(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 741 050**

(51) Int. CI.:

G01N 30/86 (2006.01)
G16H 50/20 (2008.01)
G16H 50/70 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **12.02.2016 PCT/GB2016/050344**

(87) Fecha y número de publicación internacional: **18.08.2016 WO16128764**

(96) Fecha de presentación y número de la solicitud europea: **12.02.2016 E 16704919 (6)**

(97) Fecha y número de publicación de la concesión europea: **15.05.2019 EP 3256848**

(54) Título: **Método para crear un clasificador indicativo de una presencia de una condición médica**

(30) Prioridad:

13.02.2015 GB 201502447

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
07.02.2020

(73) Titular/es:

THE UNIVERSITY OF LIVERPOOL (100.0%)
Foundation Building 765 Brownlow Hill
Liverpool, L69 7ZX, GB

(72) Inventor/es:

PROBERT, CHRISTOPHER SIMON JOHN y
AGGIO, RAPHAEL BASTOS MARESCHI

(74) Agente/Representante:

SÁEZ MAESO, Ana

DESCRIPCIÓN

Método para crear un clasificador indicativo de una presencia de una condición médica

La presente invención se relaciona con un método para determinar una presencia de una condición médica en un sujeto. En particular, aunque no exclusivamente, algunos aspectos de la presente invención se relacionan con un método para determinar una presencia de cáncer, incluyendo cáncer de próstata, en un sujeto. La presente invención proporciona un método para crear un clasificador indicativo de una presencia de una condición médica en un sujeto.

Antecedentes

VAN BERKEL ET AL, Journal of Chromatography B, 861(2008)101-107 divulga un método de clasificación con base en el análisis de compuestos orgánicos volátiles de aire exhalado humano usando GC/TOF-MS.

El cáncer de próstata es la segunda enfermedad más común en el mundo para los hombres con aproximadamente 1,111,000 casos nuevos cada año. Muchos hombres con síntomas de flujo de salida de vejiga a menudo son investigados por cáncer de próstata cuando se descubre que tienen niveles elevados de PSA en suero. Sin embargo, los niveles de PSA carecen de especificidad y, por consiguiente, estos hombres tienen que someterse a pruebas invasivas para confirmar o refutar el diagnóstico de cáncer de próstata. En muchos, el cáncer no se encuentra. Esto a menudo deja a los hombres preocupados, en vez de tranquilizados, y puede seguir un ciclo interminable de mediciones repetidas de nivel de PSA. Actualmente, el PSA no se considera un marcador de diagnóstico y no se ha aprobado para uso en programas de detección en la mayoría de los países. El cáncer de vejiga es el noveno cáncer más común en el mundo y el más costoso de manejar. No hay biomarcadores aprobados para el seguimiento y se realizan cistoscopias repetidas que son invasivas, costosas y no exentas de riesgos. La enfermedad de intestino inflamatorio (IBD) es una enfermedad gastrointestinal crónica causada por una respuesta inmune aberrante en el intestino, mientras que el síndrome de intestino irritable (IBS) es un trastorno del tracto digestivo sin causa conocida. Existe una necesidad clínica apremiante de un mejor biomarcador que pueda usarse para el diagnóstico y detección de condiciones médicas incluyendo cáncer de próstata, IBD e IBS. Ahorraría dinero a los proveedores de atención médica, penalidades para paciente, y también aceleraría el tratamiento muy necesario para el paciente.

Es un objeto de realizaciones de la invención mitigar al menos uno o más de los problemas de la técnica anterior.

VAN BERKEL ET AL, Journal of Chromatography B, 861 (2008) 101-107 divulga la construcción de un clasificador para la identificación de VOCs en muestras de aire exhalado. El método se prueba en una población de estudio de fumadores/no fumadores.

Declaración de la invención

De acuerdo con la presente invención, se proporciona un método para determinar una presencia de una condición médica en un sujeto, de acuerdo con la reivindicación 1. Aspectos adicionales de la invención se definen en las reivindicaciones 2-14.

Breve descripción de los dibujos

Las realizaciones de la invención se describirán ahora solo a modo de ejemplo, con referencia a las figuras acompañantes, en las que:

La figura 1 muestra un método de acuerdo con una realización de la invención;

La figura 2 muestra un sistema para realizar el método de la invención;

La figura 3 muestra una ilustración de datos de cromatograma;

La figura 4 muestra una ilustración de datos de cromatograma invertidos en una realización de la invención;

La figura 5 muestra datos de cromatograma preprocesados en una realización de la invención;

La figura 6 muestra datos de cromatograma normalizados en una realización de la invención;

La figura 7 muestra datos de cromatograma alineados en una realización de la invención;

La figura 8 muestra un método para seleccionar la muestra de cromatograma de referencia para la alineación de datos en una realización de la invención;

La figura 9 muestra un método para alinear datos de cromatograma en una realización de la invención;

La figura 10 muestra los coeficientes de ondícula determinados para datos de cromatograma en una realización de la invención;

La figura 11 muestra datos de cromatograma transformados en una realización de la invención;

La figura 12 ilustra un método para determinar una presencia de una condición médica en un sujeto de acuerdo con un aspecto de la invención; y

La figura 13 muestra un método para alinear datos de cromatograma recibidos.

Descripción detallada de realizaciones de la invención

La figura 1 ilustra un método 100 de acuerdo con una realización de la invención. El método 100 es un método para crear un clasificador indicativo de si un sujeto tiene una o más condiciones médicas. Las condiciones médicas pueden comprender uno o más de cáncer, que comprenden cáncer de vejiga y/o próstata, enfermedad de intestino irritable (IBD), síndrome de intestino irritable (IBS), una presencia de una o más bacterias predeterminadas tal como Clostridium difficile (C-dif), uno o más parásitos predeterminados, uno o más hongos predeterminados. El método 100 es un método basado en ordenador para crear el clasificador y almacenar el clasificador en un medio legible por ordenador, tal como medio legible por ordenador no transitorio.

El método puede realizarse mediante un aparato 200 de acuerdo con la figura 2. El aparato 200 comprende una unidad 210 de control que comprende una unidad 220 de procesamiento y una unidad 230 de memoria. El aparato 210 está dispuesto para recibir datos de cromatograma de una unidad 240 de detección. Los datos de cromatograma son indicativos de una presencia de compuestos volátiles en una muestra tomada u obtenida de un sujeto. La muestra puede ser una muestra de aliento, orina o heces del sujeto, aunque se comprenderá que esta lista no es exhaustiva.

La unidad de detección puede comprender uno o más sensores de Oxido Metálico (MO). La unidad 240 de detección puede estar asociada con un aparato tal como se describe en el documento WO/2011/061308. El aparato 200 puede comprender una columna de cromatografía de gases acoplada a uno o más sensores. La columna puede estar asociada con un horno para calentar la columna de acuerdo con un protocolo predeterminado.

Los datos de cromatograma pueden comunicarse entre la unidad 240 de detección y la unidad 210 de control por medio de un canal de comunicación dedicado es decir una conexión eléctrica directa, o por medio de un canal de comunicación formado a través de una o más redes de ordenadores. Los datos de cromatograma pueden recibirse en la unidad 210 de control en la forma de uno o más archivos comprendiendo cada uno datos de cromatograma para una muestra respectiva.

Para producir los datos de cromatograma, la muestra se puede calentar de acuerdo con un protocolo predeterminado. El protocolo puede definir un período de calentamiento de la muestra a una o más temperaturas predeterminadas antes de muestrear un volumen predeterminado de gas de la muestra.

La temperatura inicial del horno se puede mantener a 40°C durante 13.4 minutos, aumentar a 100°C a una tasa de 5°C/min, mantener durante 30 minutos y enfriar a 40°C usando una rampa de temperatura de 10°C/minuto. Se comprenderá que se pueden usar otros protocolos para el calentamiento de horno.

La resistencia del sensor de MO se determina durante un período de tiempo. Los datos de cromatograma pueden comprender datos indicativos de una resistencia del uno o más sensores de MO a intervalos predeterminados tales como 0.5 segundos, aunque se comprenderá que pueden usarse otros intervalos.

La figura 3 ilustra datos de cromatograma en una realización de la invención.

La figura 3 comprende una gráfica de una pluralidad de ítems de datos de cromatograma de muestras respectivas. Los datos de cromatograma se trazan sobre el tiempo (eje x) e indican una resistencia (eje y) del sensor en cada tiempo de muestra respectivo. Los datos de cromatograma son recibidos por la unidad 210 de control en la etapa 105. Los datos de cromatograma pueden almacenarse en la unidad 230 de memoria de la unidad 210 de control.

Con el fin de crear el clasificador indicativo de si un sujeto tiene una o más condiciones médicas, los datos de cromatograma de una pluralidad de muestras se proporcionan a partir de sujetos que tienen la una o más condiciones médicas respectivas. El clasificador se basa en los datos de cromatograma de esos sujetos, como se explicará. De este modo se recibe un conjunto de datos de cromatograma de la pluralidad de muestras que tienen la una o más condiciones médicas en la etapa 105. Se proporciona un conjunto adicional de datos de cromatograma a partir de una pluralidad de muestras que no tienen la una o más condiciones médicas que pueden denominarse como un conjunto de control de datos de cromatograma.

En la etapa 110, las señales de resistencia de los datos de cromatograma recibidos en la etapa 105 se invierten con el fin de facilitar su procesamiento usando herramientas metabolómicas. Esta inversión se realiza individualmente para cada muestra usando la siguiente ecuación matemática:

$$x = |x - (\max(x) + 1)|$$

donde x contiene los valores de resistencia registrados para una única muestra. La figura 4 comprende una gráfica de los datos de cromatograma invertidos.

En la etapa 120 se procesan los datos de cromatograma recibidos. La etapa 120 comprende un proceso de eliminación de línea base. La línea base es un nivel de resistencia de línea base de los datos de cromatograma. La línea base puede ser contribuida como una mayoría, o solo por, una fase móvil. La fase móvil es el gas que transporta los metabolitos a través de una columna de un cromatograma de gases. En algunas realizaciones el gas puede ser aire sintético.

También se puede determinar un umbral en la etapa 120. En algunas realizaciones la línea base de los datos de cromatograma se elimina mediante un proceso de ajuste por mínimos cuadrados.

