16, 18,21	1,3,6, 7,22	2,3,7, 19,21	3,6,7, 10,19	3,7,16, 17,21
	Sen	.8607	.8249	.8607	.7659	.8512	.8158	.8668	.7769
	Spe	.8402	.8673	.8402	.8753	.8535	.8723	.8358	.8755
	Error	.1536	.1462	.1536	.1597	.1466	.1452	.1550	.1554
	Rej	.3048	.3041	.3048	.2949	.2945	.2828	.2899	.2550
7	Reg	3,5,6, 7,10, 16,21	3,6,7, 10,16, 18,21	3,5,6, 7,8, 10,22	3,6,7, 10,16, 18,21	1,2,3, 6,7, 10,22	1,2,3, 6,7, 10,22	3,5,6, 7,10, 13,19	2,3,7, 13,16, 18,21
	Sen	.8477	.8260	.8529	.8260	.8417	.8417	.8510	.7900
	Spe	.8556	.8674	.8394	.8674	.8590	.8590	.8398	.8700

	Error	.1472	.1455	.1563	.1455	.1458	.1458	.1569	.1548
	Rej	.2869	.2789	.2972	.2789	.3019	.3019	.2759	.2627
9	Reg	1,3,5, 6,7, 8,10, 16,21	3,5,6, 7,10, 16,18, 21,22	1,3,4, 5,6, 7,10, 13,16	2,3,6, 7,10, 15,16, 18,21	1,3,5, 6,7, 10,13, 16,22	1,2,3, 6,7, 11,13, 16,22	1,3,4, 5,6, 7,8, 10,16	2,3,5, 7,13, 15,16, 17,21
	Sen	.8393	.8284	.8483	.8140	.8455	.8316	.8482	.7896
	Spe	.8554	.8616	.8328	.8650	.8532	.8613	.8382	.8679
	Error	.1494	.1487	.1626	.1508	.1493	.1475	.1588	.1565
	Rej	.2848	.2743	.2738	.2712	.2729	.2862	.2761	.2572

Tabla 10: Resumen de resultados con regla de combinación de votación mayoritaria y clasificadores de base con tasa de rechazo de 0.50.

L		Linear Núcleo				Núcleo RBF			
		ARR	Error	Sen	Spe	ARR	Error	Sen	Spe
3	Reg	3,7,10	6,10,21	6,7,10	3,10,21	1,7,10	1,10,16	6,7,10	1,10,21
	Sen	.8892	.8702	.9001	.8564	.8983	.8651	.9043	.8612
	Spe	.8661	.8807	.8440	.8830	.8764	.8984	.8442	.8988
	Error	.1256	.1216	.1391	.1234	.1154	.1106	.1368	.1115
	Rej	.4949	.4818	.5023	.4950	.5219	.5237	.5127	.5196
5	Reg	3,6,7, 16,21	3,7,10, 16,21	3,6,7, 10,14	3,7,10, 16,21	1,3,7, 10,16	1,3,7, 10,16	1,6,7, 10,14	1,2,9, 10,16
	Sen	.8764	.8625	.8979	.8625	.8888	.8888	.9024	.8443
	Spe	.8851	.8935	.8635	.8935	.8874	.8874	.8650	.9009
	Error	.1165	.1148	.1245	.1148	.1109	.1109	.1216	.1156
	Rej	.5087	.4996	.4919	.4996	.5145	.5145	.5133	.5296
7	Reg	3,5,6, 7,10, 16,21	3,6,7, 10,11, 16,21	1,5,6, 7,8, 10,14	2,3,7, 10,16, 20,21	1,3,5, 7,10, 11,16	1,3,5, 7,10, 11,16	3,5,6, 7,10, 11,14	1,2,3, 9,10, 16,21
	Sen	.8873	.8780	.8926	.8556	.8942	.8942	.8995	.8415
	Spe	.8847	.8919	.8574	.8936	.8896	.8896	.8639	.9032
	Error	.1136	.1110	.1310	.1171	.1075	.1075	.1237	.1146
	Rej	.5034	.5048	.4970	.5041	.5117	.5117	.5067	.5228
9	Reg	2,3,5, 6,7, 10,16, 20,21	2,3,5, 6,7, 10,16, 20,21	3,4,5, 6,7, 10,11, 16,19	1,2,3, 6,7, 10,16, 20,21	1,3,5, 6,7, 10,11, 12,16	1,3,5, 6,7, 10,11, 16,21	3,5,6, 7,8, 10,11, 14,19	1,2,6, 9,10, 11,16, 21,22
	Sen	.8679	.8679	.8819	.8556	.8903	.8870	.8904	.8439
	Spe	.8887	.8887	.8616	.8918	.8884	.8903	.8692	.9034
	Error	.1386	.1386	.1308	.1181	.1094	.1088	.1223	.1141
	Rej	.5085	.5085	.4968	.5005	.5136	.5073	.5005	.5147

Tabla 11: Resumen de los resultados con la regla de combinación Bayes ingenua y clasificadores de base con una tasa de rechazo de 0.50.

L		Linear Núcleo				Núcleo RBF			
		ARR	Error	Sen	Spe	ARR	Error	Sen	Spe
3	Reg	1,7,10	6,10,21	1,7,10	6,11,21	2,7,10	1,10,16	2,7,10	1,10,16
	Sen	.8657	.8657	.8953	.8172	.8973	.8585	.9016	.8585
	Spe	.8902	.8902	.8488	.9027	.8817	.9064	.8479	.9064
	Error	.1160	.1160	.1372	.1242	.1122	.1073	.1351	.1073
	Rej	.5273	.5273	.5023	.5227	.5219	.5237	.5127	.5237
5	Reg	3,6,7, 11,16	3,6,7, 11,16	3,6,7, 10,14	3,7,17, 18,21	1,3,7, 10,16	1,6,10, 16,18	1,6,7, 10,14	1,6,10, 16,18
	Sen	.8855	.8855	.8995	.7789	.8888	.8603	.9034	.8603
	Spe	.8869	.8869	.8688	.8972	.8894	.9036	.8662	.9036
	Error	.1126	.1126	.1205	.1407	.1094	.1082	.1207	.1082
	Rej	.5060	.5060	.4919	.5073	.5145	.5158	.5133	.5158
7	Reg	3,6,7, 10,11, 16,21	3,6,7, 10,11, 16,21	1,5,6, 7,8, 10,14	2,3,6, 10,16, 20,21	1,3,5, 7,10, 11,16	1,3,5, 7,10, 11,16	3,5,6, 7,10, 11,14	1,2,6, 11,13, 16,18

	Sen	.8733	.8733	.8909	.8440	.8942	.8942	.8984	.8348
	Spe	.8948	.8948	.8646	.8963	.8935	.8935	.8661	.9031
	Error	.1109	.1109	.1267	.1190	.1047	.1047	.1227	.1163
	Rej	.5048	.5048	.4970	.4950	.5117	.5117	.5067	.5120
9	Reg	2,3,5, 6,7, 10,16, 20,21	3,5,6, 7,10, 11,16, 18,21	2,3,5, 6,7, 8,10, 11,14	2,3,6, 7,13, 16,18, 21,22	1,3,5, 6,7, 10,11, 12,16	1,5,6, 7,10, 11,16, 18,21	3,5,6, 7,8, 10,11, 14,19	1,2,6, 9,10, 11,16, 21,22
	Sen	.8675	.8639	.8788	.8333	.8903	.8802	.8915	.8439
	Spe	.8902	.8919	.8708	.8943	.8904	.8959	.8695	.9052
	Error	.1160	.1158	.1251	.1249	.1081	.1074	.1220	.1130
	Rej	.5085	.5048	.5101	.5007	.5136	.5090	.5005	.5147

5 El uso de más clasificadores base redujo la tasa de rechazo sin reducir significativamente la sensibilidad y especificidad. Esto es particularmente cierto en los casos en que se utilizaron clasificadores de base con una tasa de rechazo de 0.33, aunque también es observable, en menor medida, en los casos en que se utilizaron 0.50.

10 La mejora obtenida al utilizar el enfoque integrado puede apreciarse realmente cuando se compara con los resultados de clasificación obtenidos usando el enfoque estándar. Con el enfoque estándar se obtiene una sensibilidad y especificidad de 0.74 y 0.75 mientras se utiliza el enfoque integrado se obtiene una sensibilidad, especificidad y tasa de rechazo de 0.85, 0.85, 0.27 con clasificadores base que tienen una tasa de rechazo de alrededor de 0.33 y de 0.89, 0.89, 0.51 con clasificadores base que tienen una tasa de rechazo de alrededor de 0.50.

Conclusión

15 Se evaluaron métodos de extracción de características y selección de características, específicamente PCA, SFFS y SVDSS, con el fin de tratar la alta naturaleza dimensional de los datos ESS. Si bien no se observó una mejora significativa en el rendimiento de clasificación cuando se aplicaron los diferentes métodos de reducción de la dimensionalidad a los datos de prueba, el uso de selección de características permitió comprender mejor cuáles áreas del espectro de ESS eran más útiles para fines de clasificación. Se vio que las longitudes de onda más cortas eran más adecuadas para esto. Esta observación fue confirmada posteriormente como clasificadores fueron diseñados en diferentes regiones del espectro, los diseñados en regiones en las longitudes de onda más cortas comparables, y en algunos casos mejor, el rendimiento en comparación con el uso de todo el espectro, en lugar de utilizar longitudes de onda más largas donde el desempeño de clasificación no fue tan bueno.

25 Un desafío importante que surge en esta aplicación es la variabilidad biológica inherente presente. Esto hace que la distinción de patologías, con alta precisión, sea particularmente difícil. Utilizando el paradigma estándar de reducción de la dimensionalidad y la clasificación en el espectro ESS, con SVM con núcleo lineal, se obtuvieron sensibilidades y especificidad de ~0.75. Por otra parte, el uso de núcleos no lineales no mejoró significativamente este resultado, lo que demuestra que, de hecho, el conjunto de datos no fue linealmente ni no linealmente separable. Para superar estos elementos se utilizó un marco de rechazo de errores. El clasificador identifica la región donde las muestras tienen más probabilidades de ser clasificadas erróneamente y evita clasificar o rechazar las muestras que se encuentran en esta región. Estas muestras pueden ser examinadas por otros métodos. Este marco particular se adoptó ya que es muy adecuado para la aplicación de cribado de cáncer de colon. Si el clasificador no toma una decisión sobre una muestra, se pueden tomar otras medidas ópticas o se puede tomar una biopsia y examinarla un patólogo. En particular, el clasificador es una máquina de vector de apoyo desarrollada con rechazo de error incorporado. Tanto la orientación como la anchura de la región de decisión, de acuerdo con pesos que reflejan el equilibrio entre muestras de rechazo y de clasificación errónea, se obtienen durante la fase de entrenamiento. El problema de formación se formula como un problema de optimización convexa, un algoritmo de entrenamiento dual, basado en la optimización mínima secuencial (SMO), así como permite el uso de núcleos para obtener regiones de decisión no lineal. Los resultados mostraron que este enfoque produjo mejores resultados que los métodos que rechazan las muestras mediante el umbral de las salidas de los clasificadores estándar. Se obtuvieron mejoras adicionales en los resultados utilizando núcleos no lineales. Cuando se comparó con el paradigma de clasificación estándar, que produjo sensibilidades de ~0.74 y especificidades de ~0.77 con un núcleo RBF, utilizando el rechazo de error, se pueden obtener una sensibilidad de ~0.80 y especificidad de ~0.85 con una tasa de rechazo de ~0.33, y una sensibilidad de ~0.82 y especificidad de ~0.89 con una tasa de rechazo de ~0.50, también usando un núcleo RBF.

45 Otra metodología de clasificación que se empleó para mejorar el rendimiento fue un sistema clasificador múltiple. El sistema clasificador múltiple toma una decisión final combinando la decisión de varios clasificadores base. Cada clasificador de base se diseña utilizando características extraídas de una región diferente del espectro de ESS. Las regiones utilizadas fueron elegidas heurísticamente, utilizando observaciones e intuición recogidas de los resultados de los experimentos de selección de características. Cabe destacar que las regiones como 330-600 nm, con

5 sensibilidad y especificidad de ~0.76, 330-400 nm, con sensibilidad de ~.8 y especificidad de ~0.73, y 360-600 nm, con sensibilidad de ~0.77 y especificidad de ~0.74, mostró un rendimiento comparable, si no mejor, al uso de todo el espectro, 330-760 nm, cuya sensibilidad fue de ~0.74 y especificidad de ~0.77. Cuando se usan las sensibilidades y especificidades del clasificador del conjunto de -0.80 y -0.78, -0.81 y -0.76, -0.75 y -0.83, y finalmente -0.78 y -0.80 se pueden obtener usando diferentes combinaciones de regiones. Estos resultados muestran claramente una mejora general en el rendimiento de cualquier clasificador único dado utilizado como clasificadores base.

10 Por último, los sistemas de rechazo de errores y clasificadores múltiples se incorporaron en un marco unificado. Esto se logró integrando el SVM desarrollado con el rechazo de errores incorporado con el sistema de clasificadores múltiples descrito anteriormente. Los clasificadores con rechazo de errores fueron entrenados en las regiones espectrales de ESS usadas anteriormente y luego combinados para obtener la decisión final. Una vez más, se buscan mejoras sobre la sensibilidad de 0.80, la especificidad de 0.85 con la tasa de rechazo de 0.32 y la sensibilidad de 0.82, la especificidad de 0.89 con la tasa de rechazo de 0.51 obtenida usando rechazo de error con todo el espectro, 330-760 nm, u otra regiones de buen funcionamiento como 330-400 nm con sensibilidades, especificidades y tasas de rechazo de 0.86, 0.80, 0.32 y 0.88, 0.83, 0.49 y 360-600 nm con sensibilidades, especificidades y tasas de rechazo de 0.83, 0.83, 0.33 y 0.88, 0.87, 0.53. Usando el marco unificado, diferentes conjuntos de clasificadores con sensibilidades, especificidades y tasas de rechazo de 0.85, 0.85, 0.27 u 0.86, 0.86, 0.33 u 0.87, 0.84, 0.29 u 0.88, 0.81, 0.32 pueden obtenerse por clasificadores de base entrenados para tener una tasa de rechazo de 0.33. Usando alrededor de 0.50 como la tasa de rechazo del clasificador base, conjuntos con sensibilidades, especificidades y tasas de rechazo de 0.89, 0.89, 0.51 o 0.90, 0.87, 0.49 u 0.86, 0.91, 0.52 u 0.88, 0.90, 0.51 u 0.87, 0.88, 0.46, mostrando claramente mejoras respecto a los clasificadores individuales que utilizan todo el espectro o cualquier región individual.

AdaBoost

El algoritmo AdaBoost es un método que intenta "aumentar" la precisión de cualquier algoritmo de aprendizaje dado.

25 Sea $x_i \in \mathcal{R}^d$, sea un vector de entrada y $y_i \in \{1,-1\}$ su etiqueta correspondiente para $i=1, \dots, l$, donde l es el número total de muestras de entrenamiento. El objetivo es construir una función de decisión compuesta por un conjunto de clasificadores, $g_i: x \rightarrow \{\pm 1\}$ denominados *aprendices de base o aprendices débiles*, de la forma:

$$f_T(\mathbf{x}) = \sum_{t=1}^T c_t g_t(\mathbf{x}). \quad (63)$$

30 El algoritmo llama al aprendiz de base repetidamente sobre una serie de rondas $t=1, \dots, T$, manteniendo una distribución o un conjunto de pesos sobre el conjunto de entrenamiento. Los pesos de las muestras incorrectamente clasificadas se incrementan para que el aprendiz débil se enfoque en las muestras difíciles de clasificar. En cada ronda t , el aprendiz de base $g_t(x)$ se aplica al conjunto de entrenamiento con la distribución de peso actual. Entonces el algoritmo asigna un peso a esa hipótesis que intuitivamente mide la importancia asignada a $g_t(x)$. Los detalles del algoritmo de entrenamiento AdaBoost se muestran en la Figura 14.

35 El problema de la inclusión de una opción de rechazo de errores en sistemas clasificadores múltiples es una combinación lineal de las salidas de clasificadores entrenados. La mejor relación de error-rechazo alcanzable por una combinación lineal de clasificadores no siempre es peor que la del mejor clasificador individual. Esta compensación depende no sólo del umbral de rechazo sino también de los coeficientes de la combinación lineal de clasificadores. Por lo tanto, el problema del entrenamiento de un sistema de clasificación múltiple con opción de rechazo es el de encontrar los coeficientes y el umbral de rechazo dados los costes de error y rechazo.

40 Las realizaciones de la presente invención también aplican rechazo de errores al clasificador AdaBoost. Se aprovechó la similitud entre el problema de entrenamiento con la máquina de vectores de soporte y una formulación planteada para introducir márgenes blandos en AdaBoost. Este nuevo problema de entrenamiento no sólo busca obtener el umbral de rechazo sino también recalculan los pesos de conjunto, formando así las regiones de decisión. Una vez más, estos dos dependen de los costes de rechazo y error.

45 AdaBoost alcanza asintóticamente un margen duro, es decir, el algoritmo se concentra en el patrón de aprendizaje difícil sin permitir errores. Por lo tanto, es deseable algún tipo de regularización que permita al algoritmo desconfiar de los datos y lograr un margen suave, es decir, se permiten algunos errores. Empíricamente se observó que cuanto más diferentes son las ponderaciones de las hipótesis, b_i , mayor es la complejidad del conjunto.