En algunas realizaciones la etapa 120 comprende además determinar un umbral de resistencia. El umbral de resistencia se define como un valor de resistencia promedio en los datos de cromatograma de una muestra menos la desviación estándar de sus valores de resistencia. Cualquier valor de resistencia inferior al umbral de resistencia se ajusta entonces a un valor predeterminado, que puede ser cero. La figura 5 ilustra los datos de cromatograma procesados de acuerdo con una realización de la etapa 120.

En la etapa 130 se normalizan los valores en los datos de cromatograma para cada muestra. En una realización, los valores de resistencia de una muestra se normalizan al dividir sus valores por el valor de resistencia más alto registrado para la muestra particular. La figura 6 ilustra los datos de cromatograma procesados de acuerdo con una realización de la etapa 130.

En la etapa 140 se selecciona una muestra de cromatograma de referencia para alineación de datos. La etapa 140 comprende seleccionar datos de cromatograma de referencia a partir de los datos de cromatograma proporcionados de la etapa 130. En algunas realizaciones seleccionar los datos de cromatograma de referencia comprende determinar un coeficiente indicativo de correlación entre cada par de datos de cromatograma. El coeficiente puede ser un coeficiente de correlación del producto-momento de Pearson, a menudo denominado como un coeficiente de Pearson, como será evidente para la persona experimentada.

Se ilustra un método 700 para seleccionar el cromatograma de referencia para alinear los datos de cromatograma en una realización de la invención en la figura 8.

Refiriéndose a la figura 8, en la etapa 705 se crean dos listas que contienen todas las muestras en la condición experimental 1, por ejemplo, muestras de cáncer. Una de estas listas puede denominarse como SampleListRef, mientras que la segunda lista puede denominarse como SampleListTest.

En la etapa 710 una muestra puede seleccionarse de manera aleatoria de SampleListRef, cargarse en la memoria y eliminarse de SampleListRef. Por claridad, esta muestra se describirá aquí como SampleRef.

En la etapa 715 una muestra puede seleccionarse de manera aleatoria de SampleListTest, cargarse en la memoria y eliminarse de SampleListTest. Por claridad, esta muestra se describirá aquí como SampleTest. En una primera iteración de etapas 710 y 715 las muestras seleccionadas pueden ser los primeros cromatogramas en el conjunto de datos. Por ejemplo cuando a todos los datos de cromatograma se les asigna una ID, se puede seleccionar un cromatograma que tenga un valor más bajo de ID en la primera iteración de etapas 710 y 715.

En las etapas 720 y 725 el coeficiente de correlación de Pearson entre SampleRef y SampleTest se determina y almacena en una matriz, que puede denominarse R.

En las etapas 730 a 765, la SampleRef se desplaza un número predeterminado de puntos de muestreo con un coeficiente de correlación con SampleTest que se calcula después de cada desplazamiento de punto de muestreo y el coeficiente de correlación resultante almacenado en la matriz R. Se apreciará que la SampleRef, en algunas realizaciones, se desplazará en ambas direcciones de punto de tiempo positiva y negativa con respecto a la SampleTest. En una realización la ventana de desplazamiento es ±15 puntos de muestreo, aunque se comprenderá que se pueden elegir otros tamaños de ventana de desplazamiento.

Cuando la SampleRef se ha desplazado hasta la extremidad o extremidades de la ventana de desplazamiento, el método avanza a la etapa 775. Se apreciará que al llegar a la etapa 775, en algunas realizaciones, cada cromatograma se asocia con coeficientes P como:

$$P = (2s+1) \times (n-1)$$

donde s es una magnitud de la ventana de desplazamiento, tal como 15 (de ahí que 2s calcula el rango de desplazamientos de negativo a positivo), y n es el número de muestras en condición experimental 1. Por lo tanto, en una realización, cada dato de cromatograma está asociado con 31 coeficientes de correlación para cada uno de los datos de cromatograma restantes en la condición experimental 1.

En la etapa 775 se obtiene el valor máximo en la matriz R, se almacena en una nueva matriz denominada M y se eliminan o restablecen los contenidos de R. Las etapas 715 a 775 se repiten hasta que la SampleListTest esté vacía y el método avanza a la etapa 785.

En la etapa 785 se calcula el valor medio de todos los valores almacenados en M, se almacena en una matriz denominada C junto con la información que identifica la muestra de referencia, tal como la ID de SampleRef y los contenidos de M se eliminan. Las etapas 710 a 785 se repiten hasta que SampleListRef está vacía y el método avanza a la etapa 795. En la etapa 795 la muestra asociada con el valor positivo más alto en matriz C se determina como muestra de referencia para la alineación de cromatograma. La etapa 795 puede comprender almacenar la ID asociada con el cromatograma seleccionado como la muestra de cromatograma de referencia para permitir que otros datos de cromatograma se alineen en un momento posterior, como se explicará.

Volviendo a la figura 1, en la etapa 150 se alinean los datos de cromatograma. El objetivo de alineación es asegurar que se comparen las mismas características entre las muestras de las diferentes clases de datos o condiciones médicas bajo análisis. La etapa 150 comprende alinear los datos de cromatograma en relación con la muestra de cromatograma de referencia seleccionada en la etapa 140. Se ilustra un método 800 para alinear los datos de cromatograma de acuerdo con una realización de la invención en la figura 9.

Refiriéndose a la figura 9, en el método 800 todos los datos de cromatograma se alinean en relación con el cromatograma de referencia seleccionado en la etapa 140 del método 100 de figura 1.

En la etapa 805 la muestra de cromatograma de referencia seleccionada en la etapa 140 se carga en la memoria. Por claridad, la muestra de cromatograma de referencia se describirá aquí como RefSample. En la etapa 810 se crea una lista que contiene todas las muestras en el uno o más conjuntos de datos bajo análisis, por ejemplo, muestras de Cáncer y Control. Por claridad, esta lista se describirá aquí como SamplesToAlign.

En la etapa 815 se carga una muestra aleatoria de SamplesToAlign. Por claridad, esta muestra se describirá aquí como SampleAlign. En las etapas 820 a 870 la SampleAlign se desplaza un número predeterminado de puntos de muestreo con un único coeficiente de correlación que se calcula entre RefSample y SampleAlign después de cada desplazamiento de punto de muestreo y el coeficiente de correlación resultante almacenado en la matriz R. En una realización la ventana de desplazamiento es ±15 puntos de muestreo, aunque se comprenderá que se puede elegir otro número de puntos de tiempo. Se apreciará que la SampleAlign, en algunas realizaciones, se desplazará en ambas direcciones de punto de tiempo positivo y negativo con respecto a la RefSample. Cuando la SampleAlign se ha desplazado hasta la extremidad o extremidades de la ventana de desplazamiento, el método avanza a la etapa 875. Se apreciará que al llegar a la etapa 875 en algunas realizaciones la SampleAlign se asocia con coeficientes P como:

$$P = 2s + 1$$

donde s es una magnitud de la ventana de desplazamiento de tiempo, tal como 15 (de ahí que 2s calcula el rango de desplazamientos de tiempo de negativo a positivo). Por lo tanto, en una realización, SampleAlign está asociada con 31 coeficientes de correlación. En la etapa 875 el punto de muestreo de desplazamiento asociado con el valor más alto en R se determina y almacena como SamplingPointsToShift. En la etapa 880 la SampleAlign se desplaza el número de puntos de muestreo definidos en SamplingPointsToShift y se eliminan los contenidos de matriz R. Las etapas 815 a 880 se repiten hasta que la lista de SamplesToAlign esté vacía. La figura 7 ilustra datos de cromatograma alineados de acuerdo con una realización de la etapa 150.

Volviendo a la figura 1, en la etapa 160 los valores de los datos de cromatograma alineados se transforman en coeficientes de ondícula usando una ondícula madre del tipo conocido como sombrero mexicano, que también puede conocerse como una ondícula Ricker. Se pueden usar otras ondículas madre. En una realización los coeficientes de ondícula pueden determinarse usando una pluralidad de escalas de ondícula madre del tipo conocido como sombrero mexicano. La pluralidad de escalas puede ser escalas entre límites inferior y superior. En una realización los límites superior e inferior pueden ser 100 y 1, respectivamente. En una realización puede determinarse un coeficiente en cada escala de enteros entre los límites inferior y superior. Los coeficientes pueden determinarse como un módulo de un coeficiente calculado. Es decir, los valores de los datos de cromatograma para cada muestra se convierten al módulo de sus coeficientes de ondícula usando la escala de ondícula madre del tipo conocido como sombrero mexicano, aunque se pueden usar los valores originales extraídos mediante una ondícula madre del tipo conocido como sombrero mexicano. Los coeficientes de ondícula entonces se almacenan para uso futuro, como se explicará. Uno de los valores de escala de ondícula se elige como una mejor coincidencia para los datos de cromatograma. La mejor coincidencia puede ser la escala de ondícula que tiene la precisión de clasificación más alta, como se explicará. La precisión de cada escala de ondícula puede determinarse con base en uno o más de precisión mínima, mediana, media y máxima de un proceso de validación. La figura 10 ilustra los datos de cromatograma transformados en coeficientes de ondícula de acuerdo con una realización de la etapa 160.

En la etapa 170 uno o más de los procesos de transformación de registro, rango y SpatialSign se aplican a los datos de cromatograma. En una realización, antes de los procesos de transformación de registro, rango y SpatialSign, cada valor de los datos de cromatograma tiene un valor predeterminado, tal como el valor 1 agregado. Los datos de cromatograma pueden entonces someterse a transformación de registro usando un logaritmo natural como base, aunque se comprenderá que se pueden usar otros valores de base para la transformación de registro. En una realización la transformación de rango se aplica entonces para establecer los valores de los datos de cromatograma en un rango predeterminado tal como un rango entre 0 y 1. La transformación de rango puede determinar un valor transformado x_t en cada punto de tiempo de los datos de cromatograma donde x es un valor de datos de los datos de

cromatograma y min(x) y max(x) son valor mínimo y máximo de los datos de cromatograma, respectivamente. La transformación de rango se puede realizar usando la ecuación:

$$x_t = \frac{(x - \min(x))}{(\max(x) - \min(x))}$$

En algunas realizaciones se puede aplicar una transformada adicional que puede conocerse como una transformada de SpatialSign como se describe en S. Serneels, E. De Nolf, P. J. Van Espen, Spatial sign preprocessing: A simple way to impart moderate robustness to multivariate estimators. Journal of Chemical Information and Modeling 46, 1402-1409 (2006). La figura 11 ilustra los datos de cromatograma transformados de acuerdo con una realización de la etapa 170.

En la etapa 180 se seleccionan una o más características de los datos de cromatograma. La una o más características se seleccionan para ser indicativas de la presencia de la una o más condiciones médicas. En realizaciones de la invención, la una o más características se seleccionan mediante un algoritmo de selección de características usando bosque aleatorio. En este algoritmo, los árboles de decisión se desarrollan con basa en diferentes conjuntos de muestras y el bosque aleatorio se usa para calcular una pérdida de precisión de clasificación cuando los valores de características se permutan de manera aleatoria entre conjuntos de muestras. Entonces se seleccionan una o más características asociadas con una pérdida de precisión de clasificación.

En algunas realizaciones de la invención, uno de dos algoritmos diferentes conocidos como Boruta y rfe basados en bosques aleatorios se aplican en la etapa 180 con el fin de seleccionar las características para ser usadas. El algoritmo Boruta involucra el desarrollo de árboles de decisión con base en diferentes conjuntos de muestras. Entonces se aplica bosque aleatorio para calcular la pérdida de precisión de clasificación cuando los valores de características se permutan de manera aleatoria entre conjuntos de muestras. Las características asociadas con la pérdida de precisión se seleccionan entonces como características indicativas. El algoritmo rfe funciona de manera similar a Boruta, sin embargo, elimina las características que no producen cambios en el nivel de precisión, en vez de seleccionar características que producen pérdida de precisión. Los algoritmos Boruta y rfe se describen en “Feature Selection with the Boruta Package" Journal of Statistical Software 36(11): 1-13; y Anderssen, E., K. Dyrstad, F. Westad and H. Martens (2006), "Reducing over-optimism in variable selection by cross-model validation" Chemometrics and Intelligent Laboratory Systems 84(1-2): 69-74. En la etapa 180 la una o más características seleccionadas se almacenan para uso posterior.

En la etapa 190 se determina un clasificador. El clasificador es para clasificar una muestra ya sea como siendo una muestra de un sujeto que tiene la una o más condiciones médicas o una muestra que no tiene la una o más condiciones médicas. El clasificador se puede determinar de acuerdo con uno de: análisis discriminante lineal (LDA); mínimos cuadrados parciales (PLS); bosque aleatorio, k vecindario más cercano (KNN); máquina de vectores de soporte (SVM) con núcleo de función de base radial (SVMRadial); SVM con núcleo de función de base lineal (SVMLineal); y SVM con núcleo de función de base polinomial (SVMPoly). El clasificador puede determinarse usando, por ejemplo, un paquete de software tal como R Package Caret (Kuhn, M., caret: Classification and Regression Training. 2014).

La construcción y prueba del clasificador en el mismo conjunto de datos puede producir resultados sesgados y desmedidamente optimistas debido a potencial sobreajuste. Por lo tanto se puede usar en la etapa 190 un proceso de validación para prevenir tal sobreajuste. El proceso de validación puede ser uno de validación cruzada k-veces repetida y validación cruzada doble repetida. En particular, en realizaciones de ejemplo de la invención se usan dos procesos de validación: 30 repeticiones de validación cruzada de 10 veces y 30 repeticiones de la validación cruzada doble de 3 veces con un bucle interno de 10 veces repetido 5 veces. Además, estos dos procesos de validación cruzada se repiten en los mismos conjuntos de datos, sin embargo, aplicando una permutación aleatoria Monte Carlo de etiquetas de clase en cada repetición.

Como se menciona en la descripción anterior de la etapa 160, el método 100 se repite para una pluralidad de escalas de ondícula. La escala que produce la precisión de clasificación más alta entonces se selecciona como la mejor coincidencia para los datos de cromatograma procesados. Como un resultado de realizaciones del método 100 ilustrado en la figura 1, se produce un clasificador que es capaz de clasificar datos de cromatograma como originados de una muestra que tiene la una o más condiciones médicas o que no tiene la una o más condiciones médicas.

La figura 12 ilustra un método 1000 para determinar una presencia de una condición médica en un sujeto de acuerdo con un aspecto de la invención. El método se realiza sobre una muestra tomada del sujeto. Los datos de cromatograma pueden proporcionarse desde un aparato como se describe anteriormente con referencia a la figura 2. La misma puede ser material excretado del sujeto. La muestra puede ser una muestra de aliento, orina o heces del sujeto, aunque se comprenderá que esta lista no es exhaustiva. Como se anotó anteriormente, la condición médica puede comprender uno o más de cáncer, que comprende cáncer de vejiga y/o próstata, enfermedad de intestino irritable (IBD), síndrome de intestino irritable (IBS), una presencia de una o más bacterias predeterminadas tal como Clostridium difficile (C-dif), uno o más parásitos predeterminados, uno o más hongos predeterminados.

Un número de etapas del método 1000 se describen como en conjunto con el método 100 ilustrado en la figura 1. Por lo tanto se omitirá la descripción repetida de estas etapas y el lector se referirá a la descripción asociada con la etapa equivalente en la figura 1.

En la etapa 1050 se reciben los datos de cromatograma. Por claridad, los datos de cromatograma recibidos se describirán aquí como newSample. En algunos aspectos, como se describió previamente, en la etapa 1100 la newSample tiene su línea base eliminada y sus valores de datos se normalizan en la etapa 1150. En la etapa 1200 la newSample entonces se alinea. Se ilustra un método 2000 para alinear la newSample en la figura 13.

Refiriéndose a la figura 13, en la etapa 2050 la muestra de cromatograma de referencia seleccionada en la etapa 140 de método 100 se carga en la memoria. Por claridad, los datos de cromatograma de referencia se describirán aquí como RefSample. En la etapa 2100 los datos de cromatograma de newSample se cargan en la memoria.

En las etapas 2150 a 2650 el tiempo de retención de la newSample se desplaza un número predeterminado de puntos de muestreo con un único coeficiente de correlación que se calcula entre RefSample y newSample después de cada desplazamiento de punto de muestreo y el coeficiente de correlación resultante almacenado en la matriz R. En una realización la ventana de desplazamiento es ±15 puntos de muestreo, aunque se comprenderá que se puede elegir otro número de puntos de desplazamiento. Se apreciará que los datos de cromatograma de newSample, en algunas realizaciones, se desplazarán en ambas direcciones de punto de tiempo positiva y negativa con respecto a la RefSample. Cuando los datos de cromatograma de newSample se han desplazado hasta la extremidad o extremidades de la ventana de desplazamiento, el método avanza a la etapa 2700. Se apreciará que al llegar al llegar a la etapa 2700 en algunas realizaciones los datos de cromatograma de newSample se asocian con coeficientes P como:

$$P = 2s + 1$$

donde s es una magnitud de la ventana de desplazamiento de tiempo, tal como 15 (de ahí que 2s calcula el rango de desplazamientos de tiempo de negativo a positivo). Por lo tanto, en una realización, los datos de cromatograma de newSample están asociados con 31 coeficientes de correlación. En la etapa 2700 el punto de muestreo asociado con el coeficiente en R se determina y almacena como SamplingPointsToShift. En la etapa 2750 los datos de cromatograma de newSample se desplazan el número de puntos de muestreo definidos en SamplingPointsToShift para alinear los datos de cromatograma de nueva muestra con los datos de cromatograma de referencia del método ilustrado en la figura 1.

Volviendo a la figura 12, en la etapa 1250 los datos de cromatograma de newSample se transforman en coeficientes de ondícula usando una ondícula del tipo conocido como sombrero mexicano y una escala predeterminada. La escala predeterminada puede ser esa escala determinada para haber producido una precisión más alta en método 100 descrito con referencia a la figura 1, como se explicó anteriormente.

En la etapa 1300 se cargan los coeficientes de ondícula producidos por una escala de ondícula predeterminada, que puede ser la escala de ondícula asociada con una precisión más alta y almacenada en la etapa 160 de método 100. El valor de la escala de ondícula usada en la etapa 160 de método 100 es el mismo como el valor de la escala de ondícula usada en la etapa 1250 de método 1000. Por claridad, los coeficientes de ondícula producidos en la etapa 160 de método 100 se describirán aquí como datos preProcessed. En la etapa 1350 la newSample se combina con los datos preProcessed en un único conjunto de datos denominado transformData.

En la etapa 1400 los transformData se transforman entonces como se describe en la etapa 170 de método 100. Las características definidas en la etapa 180 de método 100 se seleccionan entonces de transformData. La newSample se aísla de los transformData y se predice o clasifica mediante el modelo determinado en la etapa 190 de método 100.

Los métodos descritos anteriormente se aplicaron a dos conjuntos de datos diferentes. Primero, se aplicaron para clasificar muestras de orina de pacientes con cáncer de próstata, cáncer de vejiga y pacientes con una mezcla de síntomas urológicos - síntomas de hematuria y/o prostáticos (Control). La tabla 1 muestra los resultados de la validación cruzada doble 30 veces repetida para los siete clasificadores construidos. SVMRadial pudo clasificar el cáncer de próstata y muestras de cáncer de vejiga con 89.6% y 96.2% de precisión, respectivamente. Las muestras de cáncer de próstata y vejiga se diferenciaron con 93.5% de precisión. Entonces, los métodos descritos anteriormente se aplicaron para clasificar muestras de heces de pacientes con enfermedad intestinal inflamatoria (IBD), síndrome de intestino irritable (IBS) y donantes sanos (Control). Las tablas 2 y 3 muestran los resultados de la validación cruzada doble 30 veces repetida para los siete clasificadores construidos. IBD e IBS se diferenciaron de las muestras de Control con 88.9% y 94.4%, respectivamente. Las muestras de IBD se diferenciaron de muestras de IBS con 85.2% de precisión. Las muestras de IBD se diferenciaron de muestras de no IBD con 84.9% de precisión. Las muestras de IBS se diferenciaron de muestras de no IBS con 92.1% de precisión. Finalmente, las muestras de Control se diferenciaron de las muestras de no Control con 86.8% de precisión. De este modo se puede apreciar que la invención permite determinar si una muestra es de una persona que tiene una condición predeterminada con precisión.

Los métodos que forman realizaciones de la invención pueden implementarse por ordenador.