50 De este modo, la norma del vector de pesos $\|b\|$ puede considerarse como una medida de complejidad. Si $\|b\|$ tiene un valor alto, entonces algunas de las hipótesis se enfatizan fuertemente. Para lograr un margen suave, sería deseable que $\|b\|$ tuviera un valor menor, ya que esto significaría que el énfasis en las hipótesis es más distribuido. Con esto en mente y por la introducción de las variables de holgura ξ_i , se proporciona el siguiente problema de optimización que encuentra los valores de b_i para conseguir márgenes de clasificación:

minimizar

$$\|\mathbf{b}\|^2 + C \sum_i \xi_i, \quad (64)$$

con las siguientes limitaciones

$$\begin{aligned} \sum_{t=1}^T b_t y_t g_t(\mathbf{x}_i) &\geq 1 - \xi_i, \quad i = 1, \dots, l, \\ \xi_i &\geq 0, \quad i = 1, \dots, l, \\ b_t &\geq 0, \quad t = 1, \dots, T. \end{aligned} \quad (65)$$

5 Este problema se resuelve después de que el algoritmo de AdaBoost mostrado en la Figura 3 se ejecute para obtener la hipótesis $T g_1, \dots, g_T$.

El problema de entrenamiento incorpora una opción de rechazo de errores que se puede formular usando h_c funcional (ξ_i, ε) , descrito anteriormente. El problema es entonces minimizar

$$\|\mathbf{b}\|^2 + C \sum_i h_c(\xi_i, \varepsilon), \quad (66)$$

10 con restricciones

$$\begin{aligned} \sum_{t=1}^T b_t y_t g_t(\mathbf{x}_i) &\geq 1 - \xi_i, \quad i = 1, \dots, l, \\ \xi_i &\geq 0, \quad i = 1, \dots, l, \\ b_t &\geq 0, \quad t = 1, \dots, T, \\ 0 &\leq \varepsilon \leq 1. \end{aligned} \quad (67)$$

La función de decisión se obtiene así:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_{t=1}^T b_t g_t(\mathbf{x}) \geq \varepsilon \\ -1, & \text{if } \sum_{t=1}^T b_t g_t(\mathbf{x}) \leq -\varepsilon . \\ 0, & \text{if } -\varepsilon < \sum_{t=1}^T b_t g_t(\mathbf{x}) < \varepsilon \end{cases} \quad (68)$$

15 De este modo, para conseguir el rechazo de errores en un clasificador de conjuntos, se calculan los pesos de conjunto y el umbral de rechazo dados los costes de rechazo y clasificación errónea de las muestras.

A menos que se indique específicamente lo contrario, a lo largo de la presente divulgación, términos tales como "procesamiento", "cómputo", "cálculo", "determinación", o similares, pueden referirse a las acciones y/o procesos de un ordenador o sistema informático o dispositivo de cómputo electrónico similares, que manipulan y/o transforman datos representados como cantidades físicas, tales como electrónicas, dentro de los registros y/o memorias del sistema informático en otros datos representados de manera similar como cantidades físicas dentro de las memorias, registros u otros tales dispositivos del sistema informático de almacenamiento, transmisión o visualización de información.

20 Las realizaciones de la presente invención pueden incluir un aparato para llevar a cabo las operaciones en el mismo. Dicho aparato puede estar especialmente construido para los fines deseados, o puede comprender un ordenador de uso general selectivamente activado o reconfigurado por un programa informático almacenado en el ordenador.

25 La Figura 15 muestra una representación esquemática de una máquina en la forma de ejemplo de un sistema 1500 informático dentro del cual se puede ejecutar un conjunto de instrucciones, para hacer que la máquina realice cualquiera o más de las metodologías aquí discutidas. En realizaciones alternativas, la máquina funciona como un dispositivo autónomo o puede estar conectada (por ejemplo, conectada en red) a otras máquinas. En una

- 5 implementación en red, la máquina puede funcionar en la capacidad de un servidor o una máquina cliente en un entorno de red cliente-servidor, o como una máquina igual en un entorno de red de igual a igual (o distribuido). La máquina puede ser un servidor, una computadora personal (PC), una tableta, un decodificador (STB), un Asistente Personal Digital (PDA), un teléfono celular, un dispositivo web, un enrutador de red, un conmutador o un puente, o cualquier máquina capaz de ejecutar un conjunto de instrucciones (secuenciales o de otro tipo) que especifiquen las acciones a tomar por esa máquina. Además, aunque sólo se ilustra una única máquina, se considerará que el término "máquina" incluye cualquier colección de máquinas que individual o conjuntamente ejecuten un conjunto (o múltiples conjuntos) de instrucciones para realizar cualquiera o más de las metodologías aquí discutidas.
- 10 El sistema 1500 informático de ejemplo incluye un procesador 1502 (por ejemplo, una unidad central de procesamiento (CPU), una unidad de procesamiento de gráficos (GPU) o ambas), una memoria principal 1504 (por ejemplo, memoria de sólo lectura (ROM), memoria instantánea dinámica aleatoria y una memoria 1506 estática (por ejemplo, memoria instantánea, memoria estática de acceso aleatorio (SRAM), etc.), que se comunican entre sí (DRAM), tales como DRAM síncrona (SDRAM) o Rambus DRAM (RDRAM) a través de un bus 1508.
- 15 El sistema 1500 informático puede incluir además una unidad 1510 de visualización de vídeo (por ejemplo, una pantalla de cristal líquido (LCD) o un tubo de rayos catódicos (CRT)). El sistema 1500 informático también incluye un dispositivo 1512 de entrada alfanumérico (por ejemplo, un teclado), un dispositivo 1514 de control de cursor (por ejemplo, un ratón), una unidad 1516 de disco, un dispositivo 1520 de generación de señal (por ejemplo, un altavoz) y un dispositivo 1522 de interfaz de red.
- 20 La unidad 1516 de accionamiento de disco incluye un medio 1524 legible por máquina en el que se almacena uno o más conjuntos de instrucciones (por ejemplo, software 1526) que incorpora cualquiera o más de las metodologías o funciones descritas en la presente memoria. El software 1526 también puede residir, total o al menos parcialmente, dentro de la memoria 1504 principal y/o dentro del procesador 1502 durante la ejecución del software 1526 por el sistema 1500 informático.
- 25 El software 1526 puede ser además transmitido o recibido a través de una red 1528 a través del dispositivo 1522 de interfaz de red.
- 30 Mientras que el medio 1524 legible por máquina se muestra en una realización de ejemplo para ser un solo medio, el término "medio legible por máquina" debe entenderse incluyendo un único medio o medios múltiples (por ejemplo, una base de datos centralizada o distribuida y/o cachés y servidores asociados) que almacenan uno o más conjuntos de instrucciones. El término "medio legible por máquina" también se tomará para incluir cualquier medio que sea capaz de almacenar, codificar o transportar un conjunto de instrucciones para su ejecución por la máquina y que haga que la máquina realice cualquiera o más de las metodologías de la presente divulgación. Por consiguiente, el término "medio legible por máquina" se incluirá, pero no se limita a, memorias de estado sólido, medios ópticos y magnéticos y ondas portadoras. Por consiguiente, el término "medio de almacenamiento legible por máquina" incluirá, pero no se limita a, las memorias de estado sólido y los medios ópticos y magnéticos (por ejemplo, cualquier tipo de disco incluyendo discos flexibles, discos ópticos, CD-ROMs, (ROM), memorias de acceso aleatorio (RAM), memorias de sólo lectura programables eléctricamente (EPROM), memorias de sólo lectura programables (EEPROM), tarjetas magnéticas o ópticas, o cualquier otro tipo de soporte adecuado para almacenar instrucciones o datos electrónicos, y apto para ser acoplado a un bus del sistema informático).
- 35
- 40 La invención ha sido descrita a través de módulos funcionales, que están definidos por instrucciones ejecutables grabadas en medios legibles por ordenador que hacen que un ordenador realice pasos de método cuando se ejecutan. Los módulos han sido segregados por función para mayor claridad. Sin embargo, debe entenderse que los módulos no necesitan corresponder a bloques discretos de código y las funciones descritas pueden ser llevadas a cabo por la ejecución de diversas porciones de código almacenadas en diversos medios y ejecutadas en diversos momentos.
- 45 Debe entenderse que los procedimientos y técnicas descritos en la presente memoria no están inherentemente relacionados con ningún aparato particular y pueden implementarse mediante cualquier combinación adecuada de componentes. Además, se pueden usar diversos tipos de dispositivos de propósito general de acuerdo con las enseñanzas descritas en la presente memoria. También puede resultar ventajoso construir un aparato especializado para realizar las etapas del método descritas en la presente memoria. La presente invención se ha descrito en relación con ejemplos particulares, que están destinados en todos los aspectos a ser ilustrativos en lugar de restrictivos. Los expertos en la técnica apreciarán que muchas combinaciones diferentes de hardware, software y firmware serán adecuadas para poner en práctica la presente invención.
- 50
- 55 Además, resultarán evidentes otras realizaciones de la invención para los expertos en la técnica a partir de la consideración de la descripción y práctica de la invención descrita en la presente memoria. Diversos aspectos y/o componentes de las realizaciones descritas pueden usarse individualmente o en cualquier combinación. Se pretende que la especificación y los ejemplos se consideren sólo como ejemplos, con un verdadero alcance de la invención indicado mediante las siguientes reivindicaciones.