Se apreciará que realizaciones de la presente invención pueden realizarse en la forma de hardware, software o una combinación de hardware y software. Cualquier software tal puede almacenarse en la forma de almacenamiento volátil o no volátil tal como, por ejemplo, un dispositivo de almacenamiento como una ROM, ya sea borrable o reescribible o no, o en la forma de memoria tal como, por ejemplo, RAM, chips de memoria, o circuitos integrados o en un medio legible de manera óptica o magnéticamente tal como, por ejemplo, un CD, DVD, disco magnético o cinta magnética. Se apreciará que los dispositivos de almacenamiento y medios de almacenamiento son realizaciones de almacenamiento legibles por máquina que son adecuados para almacenar un programa o programas que, cuando se ejecutan, implementan realizaciones de la presente invención. Por consiguiente, se proporciona un programa que comprende código para implementar un método como se reivindica y un almacenamiento legible por máquina que almacena un programa tal. Aún más, realizaciones de la presente invención pueden transmitirse electrónicamente a través de cualquier medio tal como una señal de comunicación transportada a través de una conexión por cable o inalámbrica.

La invención no está restringida a los detalles de ninguna de las realizaciones anteriores. Las reivindicaciones no deben interpretarse para cubrir simplemente las realizaciones anteriores, sino también cualquier realización que esté dentro del alcance de las reivindicaciones.

Tabla 1

Próstata vs Control									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMRadial	89.6	0.5	90.7	85.6	0.8	85.0	92.7	0.5	92.0
SVMPoly	88.8	0.4	88.6	85.5	0.8	85.0	91.4	0.6	91.7
RF	88.3	0.4	88.6	82.0	0.8	84.2	93.3	0.6	93.9
PLS	87.7	0.5	88.6	85.6	0.8	85.0	89.4	0.7	91.7
LDA	87.7	0.5	88.6	85.4	0.8	85.0	89.6	0.7	91.7
SVMLineal	83.8	0.5	83.7	81.6	1.0	82.1	85.5	0.7	87.5
KNN	83.0	0.5	83.0	81.7	0.8	84.2	84.0	0.7	83.7
Vejiga vs Control									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMPoly	96.2	0.3	96.9	87.2	1.2	87.5	99.2	0.2	100.0
SVMRadial	96.2	0.3	96.9	85.0	1.1	87.5	99.9	0.1	100.0
PLS	94.4	0.4	93.9	86.3	1.1	87.5	97.1	0.4	98.0
LDA	93.6	0.5	93.8	87.4	1.1	87.5	95.7	0.5	95.8
SVMLineal	93.6	0.3	93.8	85.6	1.1	87.5	96.3	0.4	96.0
KNN	91.0	0.5	90.8	81.3	1.4	87.5	94.2	0.5	95.8
RF	86.8	0.4	87.5	46.8	1.6	50.0	100.0	0.0	100.0
Vejiga vs Próstata									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		

Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMPoly	93.5	0.4	92.9	83.5	1.1	87.5	97.6	0.4	100.0
SVMRadial	93.0	0.4	92.9	82.8	1.1	87.5	97.2	0.4	100.0
SVMLineal	91.8	0.5	92.6	85.6	1.5	87.5	94.4	0.5	94.7
KNN	91.2	0.4	92.6	81.9	1.2	87.5	95.1	0.5	95.0
PLS	90.9	0.6	92.6	80.0	1.5	87.5	95.3	0.5	95.0
RF	89.5	0.5	88.9	70.3	1.5	75.0	97.5	0.3	100.0
LDA	87.8	0.7	88.9	77.9	1.6	75.0	91.9	0.7	94.7

Tabla 2

IBD vs Control									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMPoly	88.9	0.6	88.0	94.1	0.8	93.3	80.8	1.2	80.0
SVMRadial	86.6	0.7	87.5	92.8	0.9	93.3	77.0	1.3	77.8
SVMLineal	86.5	0.6	87.5	89.8	0.7	86.7	81.3	1.3	80.0
PLS	85.9	0.8	87.5	90.3	1.0	93.3	79.2	1.5	80.0
LDA	85.9	0.7	85.8	89.3	0.9	93.3	80.6	1.2	80.0
RF	84.9	0.6	84.0	95.6	0.5	100	68.2	1.5	70.0
KNN	82.4	0.7	83.3	91.9	0.8	93.3 :	67.6	1.5	70.0
IBS vs Control									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMRadial	94.4	0.6	94.4	93.9	1.0	100	94.9	0.8	100
SVMPoly	94.4	0.5	94.4	94.0	1.0	100	94.8	0.7	100
SVMLineal	93.4	0.6	94.4	93.2	1.2	100	93.6	0.7	90.0
PLS	92.9	0.7	94.4	90.1	1.1	87.5	95.3	0.8	100
RF	92.9	0.7	94.4	92.2	1.1	100	93.5	0.8	90.0
KNN	91.9	0.7	94.1	91.3	1.1	87.5	92.6	0.9	90.0
LDA	78.7	1.1	77.8	76.8	1.4	75.0	80.3	1.7	80.0
IBD vs IBS									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
RF	85.2	0.6	87.0	96.3	0.5	100	64.4	1.8	62.5

SVMRadial	82.2	0.7	82.6	90.7	0.9	93.3	66.1	1.8	62.5
SVMPoly	82.2	0.7	82.6	91.6	0.8	93.3	64.6	2.0	62.5
SVMLineal	81.6	0.8	82.6	85.6	1.1	86.7	74.0	1.7	75.0
PLS	80.3	0.8	82.6	89.0	0.8	86.7	64.0	1.7	62.5
KNN	77.7	0.8	78.3	91.7	0.9	93.3	51.5	1.9	50.0
LDA	75.3	0.9	78.3	82.1	1.1	86.7	62.5	2.0	62.5

Tabla 3

IBD vs no IBD									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMPoly	84.9	0.5	84.8	82.2	1.0	80.0	87.2	0.8	88.6
SVMRadial	84.0	0.5	84.4	80.1	1.0	80.0	87.3	0.8	88.2
SVMLineal	82.8	0.7	81.8	81.4	1.2	80.0	84.1	1.0	83.3
RF	81.9	0.7	81.8	79.5	1.1	80.0	84.0	1.0	83.3
LDA	81.5	0.5	81.8	80.7	1.0	80.0	82.2	0.8	83.3
PLS	80.4	0.5	81.3	78.8	1.1	80.0	81.7	0.9	82.4
KNN	76.5	0.7	75.8	75.3	1.1	73.3	77.6	1.0	77.8
IBS vs no IBS									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
PLS	92.1	0.5	90.9	80.3	1.5	81.3	96.0	0.4	96.0
SVMRadial	89.7	0.4	90.6	61.4	1.7	62.5	98.9	0.2	100.0
SVMLineal	89.6	0.5	90.6	78.6	1.6	75.0	93.2	0.5	92.0
SVMPoly	89.5	0.4	90.6	66.1	1.6	62.5	97.1	0.4	100.0
LDA	88.6	0.5	87.9	76.8	1.6	75.0	92.4	0.6	92.0
RF	83.4	0.5	84.4	36.9	1.9	37.5	98.5	0.2	100.0
KNN	82.9	0.5	81.8	39.2	1.9	37.5	97.0	0.4	96.0
Control vs no Control									
	Precisión (%)			Sensibilidad (%)			Especificidad (%)		
Clasificador	Media	SE	Mediana	Media	SE	Mediana	Media	SE	Mediana
SVMPoly	86.8	0.4	87.5	64.5	1.6	60.0	96.2	0.5	95.7
SVMRadial	85.0	0.4	84.8	61.2	1.7	60.0	95.1	0.5	95.7
LDA	85.0	0.6	86.2	74.6	1.6	77.8	89.5	0.7	91.3
SVMLineal	84.5	0.6	84.8	73.5	1.6	77.8	89.2	0.7	91.3

RF	83.5	0.5	84.4	51.0	1.9	50.0	97.2	0.3	95.7
PLS	82.8	0.7	84.4	67.3	1.5	70.0	89.4	0.8	91.3
KNN	80.2	0.6	81.3	54.0	1.9	55.6	91.2	0.6	91.3

REIVINDICACIONES

1. Un método para crear un clasificador indicativo de una presencia de una condición médica en un sujeto, que comprende:

recibir (105) datos de cromatograma indicativos de un perfil de compuestos orgánicos volátiles en una muestra de cada uno de una primera pluralidad de sujetos que tienen la condición médica y una segunda pluralidad de sujetos sin la condición médica;

seleccionar (140) uno de los datos de cromatograma como datos de cromatograma de referencia;

alinear (150) los datos de cromatograma restantes en relación con los datos de cromatograma de referencia;

caracterizado por:

extraer (160) una o más características de los datos de cromatograma alineados usando una transformada de ondícula del tipo conocido como sombrero mexicano de una o más escalas;

seleccionar (180) una o más de dichas características extraídas de los datos de cromatograma indicativos de la condición médica; y

construir (190) un clasificador de dicha selección de características para determinar un límite entre los datos de cromatograma indicativos de la condición médica y datos de cromatograma indicativos de una ausencia de la condición médica.

2. El método de la reivindicación 1, en donde la selección (140) de los datos de cromatograma de referencia comprende:

determinar un coeficiente de correlación entre cada uno de una primera pluralidad de datos de cromatograma; y

seleccionar (140) datos de cromatograma que tienen un coeficiente de correlación positivo más alto como los datos de cromatograma de referencia.

3. El método de la reivindicación 2, en donde:

el coeficiente de correlación se determina entre cada uno de la primera pluralidad de datos de cromatograma en cada uno de una pluralidad de puntos de muestra dentro de una ventana de desplazamiento predeterminada; y

la selección (140) de los datos de cromatograma comprende seleccionar un intervalo de desplazamiento de los datos de cromatograma que tienen un coeficiente de correlación positivo más alto.

4. El método de la reivindicación 3, en donde los datos de cromatograma restantes se alinean en relación con el punto de muestra de los datos de cromatograma de referencia que tienen el coeficiente de correlación positivo más alto; opcionalmente el coeficiente de correlación es un coeficiente de correlación del producto-momento de Pearson.

5. El método de cualquier reivindicación precedente, en donde la extracción (160) de la una o más características de los datos de cromatograma comprende determinar un coeficiente para los datos de cromatograma en cada una de la pluralidad de escalas de la ondícula del tipo conocido como sombrero mexicano; opcionalmente la pluralidad de escalas está entre límites superior e inferior; opcionalmente los límites superior e inferior son 100 y 1, respectivamente.