Los siguientes párrafos numerados definen realizaciones particulares de la presente descripción:

1. Un método para clasificar una muestra de tejido que comprende:

formar un clasificador para determinar una región de rechazo de datos de espectros biomédicos correspondientes a muestras de tejido que puedan ser clasificadas erróneamente;

5 clasificar la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido no están en la región de rechazo; y

detener la clasificación de la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido están en la región de rechazo.

10 2. El método del párrafo 1, que comprende además realizar una biopsia de la muestra de tejido si se detiene la clasificación.

3. El método del párrafo 1, que comprende además realizar mediciones ópticas adicionales de la muestra de tejido si la clasificación es detenida.

4. El método del párrafo 1, en el que el entrenamiento del clasificador comprende formular un problema de entrenamiento como un algoritmo de entrenamiento dual de problemas de optimización convexo.

15 5. El método del párrafo 1, en el que los datos de espectros biomédicos comprenden datos de ESS.

6. El método del párrafo 1, en el que el entrenamiento del clasificador comprende obtener la doble formación de:

$$\begin{aligned} \min_{\mathbf{w}, \beta, \xi_i, \varepsilon, z_i} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_i z_i \\ \text{s.t.} & y_i (\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i \\ & z_i \geq C w_c \xi_i \\ & z_i \geq C [w_r \xi_i + (w_r - w_c) \varepsilon + w_c - w_r] \\ & z_i \geq C [w_c \xi_i + (2w_r - w_c - w_c) \varepsilon + w_c - w_c] \\ & \xi_i \geq 0 \\ & 0 \leq \varepsilon \leq 1. \end{aligned}$$

7. El método del párrafo 6, en el que el problema dual es: 71

$$\begin{aligned} \max_{\alpha_i, \beta c_i, \beta r_i, \beta e_i} & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + C(w_c - w_r) \sum_i \beta r_i + C(w_c - w_c) \sum_i \beta e_i \\ \text{s.t.} & \sum_i \alpha_i y_i = 0 \\ & \beta c_i + \beta r_i + \beta e_i = 1 \\ & C(w_r - w_c) \sum_i \beta r_i + C(2w_r - w_c - w_c) \sum_i \beta e_i = 0 \\ & 0 \leq \alpha_i \leq C w_c \beta c_i + C w_r \beta r_i + C w_c \beta e_i \\ & \beta c_i, \beta r_i, \beta e_i \geq 0. \end{aligned}$$

20 8. El método del párrafo 1, en el que una función de decisión del clasificador comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases}$$

9. Un medio de almacenamiento legible por máquina que comprende instrucciones ejecutables por una máquina de procesamiento de datos para realizar un método para clasificar una muestra de tejido, comprendiendo el método:
- 5 entrenar un clasificador para determinar una región de rechazo de datos de espectros biomédicos correspondientes a muestras de tejido que puedan ser clasificadas erróneamente;
- clasificar la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido no están en la región de rechazo; y
- detener la clasificación de la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido están en la región de rechazo.
- 10 10. El medio de almacenamiento legible por máquina del párrafo 9, en el que el entrenamiento del clasificador comprende formular un problema de entrenamiento como un algoritmo de entrenamiento dual de problema de optimización convexa.
11. El medio de almacenamiento legible por máquina del párrafo 9, en el que los datos de espectros biomédicos comprenden datos de ESS.
- 15 12. Un sistema configurado para clasificar una muestra de tejido que comprende:
- medios para entrenar a un clasificador para determinar una región de rechazo de datos de especímenes biomédicos correspondientes a muestras de tejido que puedan ser clasificadas erróneamente; y
- medios para clasificar la muestra de tejido si los datos de los especímenes biomédicos correspondientes a la muestra de tejido no están en la región de rechazo y detención de la clasificación de la muestra de tejido si los datos
- 20 de los especímenes biomédicos correspondientes a la muestra de tejido están en la región de proyección.
13. El sistema del párrafo 12, en el que los datos de espectros biomédicos comprenden datos de ESS.
14. Un método para clasificar una muestra de tejido que comprende:
- clasificar una primera región de datos de espectros biomédicos utilizando un primer clasificador;
- clasificar una segunda región de datos de espectros biomédicos utilizando un segundo clasificador, siendo la
- 25 primera región diferente de la segunda región; y
- combinar la clasificación del primer clasificador con la clasificación del segundo clasificador para determinar una clasificación de la muestra de tejido.
15. El método del párrafo 14, en el que los datos de espectros biomédicos comprenden datos de ESS.
16. El método del párrafo 14, en el que la clasificación del primer clasificador y la clasificación del segundo
- 30 clasificador se seleccionan del grupo que consiste en positivo, negativo y rechazado.
17. El método del párrafo 14, en el que la primera región se superpone a la segunda región.
18. El método del párrafo 14, que comprende además clasificar una tercera región de los datos de espectros biomédicos utilizando un tercer clasificador y en el que la clasificación del tercer clasificador se combina con la
- 35 clasificación del primer clasificador y la clasificación del segundo clasificador para determinar la clasificación de la muestra de tejido.

19. El método del párrafo 14, en el que la clasificación de la primera región comprende detener la clasificación de la primera región.

20. El método del párrafo 14, en el que clasificar la segunda región comprende detener la clasificación de la segunda región.

5 21. El método del párrafo 14, en el que el primer clasificador y el segundo clasificador son máquinas de soporte vectorial.

22. El método del párrafo 14, en el que el primer clasificador y el segundo clasificador son máquinas de soporte vectorial con rechazo de errores incorporado.

10 23. El método del párrafo 14, en el que la combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador comprende el uso de una regla de votación mayoritaria.

24. El método del párrafo 14, en el que la combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador comprende el uso de una regla de Bayes ingenua.

15 25. El método del párrafo 14, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primeras y segundas regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \mathbf{w}^T \mathbf{x} - b \geq \varepsilon \\ -1, & \text{if } \mathbf{w}^T \mathbf{x} - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \mathbf{w}^T \mathbf{x} - b < \varepsilon \end{cases}$$

26. El método del párrafo 14, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primeras y segundas regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases}$$

20 27. Un medio de almacenamiento legible por máquina que comprende instrucciones ejecutables por una máquina de procesamiento de datos para realizar un método para clasificar una muestra de tejido, comprendiendo el método:

clasificar una primera región de datos de espectros biomédicos utilizando un primer clasificador;

25 clasificar una segunda región de datos de espectros biomédicos utilizando un segundo clasificador, siendo la primera región diferente de la segunda región; y

combinar la clasificación del primer clasificador con la clasificación del segundo clasificador para determinar una clasificación de la muestra de tejido.

28. El medio de almacenamiento legible por máquina del párrafo 27, en el que los datos de espectros biomédicos comprenden datos de ESS.

30 29. El medio de almacenamiento legible por máquina del párrafo 27, en el que la clasificación del primer clasificador y la clasificación del segundo clasificador se seleccionan del grupo que consiste en positivo, negativo y rechazado.

30. El medio de almacenamiento legible por máquina del párrafo 27, en el que la primera región se superpone a la segunda región.
- 5 31. El medio de almacenamiento legible por máquina del párrafo 27, que comprende además clasificar una tercera región de los datos de espectros biomédicos utilizando un tercer clasificador y en el que la clasificación del tercer clasificador se combina con la clasificación del primer clasificador y la clasificación del segundo clasificador para determinar la clasificación de la muestra de tejido.
32. El medio de almacenamiento legible por máquina del párrafo 27, en el que clasificar la primera región comprende detener la clasificación de la primera región.
- 10 33. El medio de almacenamiento legible por máquina del párrafo 27, en el que clasificar la segunda región comprende detener la clasificación de la segunda región.
34. El medio de almacenamiento legible por máquina del párrafo 27, en el que el primer clasificador y el segundo clasificador son máquinas de soporte vectorial.
35. El medio de almacenamiento legible por máquina del párrafo 27, en el que el primer clasificador y el segundo clasificador son máquinas de vectores de soporte con rechazo de error incorporado.
- 15 36. El medio de almacenamiento legible por máquina del párrafo 27, en el que la combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador comprende el uso de una regla de votación mayoritaria.
37. El medio de almacenamiento legible por máquina del párrafo 27, en el que la combinación de la clasificación del primer clasificador con la clasificación del segundo clasificador comprende el uso de una regla Bayes ingenua.
- 20 38. El medio de almacenamiento legible por máquina del párrafo 27, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primeras y segundas regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \mathbf{w}^T \mathbf{x} - b \geq \varepsilon \\ -1, & \text{if } \mathbf{w}^T \mathbf{x} - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \mathbf{w}^T \mathbf{x} - b < \varepsilon \end{cases}$$

- 25 39. El medio de almacenamiento legible por máquina del párrafo 27, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primera y segunda regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases}$$

40. Un sistema configurado para clasificar una muestra de tejido que comprende:
medios para clasificar una primera región de datos de espectros biomédicos;
- 30 medios para clasificar una segunda región de datos de espectros biomédicos, siendo la primera región diferente de la segunda región; y
medios para combinar la clasificación de la primera región con la clasificación de la segunda región para determinar una clasificación de la muestra de tejido.

41. El sistema del párrafo 40, en el que los datos de espectros biomédicos comprenden datos de ESS.
42. El sistema del párrafo 40, en el que la primera región se solapa con la segunda región.
43. El sistema del párrafo 40, que comprende además medios para clasificar una tercera región de los datos de espectros biomédicos y en el que la clasificación de la tercera región se combina con la clasificación de la primera región y la clasificación de la segunda región para determinar la clasificación de la muestra de tejido.
- 5 44. Un sistema configurado para clasificar una muestra de tejido que comprende:
una sonda óptica configurada para medir los datos de espectros biomédicos; y
un sistema de clasificación acoplado a la sonda óptica que comprende una máquina de vector de soporte con rechazo de error incorporado para clasificar los datos de espectros biomédicos.
- 10 45. El sistema del párrafo 44 en el que la sonda óptica es una herramienta de pinzas de biopsia integrada.
46. El sistema del párrafo 45, en el que la herramienta de pinzas de biopsia integrada comprende una fibra óptica de iluminación para dirigir la luz al tejido que se va a fotografiar y una fibra óptica de recolección para recoger la luz dispersada en el tejido.
47. El sistema del párrafo 44 en el que los datos de espectros biomédicos son datos de ESS.
- 15 48. El sistema del párrafo 44 en el que el sistema de clasificación comprende una pluralidad de máquinas de vector de soporte con rechazo de errores incorporado para clasificar los datos de espectros biomédicos.
49. El sistema del párrafo 48 en el que cada una de la pluralidad de máquinas vectoriales de soporte con rechazo de errores incorporado está configurada para clasificar una región diferente de los datos de los espectros.
- 20 50. El sistema del párrafo 44 en el que el sistema de clasificación combina el resultado de cada una de la pluralidad de máquinas vectoriales de soporte con rechazo de errores incorporado para realizar una clasificación final de los datos de los especímenes biomédicos.
51. El sistema del párrafo 44 en el que la máquina de vectores de soporte está configurada para clasificar los datos de espectros biomédicos utilizando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \mathbf{w}^T \mathbf{x} - b \geq \varepsilon \\ -1, & \text{if } \mathbf{w}^T \mathbf{x} - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \mathbf{w}^T \mathbf{x} - b < \varepsilon \end{cases}$$

- 25 52. El sistema del párrafo 44 en el que la máquina de vector de soporte está configurada para clasificar los datos de espectros biomédicos usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases}$$

Reivindicaciones

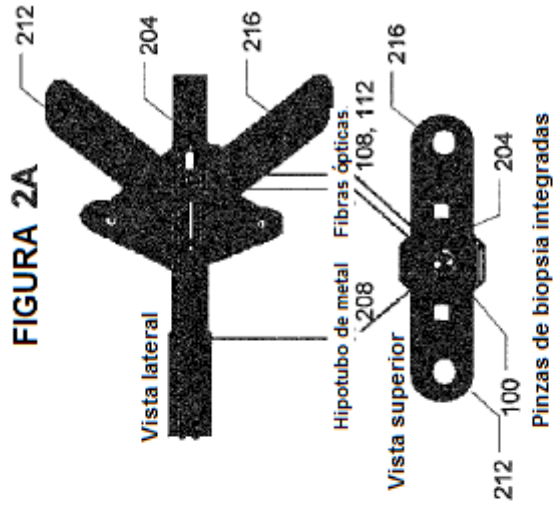
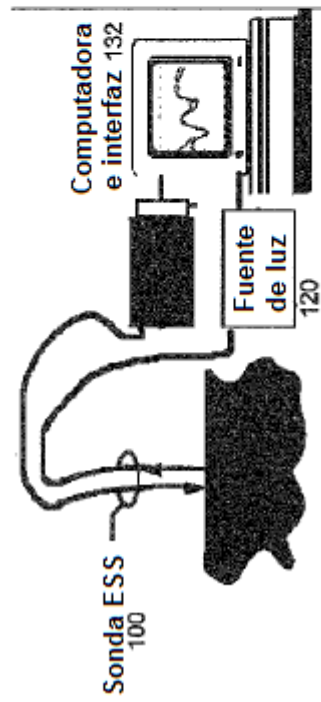
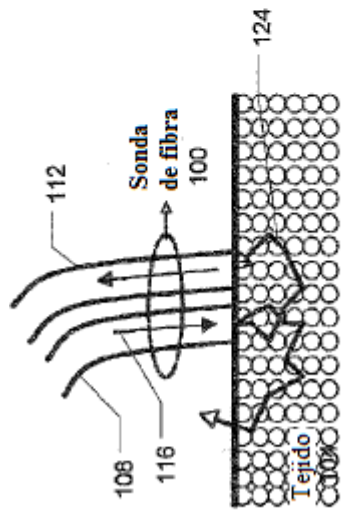
1. Un método para clasificar una muestra de tejido que comprende:
 clasificar una primera región de datos de espectroscopía óptica biomédica utilizando un primer clasificador (1316);
 clasificar una segunda región de datos de espectroscopía óptica biomédica utilizando un segundo clasificador (1316), siendo la primera región diferente de la segunda región; y
 combinar la clasificación del primer clasificador con la clasificación del segundo clasificador para determinar una clasificación de la muestra de tejido;
 caracterizado porque
 la clasificación del primer clasificador y la clasificación del segundo clasificador se seleccionan del grupo que consiste en positivo, negativo y rechazado.
2. El método de la reivindicación 1, en el que los datos de espectroscopía óptica biomédica comprenden datos de espectroscopía de dispersión elástica (ESS).
3. El método de la reivindicación 1, en el que la primera región se solapa con la segunda región.
4. El método de la reivindicación 1, que comprende además clasificar una tercera región de los datos de espectros biomédicos utilizando un tercer clasificador (1316) y en el que la clasificación del tercer clasificador se combina con la clasificación del primer clasificador y la clasificación del segundo clasificador para determinar la clasificación de la muestra de tejido.
5. El procedimiento de la reivindicación 1, en el que clasificar la primera región comprende detener la clasificación de la primera región, o en el que clasificar la segunda región comprende detener la clasificación de la segunda región.
6. El método de la reivindicación 1, en el que el primer clasificador y el segundo clasificador son máquinas de soporte vectorial.
7. El método de la reivindicación 1, en el que el primer clasificador y el segundo clasificador son máquinas de soporte vectorial que incorporan rechazo de errores.
8. El método de la reivindicación 1, en el que combinar la clasificación del primer clasificador con la clasificación del segundo clasificador comprende usar una regla de votación mayoritaria, o usar una regla de Bayes ingenua.
9. El método de la reivindicación 1, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primeras y segundas regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \mathbf{w}^T \mathbf{x} - b \geq \varepsilon \\ -1, & \text{if } \mathbf{w}^T \mathbf{x} - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \mathbf{w}^T \mathbf{x} - b < \varepsilon \end{cases}$$

10. El método de la reivindicación 1, en el que el primer clasificador y el segundo clasificador están configurados para clasificar las primeras y segundas regiones usando una función de decisión en la que la función de decisión comprende:

$$f(\mathbf{x}) = \begin{cases} +1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \geq \varepsilon \\ -1, & \text{if } \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b \leq -\varepsilon \\ 0, & \text{if } -\varepsilon < \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) - b < \varepsilon \end{cases}$$

11. Un medio de almacenamiento (1524) legible por máquina que comprende instrucciones ejecutables por una máquina de procesamiento de datos para realizar un método para clasificar una muestra de tejido, comprendiendo el método:
- 5 clasificar una primera región de datos de espectroscopía óptica biomédica utilizando un primer clasificador (1316);
 clasificar una segunda región de datos de espectroscopía óptica biomédica utilizando un segundo clasificador (1316), siendo la primera región diferente de la segunda región; y
 combinar la clasificación del primer clasificador con la clasificación del segundo clasificador para determinar una clasificación de la muestra de tejido;
- 10 caracterizado porque la clasificación del primer clasificador y la clasificación del segundo clasificador se seleccionan del grupo que consiste en positivo, negativo y rechazado.
12. El medio de almacenamiento legible por máquina de la reivindicación 11, en el que el método es como se define en una cualquiera de las reivindicaciones 2 a 10.
13. Un sistema configurado para clasificar una muestra de tejido que comprende:
- 15 medios para clasificar una primera región de datos de espectroscopía óptica biomédica;
 medios para clasificar una segunda región de datos de espectroscopía óptica biomédica, siendo la primera región diferente de la segunda región; y
 medios para combinar la clasificación de la primera región con la clasificación de la segunda región para determinar una clasificación de la muestra de tejido,
- 20 caracterizado porque
 la clasificación de la primera región y la clasificación de la segunda región se seleccionan del grupo que consiste en positivo, negativo y rechazado.
14. El sistema de la reivindicación 13, en el que los datos de espectroscopía óptica biomédica comprenden datos de ESS.
- 25