6. El método de la reivindicación 5, que comprende seleccionar (140) una de la pluralidad de escalas como una mejor coincidencia para los datos de cromatograma; opcionalmente la escala se selecciona como una mejor coincidencia con base en una precisión de un proceso de validación.

7. El método de cualquier reivindicación precedente, en donde la una o más características de los datos de cromatograma indicativos de la condición médica se seleccionan usando un algoritmo de selección basado en bosque aleatorio; opcionalmente en dicho algoritmo se seleccionan una o más características de los datos de cromatograma que, cuando se omiten, llevan a una pérdida de precisión.

8. El método de cualquier reivindicación precedente, que comprende transformar un rango de los datos de cromatograma.

9. El método de la reivindicación 8, en donde:

la transformación de rango se aplica para establecer los valores de los datos de cromatograma para estar en un rango predeterminado; opcionalmente el rango está entre 0 y 1; y/o

el rango de los datos de cromatograma se transforma de acuerdo con la ecuación:

$$x_t = \frac{(x - \min(x))}{(\max(x) - \min(x))}$$

donde un valor transformado x_t en cada punto de tiempo de los datos de cromatograma donde x es un valor de datos de los datos de cromatograma y min(x) y max(x) son el valor mínimo y máximo de los datos de cromatograma.

10. El método de cualquier reivindicación precedente, en donde el clasificador se construye de acuerdo con uno de: análisis discriminante lineal (LDA); mínimos cuadrados parciales (PLS); bosque aleatorio; k vecindario más cercano (KNN); máquina de vectores de soporte (SVM) con núcleo de función de base radial (SVMRadial); SVM con núcleo de función de base lineal (SVMLineal); y SVM con núcleo de función de base polinomial (SVMPoly).

11. Un método para determinar una presencia de una condición médica en un sujeto, que comprende:

recibir (105) datos de cromatograma indicativos de un perfil de compuestos orgánicos volátiles en una muestra del sujeto;

alinear (150) los datos de cromatograma con los datos de cromatograma de referencia;

extraer (105) una o más características predeterminadas de los datos de cromatograma usando una transformada de ondícula del tipo conocido como sombrero mexicano de una o más escalas predeterminadas en donde la una o más características predeterminadas son características seleccionadas en un método de acuerdo con cualquier reivindicación precedente; y

determinar si las características extraídas son indicativas de la presencia de una condición médica en el sujeto usando el clasificador construido con el método de acuerdo con cualquier reivindicación precedente.

12. El método de la reivindicación 11, en donde la determinación de si las características extraídas son indicativas de la presencia de la condición médica en el sujeto se basa en los valores de las características extraídas.

13. El método de la reivindicación 11 o 12, en donde la alineación (150) de los datos de cromatograma comprende:

determinar un coeficiente de correlación entre los datos de cromatograma y los datos de cromatograma de referencia en cada uno de una pluralidad de puntos de muestra dentro de una ventana de desplazamiento predeterminada; y

alinear (150) los datos de cromatograma con los datos de cromatograma de referencia en un tiempo de punto de muestra que tenga un mayor coeficiente de correlación.

14. Software para ordenador que, cuando es ejecutado por un ordenador, está dispuesto para realizar un método de acuerdo con cualquier reivindicación precedente; opcionalmente el software para ordenador se almacena en un medio legible por ordenador.

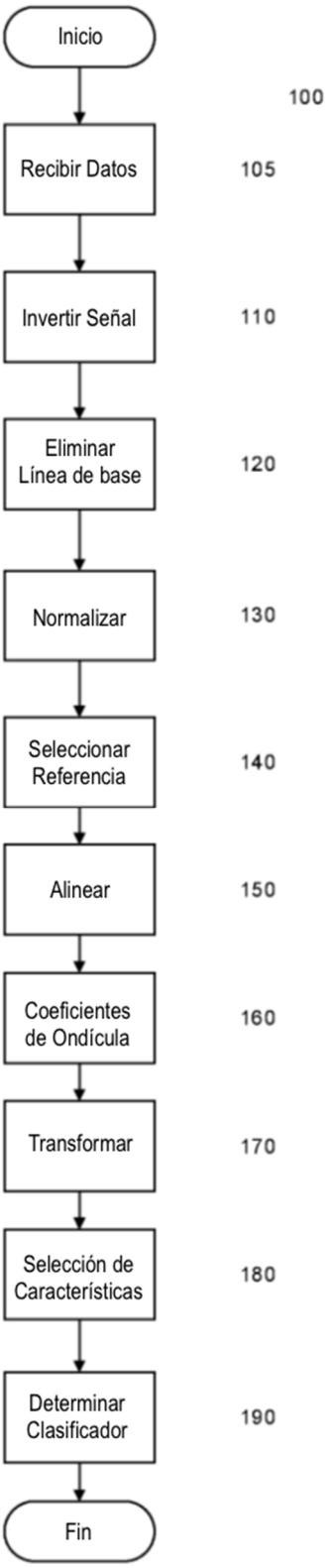
Inicio
100
Recibir Datos
105
Invertir Señal
110
Eliminar
Línea de base
120
Normalizar
130
Seleccionar
Referencia
140
Alinear
150
Coeficientes
de Ondícula
160
Transformar
170
Selección de
Características
180
Determinar
Clasificador
190
Fin

Figura 1

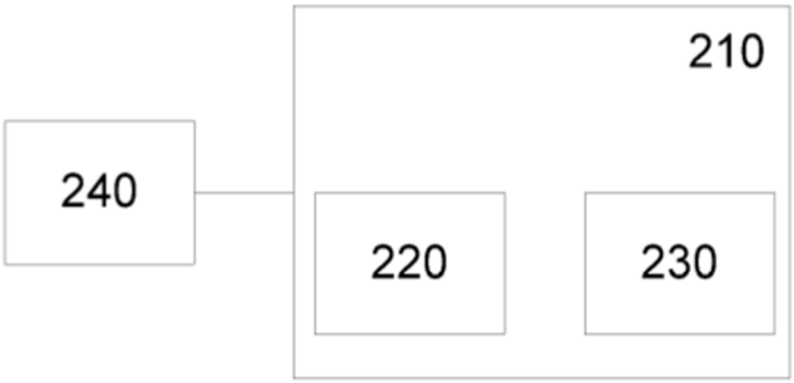

Figura 2

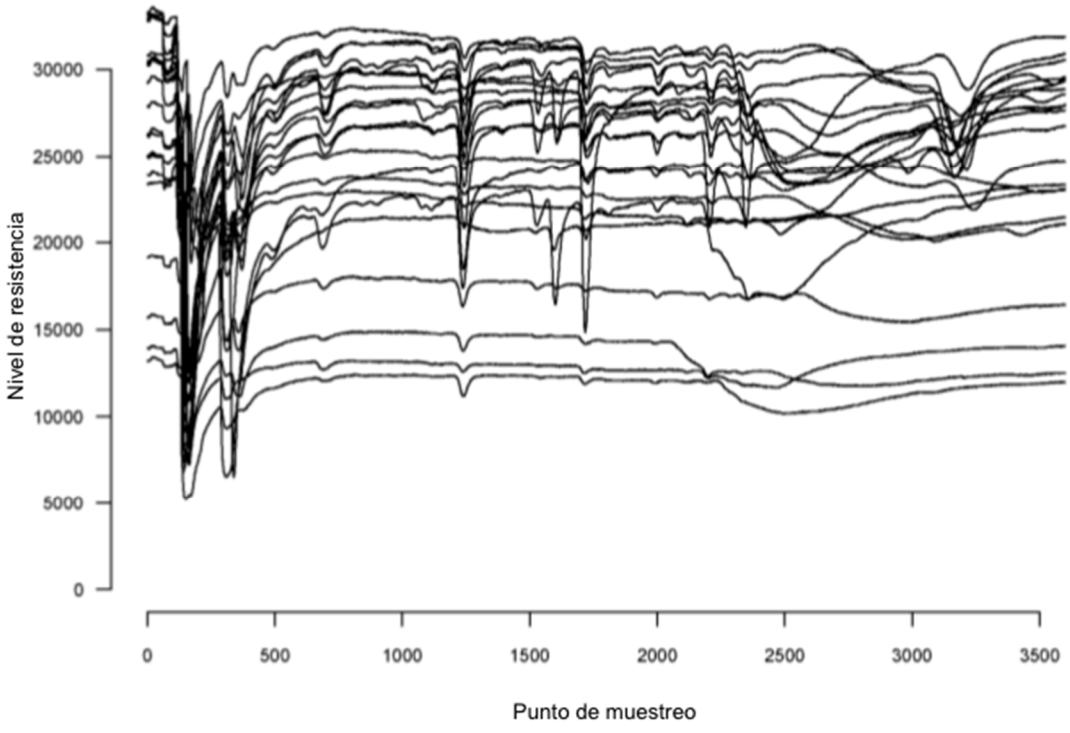

Figura 3

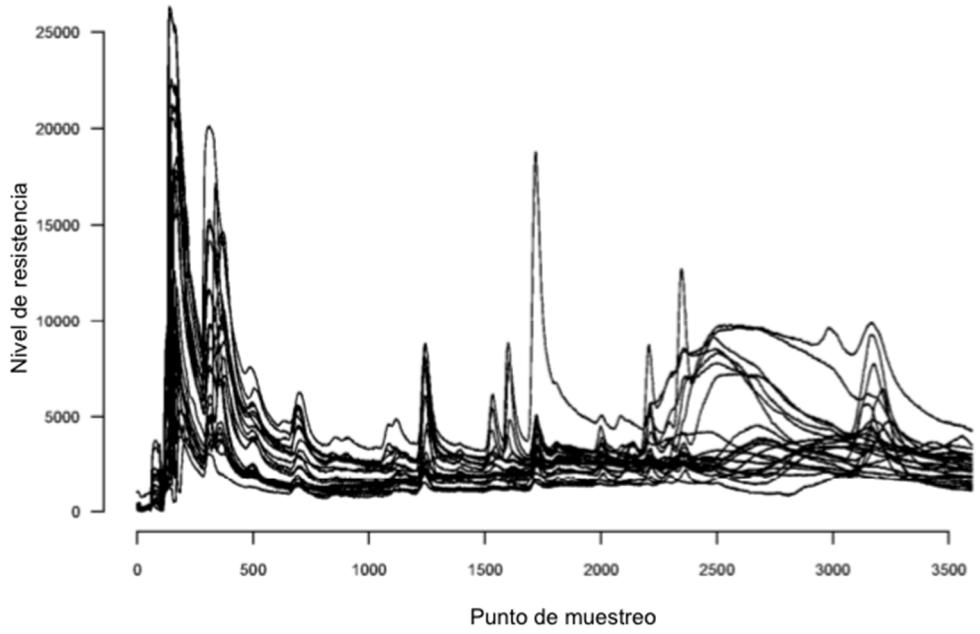

Figura 4

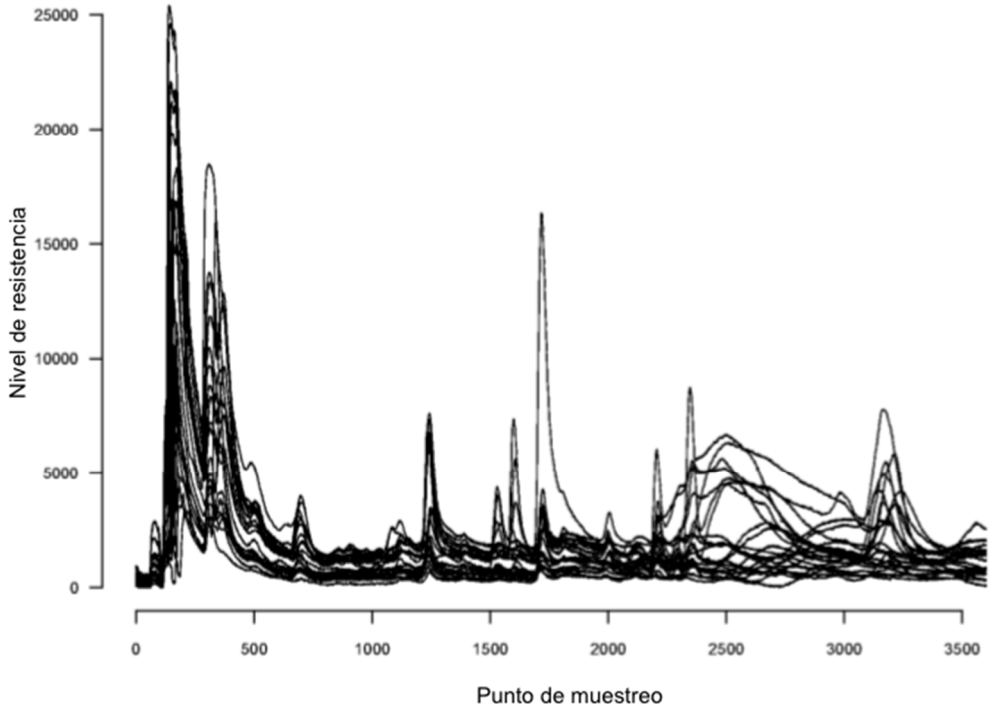

Figura 5

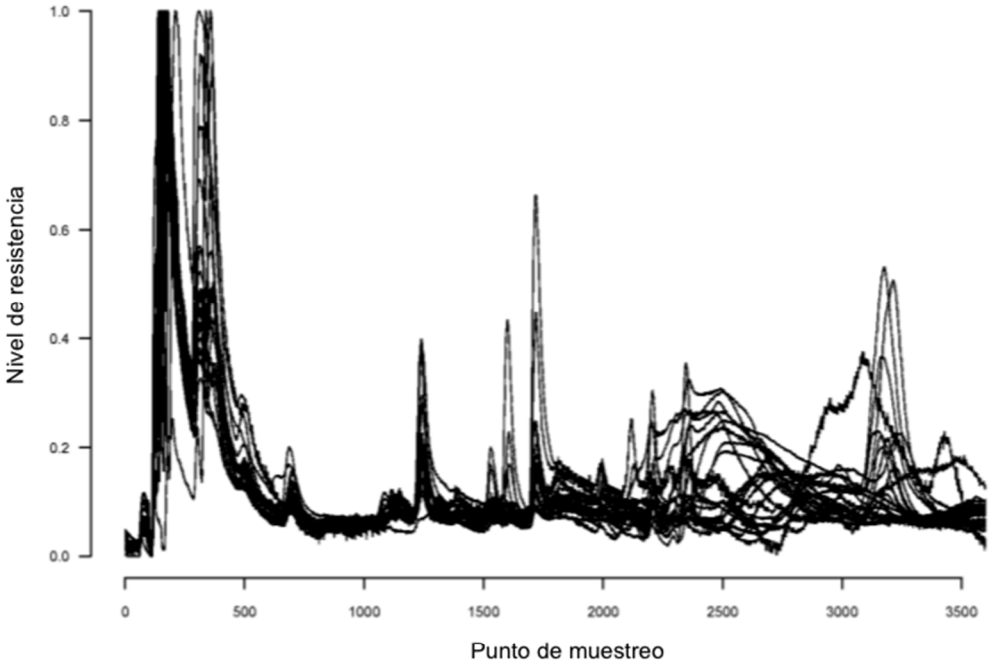

Figura 6

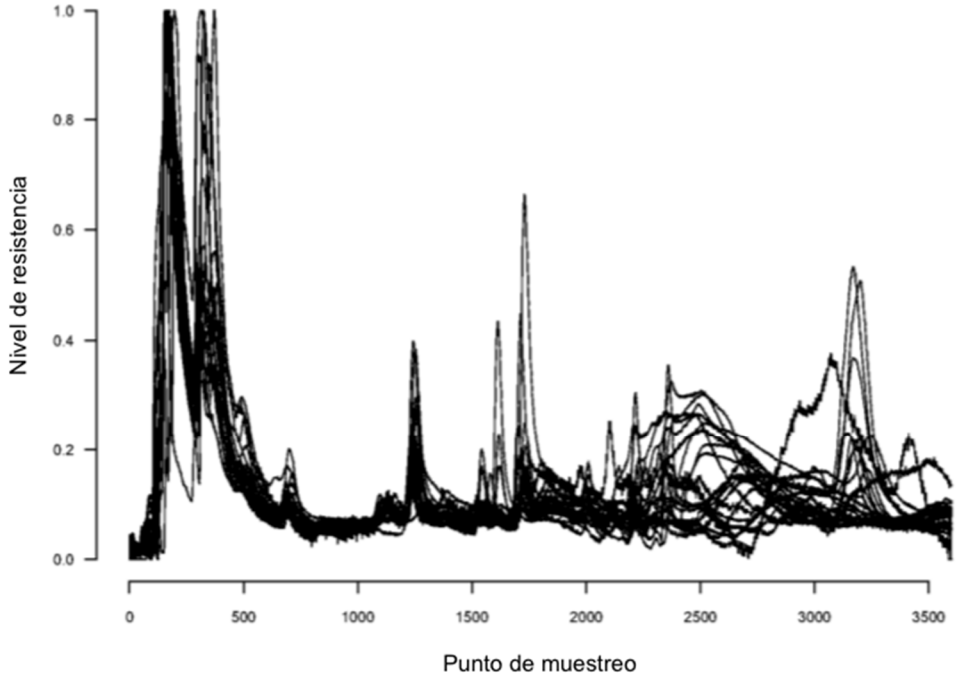

Figura 7

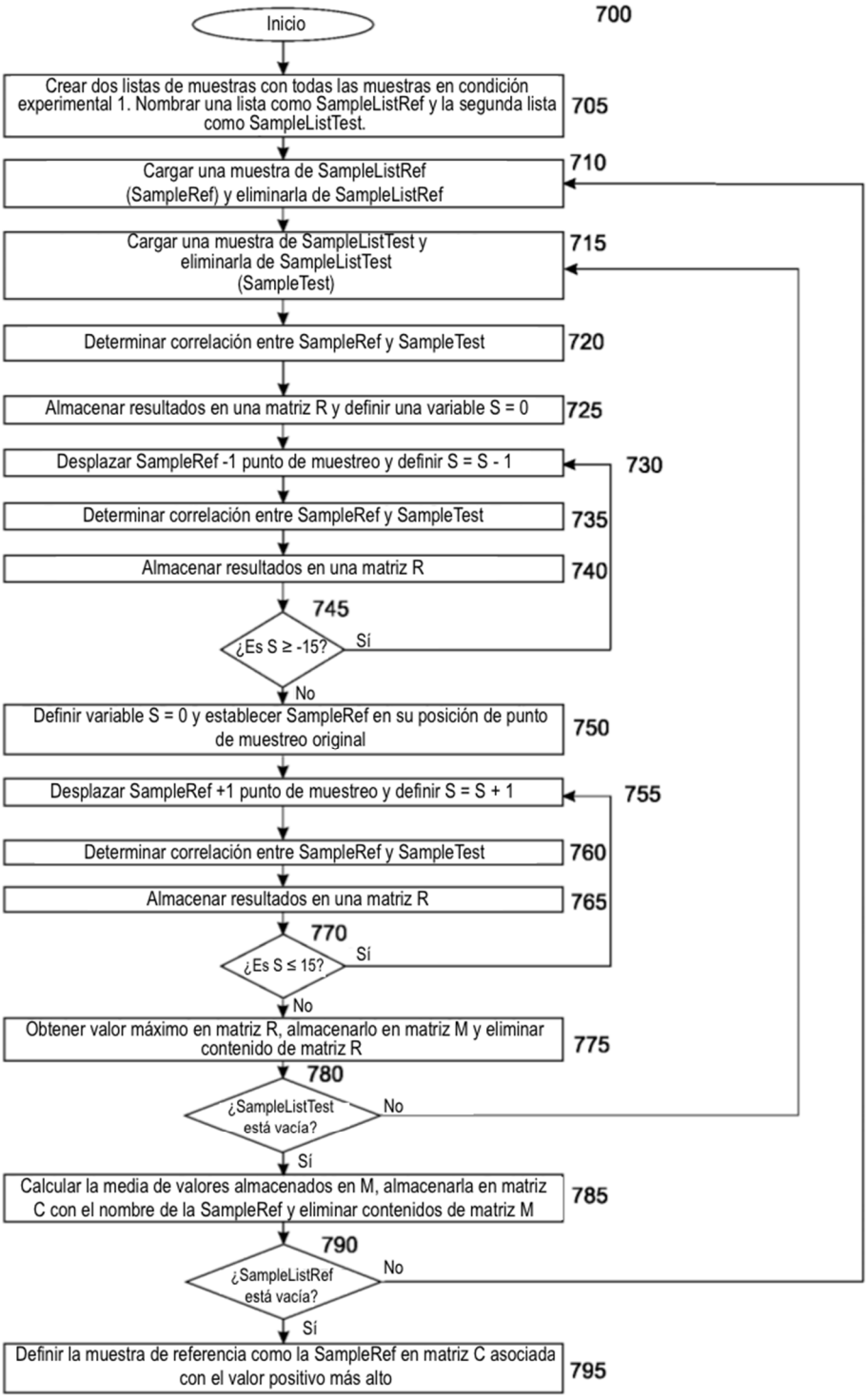
700
Inicio
Crear dos listas de muestras con todas las muestras en condición experimental 1. Nombrar una lista como SampleListRef y la segunda lista como SampleListTest.
705
Cargar una muestra de SampleListRef (SampleRef) y eliminarla de SampleListRef
710
Cargar una muestra de SampleListTest y eliminarla de SampleListTest (SampleTest)
715
Determinar correlación entre SampleRef y SampleTest
720
Almacenar resultados en una matriz R y definir una variable S = 0
725
Desplazar SampleRef -1 punto de muestreo y definir S = S - 1
730
Determinar correlación entre SampleRef y SampleTest
735
Almacenar resultados en una matriz R
740
745
¿Es S ≥ -15?
Sí
No
Definir variable S = 0 y establecer SampleRef en su posición de punto de muestreo original
750
Desplazar SampleRef +1 punto de muestreo y definir S = S + 1
755
Determinar correlación entre SampleRef y SampleTest
760
Almacenar resultados en una matriz R
765
770
¿Es S ≤ 15?
Sí
No
Obtener valor máximo en matriz R, almacenarlo en matriz M y eliminar contenido de matriz R
775
780
¿SampleListTest está vacía?
No
Sí
Calcular la media de valores almacenados en M, almacenarla en matriz C con el nombre de la SampleRef y eliminar contenidos de matriz M
785
790
¿SampleListRef está vacía?
No
Sí
Definir la muestra de referencia como la SampleRef en matriz C asociada con el valor positivo más alto
795