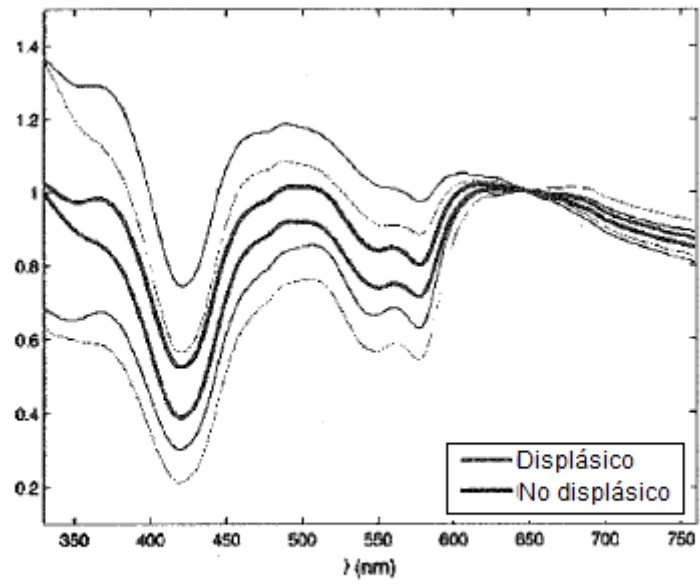


FIGURA 3

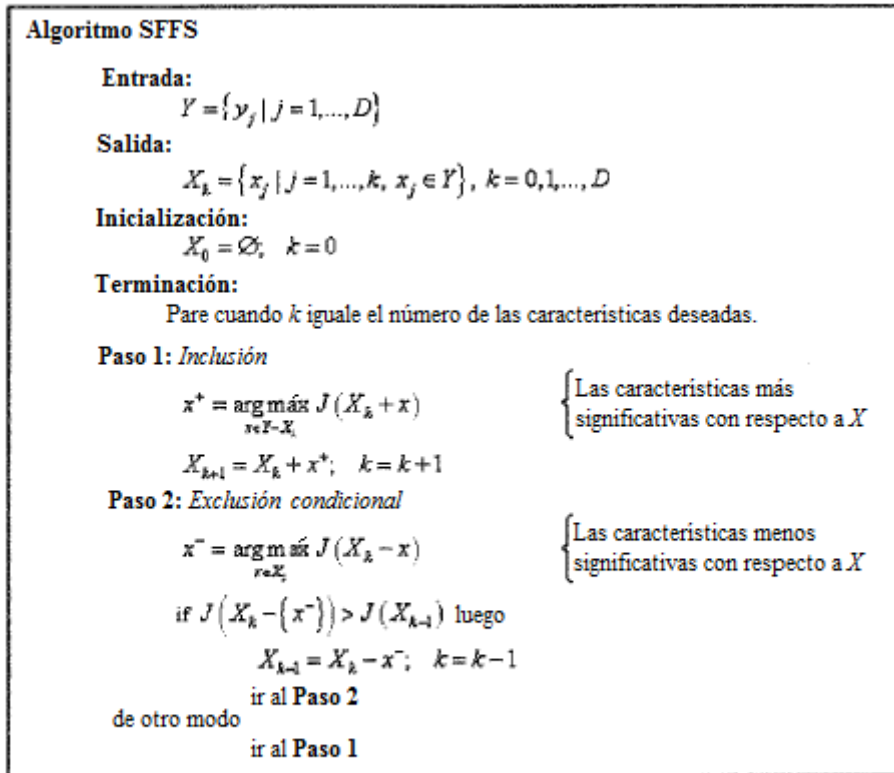


FIGURA 4

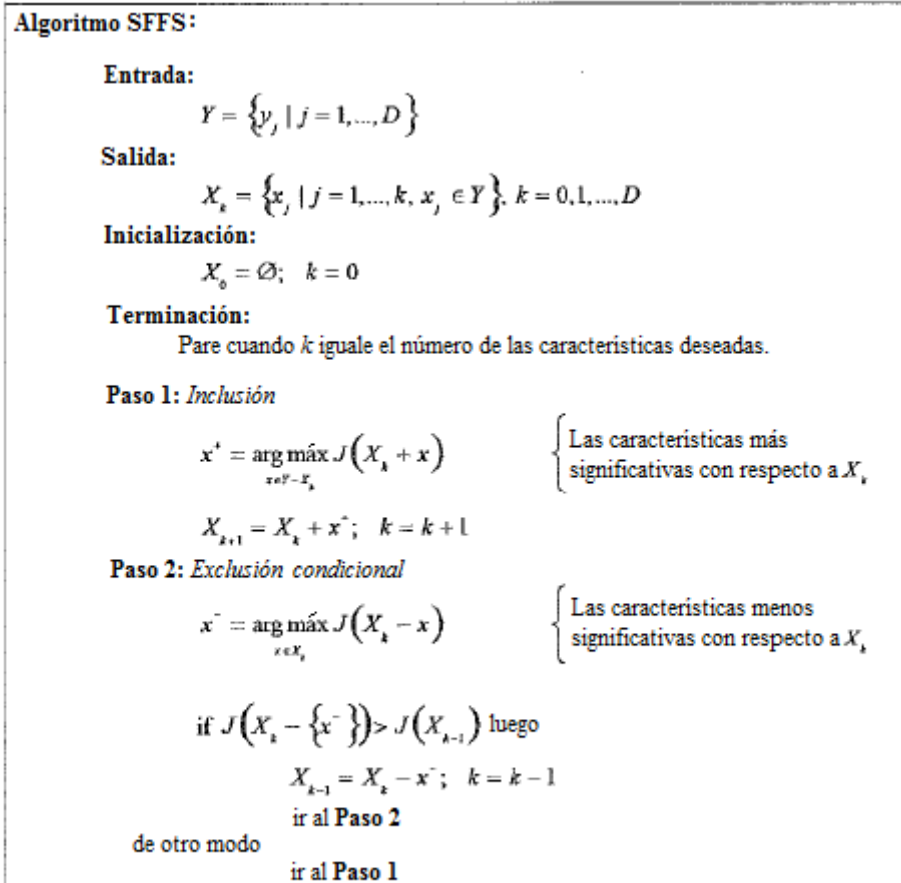


FIGURA 5

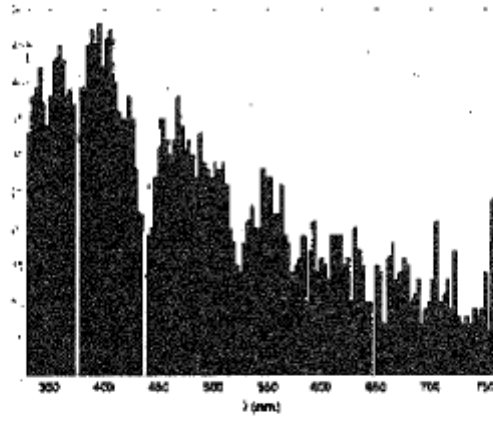


FIGURA 6A

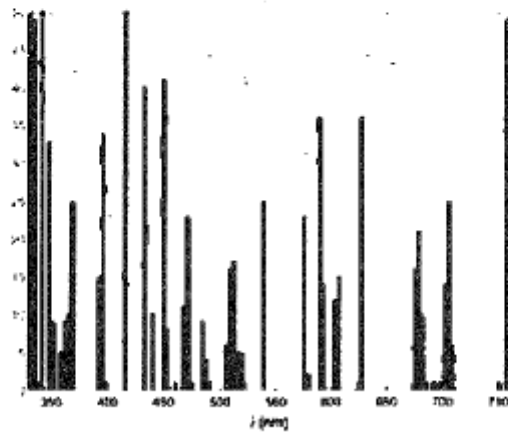


FIGURA 6B

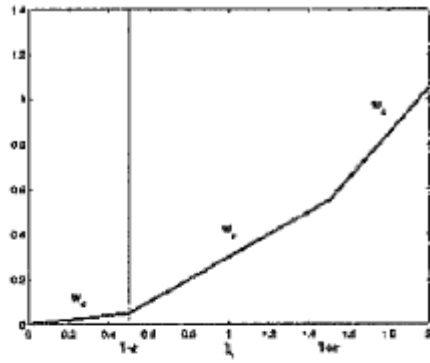


FIGURA 7A

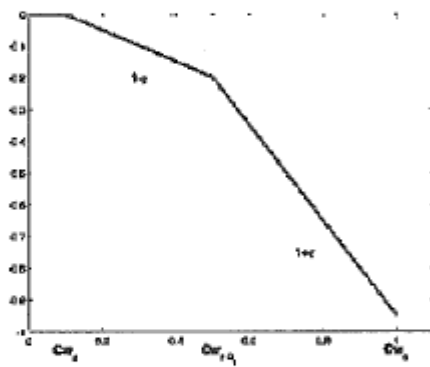


FIGURA 7B

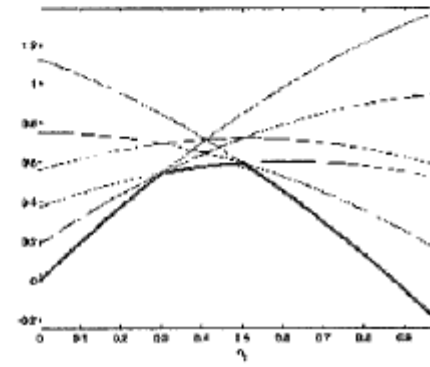


FIGURA 8

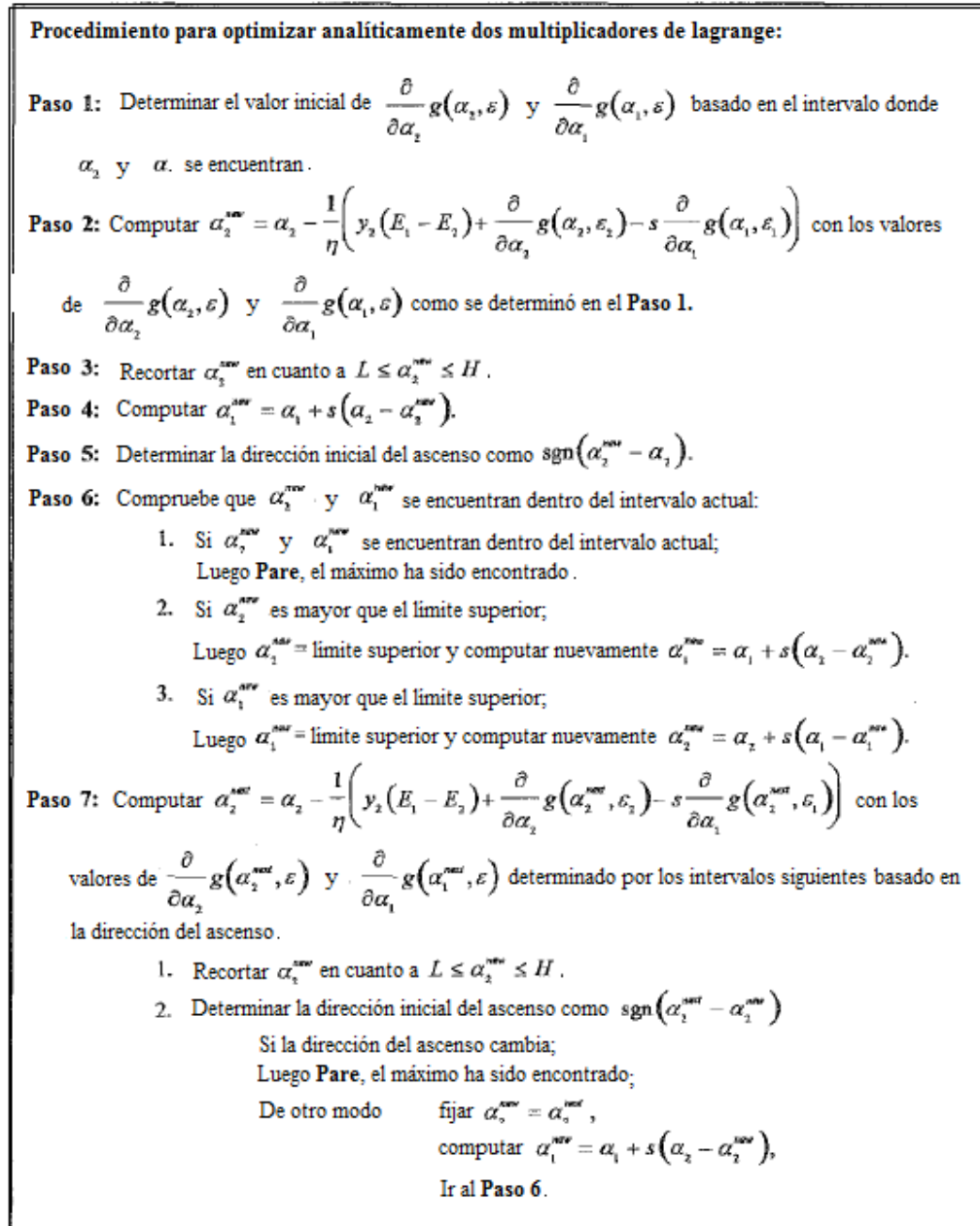


FIGURA 9

Violaciones de KKT:

$$\text{Dejar } R_i = y_i(\mathbf{w}^T \mathbf{x}_i - b) - 1,$$

- i. Si $\alpha_i > 0$ y $R_i > 0$
- ii. Si $\alpha_i < Cw_c$ y $R_i < 0$
- iii. Si $Cw_c < \alpha_i < Cw_r$ y $R_i \neq -(1-\varepsilon)$
- iv. Si $Cw_r < \alpha_i < Cw_c$ y $R_i \neq -(1+\varepsilon)$
- v. Si $\alpha_i = Cw_c$ y $R_i \geq 0$ o $R_i \leq -(1-\varepsilon)$
- vi. Si $\alpha_i = Cw_r$ y $R_i \geq -(1-\varepsilon)$ o $R_i \leq -(1+\varepsilon)$
- vii. Si $\alpha_i = Cw_c$ y $R_i \geq -(1+\varepsilon)$

FIGURA 10

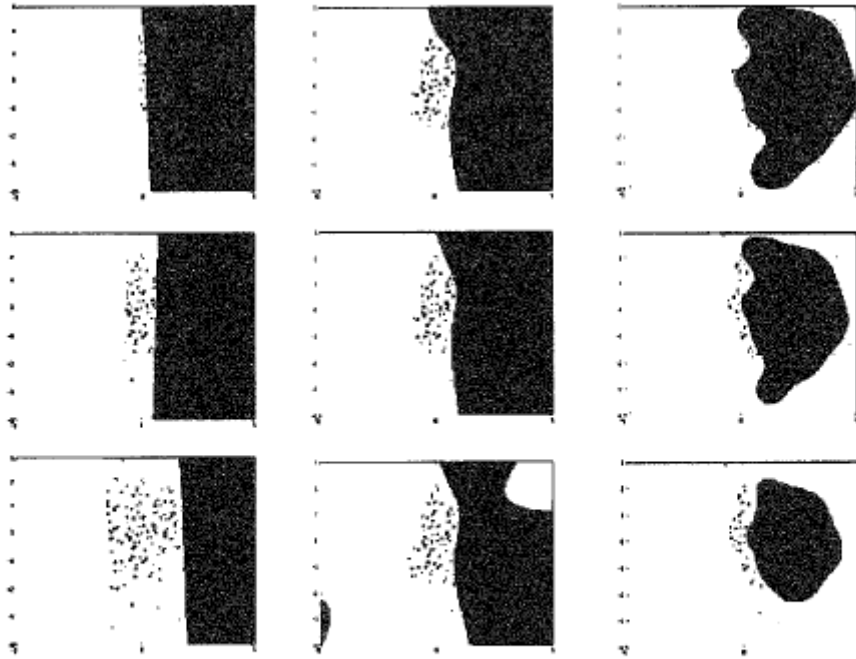


FIGURA 11

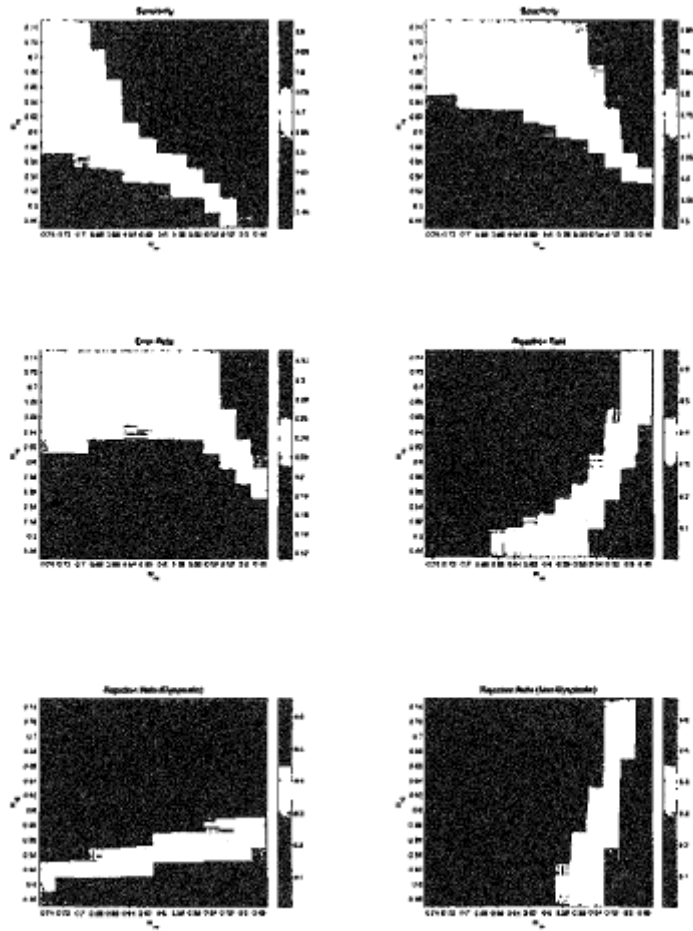


FIGURA 12

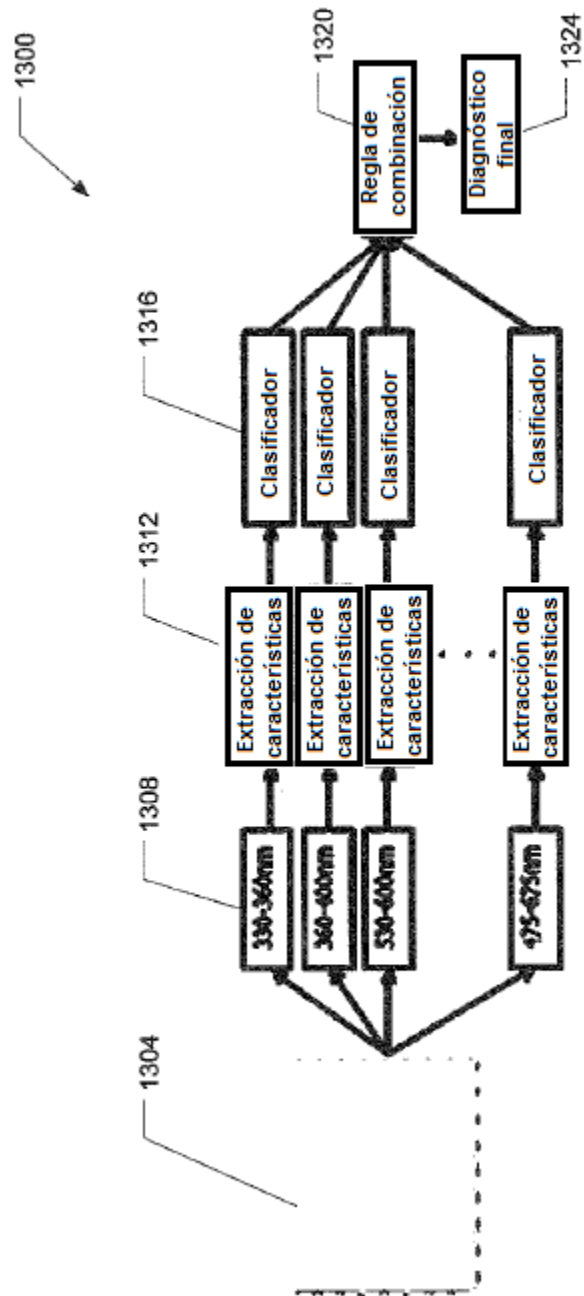


FIGURA 13

Algoritmo AdaBoost de capacitación

Entrada: Muestras de capacitación (\mathbf{x}_i, y_i) para $i = 1, \dots, l$

Inicializar: $w_i(\mathbf{x}_i) = 1/l$ para $i = 1, \dots, l$

Hacer para $t = 1, \dots, T$

1. Capacitar al alumno de la base con respecto a las muestras ponderadas fijar $\{\mathbf{x}_i, y_i, w_i\}$ y obtener hipótesis $g_t : \mathbf{x} \mapsto \{\pm 1\}$
2. Calcular el error de capacitación e_t de g_t :

$$e_t = \sum_{i=1}^l w_i(\mathbf{x}_i) I(g_t(\mathbf{x}_i) \neq y_i),$$

abortar si $e_t = 0$ o $e_t \geq 1/2 - \Delta$, donde Δ es una constante pequeña

3. Fijar

$$b_t = \log \frac{(1 - e_t)}{e_t}$$

4. Actualizar los pesos w_t :

$$w_{t+1}(\mathbf{x}_i) = w_t(\mathbf{x}_i) \exp\{-b_t I(g_t(\mathbf{x}_i) \neq y_i)\} / Z_t$$

donde Z_t es una constante de normalización, tal como $\sum_{i=1}^l w_{t+1}(\mathbf{x}_i) = 1$.

Salida: Hipótesis final

$$f(\mathbf{x}) = \sum_{i=1}^T c_i g_i(\mathbf{x}), \quad \text{donde} \quad c_i = \frac{b_i}{\sum_{i=1}^T |b_i|}$$

FIGURA 14

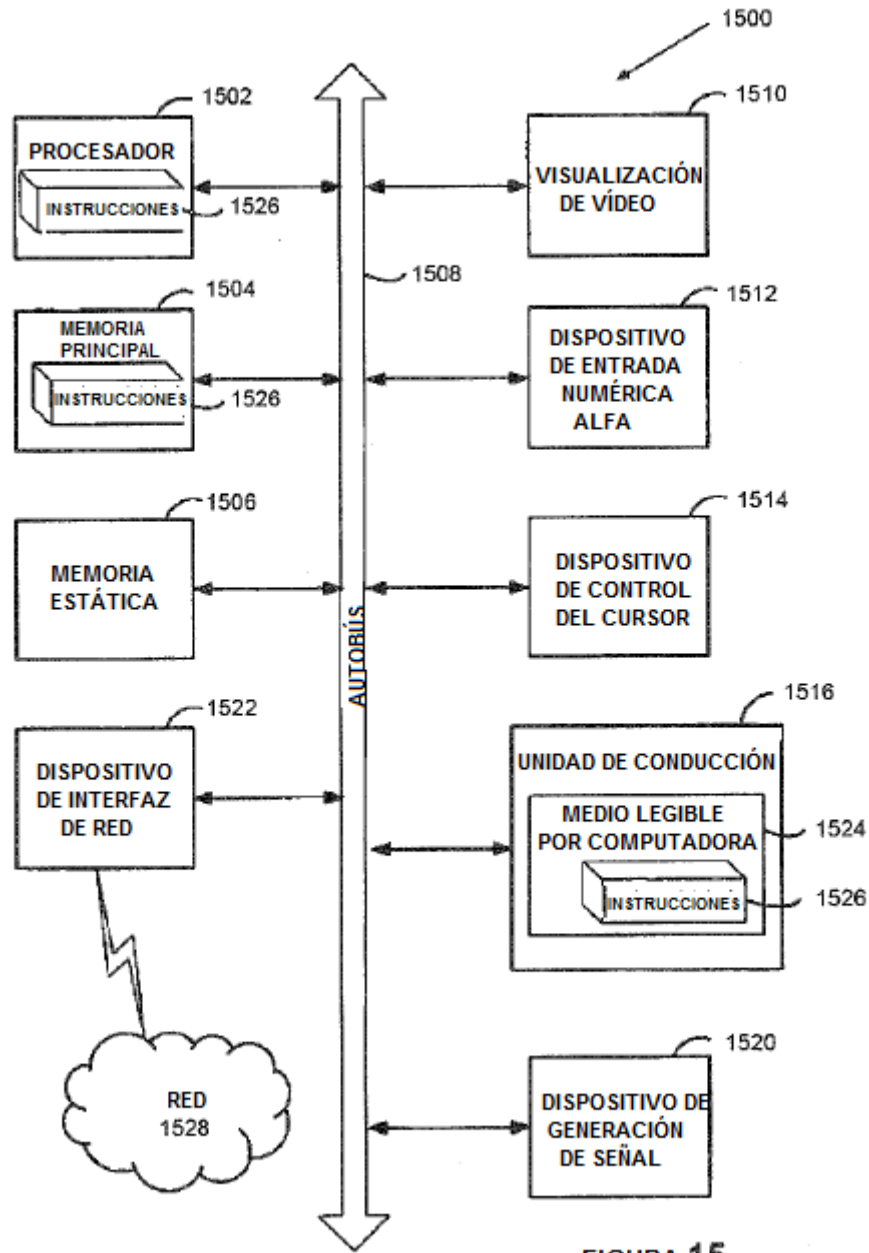


FIGURA 15