Figura 8

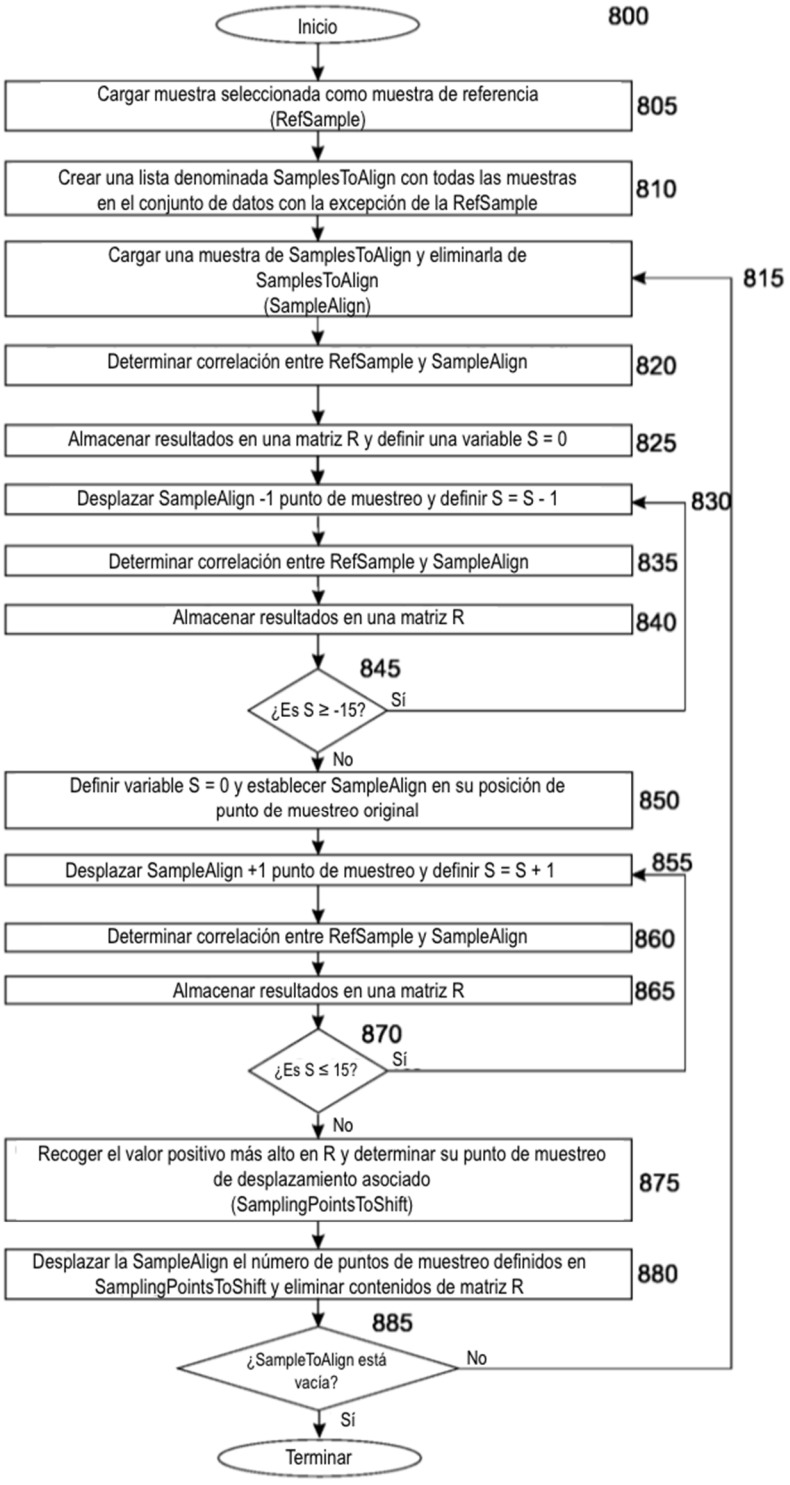
800
Inicio
Cargar muestra seleccionada como muestra de referencia (RefSample)
805
Crear una lista denominada SamplesToAlign con todas las muestras en el conjunto de datos con la excepción de la RefSample
810
Cargar una muestra de SamplesToAlign y eliminarla de SamplesToAlign (SampleAlign)
815
Determinar correlación entre RefSample y SampleAlign
820
Almacenar resultados en una matriz R y definir una variable S = 0
825
Desplazar SampleAlign -1 punto de muestreo y definir S = S - 1
830
Determinar correlación entre RefSample y SampleAlign
835
Almacenar resultados en una matriz R
840
845
¿Es S ≥ -15?
Sí
No
Definir variable S = 0 y establecer SampleAlign en su posición de punto de muestreo original
850
Desplazar SampleAlign +1 punto de muestreo y definir S = S + 1
855
Determinar correlación entre RefSample y SampleAlign
860
Almacenar resultados en una matriz R
865
870
¿Es S ≤ 15?
Sí
No
Recoger el valor positivo más alto en R y determinar su punto de muestreo de desplazamiento asociado (SamplingPointsToShift)
875
Desplazar la SampleAlign el número de puntos de muestreo definidos en SamplingPointsToShift y eliminar contenidos de matriz R
880
885
¿SampleToAlign está vacía?
No
Sí
Terminar

Figura 9

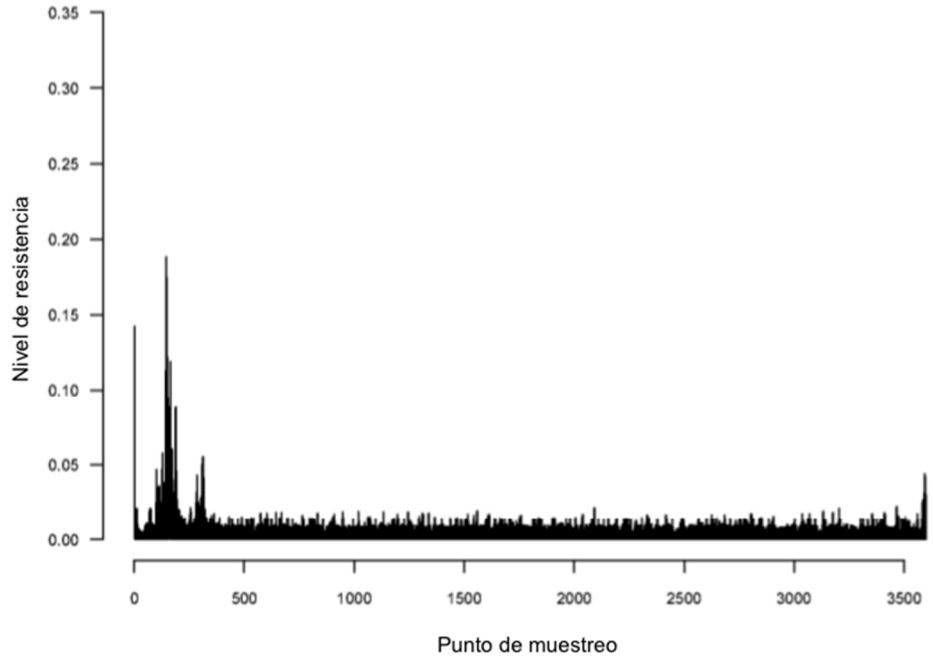

Figura 10

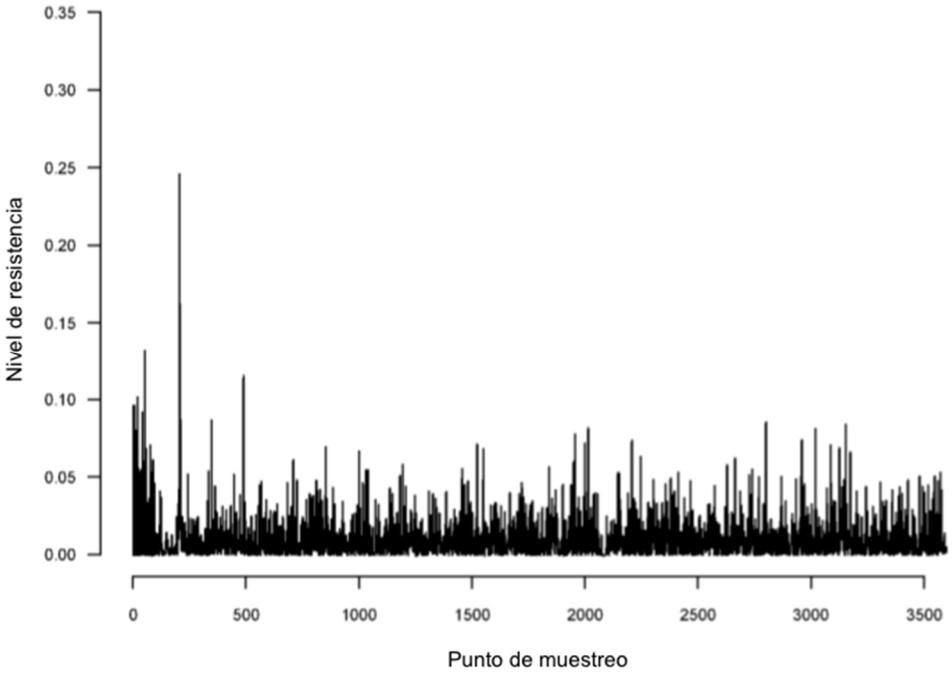

Figura 11

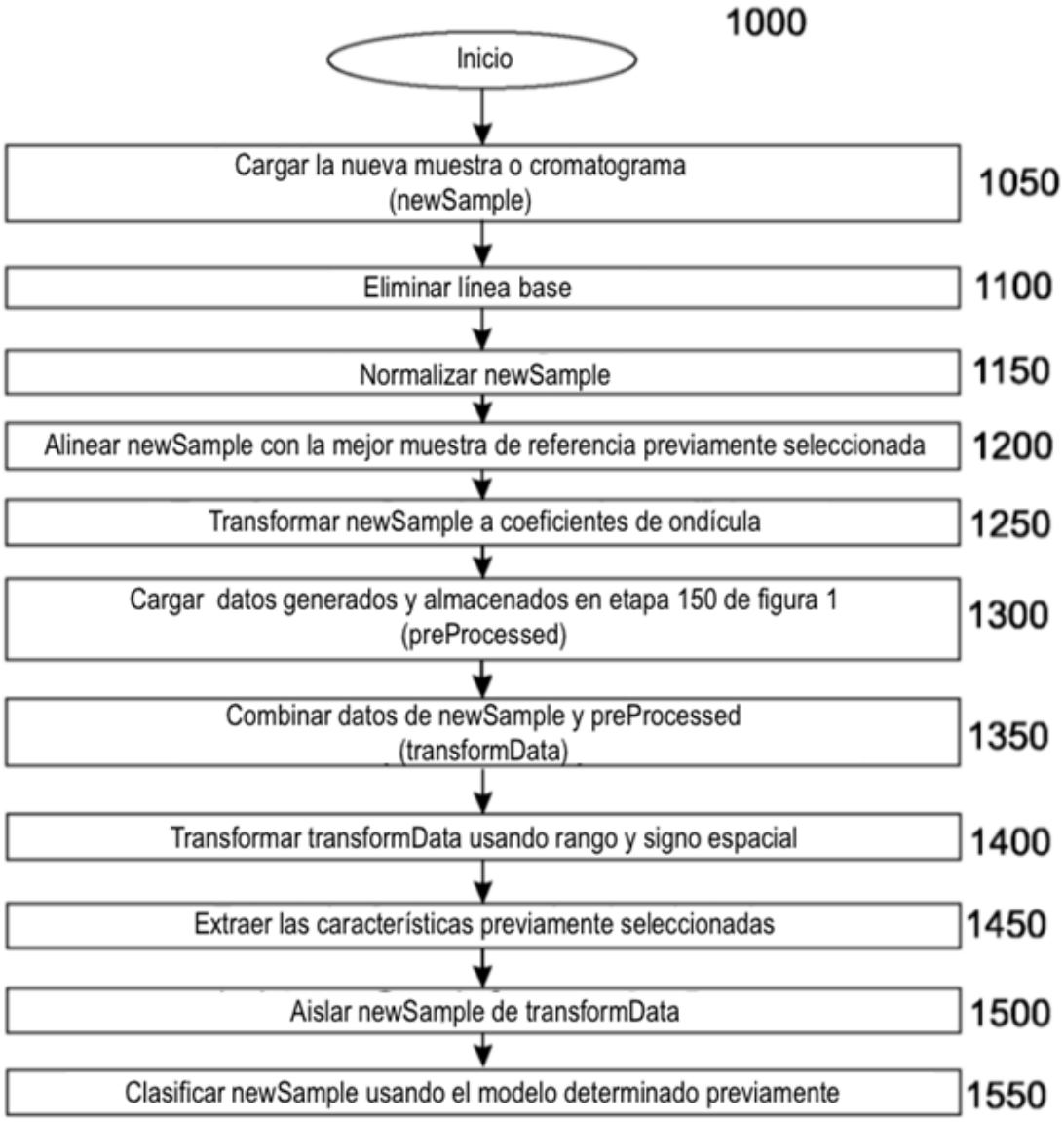
1000
Inicio
Cargar la nueva muestra o cromatograma (newSample)
1050
Eliminar línea base
1100
Normalizar newSample
1150
Alinear newSample con la mejor muestra de referencia previamente seleccionada
1200
Transformar newSample a coeficientes de ondícula
1250
Cargar datos generados y almacenados en etapa 150 de figura 1 (preProcessed)
1300
Combinar datos de newSample y preProcessed (transformData)
1350
Transformar transformData usando rango y signo espacial
1400
Extraer las características previamente seleccionadas
1450
Aislar newSample de transformData
1500
Clasificar newSample usando el modelo determinado previamente
1550

Figura 12

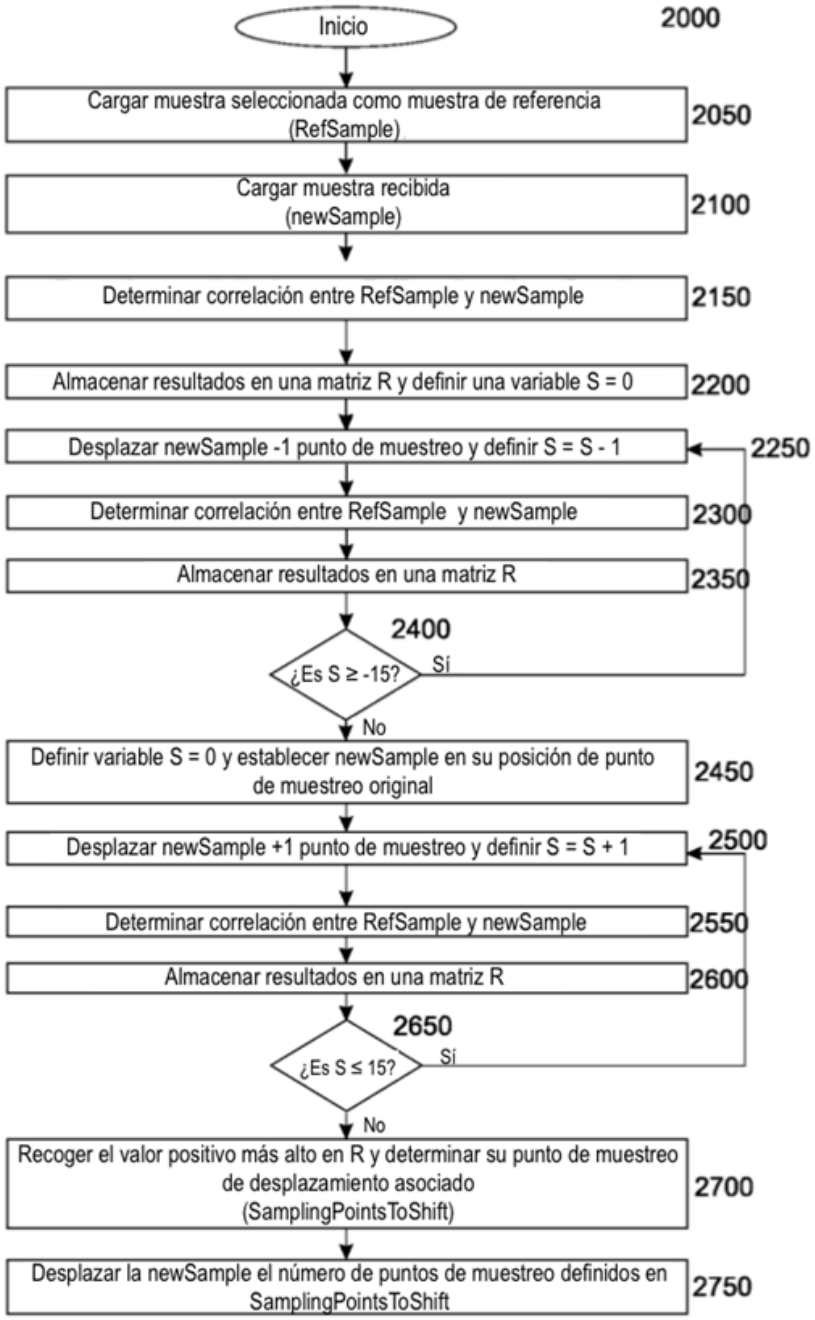
Inicio
2000
Cargar muestra seleccionada como muestra de referencia (RefSample)
2050
Cargar muestra recibida (newSample)
2100
Determinar correlación entre RefSample y newSample
2150
Almacenar resultados en una matriz R y definir una variable S = 0
2200
Desplazar newSample -1 punto de muestreo y definir S = S - 1
2250
Determinar correlación entre RefSample y newSample
2300
Almacenar resultados en una matriz R
2350
2400
¿Es S ≥ -15?
Sí
No
Definir variable S = 0 y establecer newSample en su posición de punto de muestreo original
2450
Desplazar newSample +1 punto de muestreo y definir S = S + 1
2500
Determinar correlación entre RefSample y newSample
2550
Almacenar resultados en una matriz R
2600
2650
¿Es S ≤ 15?
Sí
No
Recoger el valor positivo más alto en R y determinar su punto de muestreo de desplazamiento asociado (SamplingPointsToShift)
2700
Desplazar la newSample el número de puntos de muestreo definidos en SamplingPointsToShift
2750

Figura 13