

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 992 796**

51 Int. Cl.:

G16H 50/20 (2008.01)

G16H 50/50 (2008.01)

G16H 50/70 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.10.2018** **PCT/EP2018/079905**

87 Fecha y número de publicación internacional: **09.05.2019** **WO19086555**

96 Fecha de presentación y número de la solicitud europea: **31.10.2018** **E 18799476 (9)**

97 Fecha y número de publicación de la concesión europea: **10.07.2024** **EP 3704712**

54 Título: **Sistema médico para diagnosticar la patología y/o el resultado de una enfermedad cognitiva**

30 Prioridad:

31.10.2017 US 201762579630 P

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

18.12.2024

73 Titular/es:

GE HEALTHCARE LIMITED (100.0%)
Pollards Wood, Nightingales Lane
Chalfont St. Giles, Buckinghamshire HP8 4SP,
GB

72 Inventor/es:

AHMAD, RABIA;
FUENTES, EMMANUEL;
NGUYEN, QUANG, TRUNG;
BUCKLEY, CHRISTOPHER y
WOLBER, JAN

74 Agente/Representante:

DEL VALLE VALIENTE, Sonia

ES 2 992 796 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

Sistema médico para diagnosticar la patología y/o el resultado de una enfermedad cognitiva

5 **Campo**

El presente contenido descrito se refiere a sistemas para la determinación de la progresión de una enfermedad en un sujeto. Más específicamente, el contenido descrito se refiere a la aplicación de técnicas de aprendizaje automático para una herramienta de salud digital para ayudar a la toma de decisiones clínicas y acelerar el desarrollo clínico de nuevas opciones terapéuticas.

Antecedentes

Las tecnologías de salud digital y su implementación en el entorno clínico son temas de creciente interés. Estas tecnologías son cada vez más potentes y sofisticadas. En particular, el aprendizaje automático es muy prometedor como ayuda para los médicos y otros profesionales sanitarios en sus esfuerzos para proporcionar un mejor tratamiento y atención a los pacientes.

Es posible acumular una gran cantidad de datos sobre pacientes individuales a lo largo de su tratamiento, debido tanto al creciente número de pruebas de diagnóstico que están disponibles como a la necesidad para diferenciar entre trastornos con síntomas inespecíficos o superpuestos. Algunos de estos datos pueden ser redundantes o pueden ser cruciales para la respuesta del paciente. Sería ventajoso para la patente, así como para el sistema sanitario en general, que las pruebas que se administraran fueran solo las necesarias para permitir la toma de decisiones apropiadas dirigidas al resultado deseado.

Las tecnologías digitales también pueden utilizarse en el entorno de la investigación, en particular para abordar problemas de investigación desafiantes, por ejemplo, llevar nuevos fármacos modificadores de enfermedades (DMD) al mercado de manera más eficiente. Los requisitos reglamentarios y de reembolso son cada vez más estrictos y, aunque los organismos reguladores, como la Administración de Alimentos y Medicamentos (FDA), han implementado programas para tratar de acelerar el desarrollo de fármacos, como el programa Fast Track, los desafíos a los que se enfrenta la industria farmacéutica durante el desarrollo clínico de los posibles candidatos siguen siendo importantes y representan una enorme tarea financiera.

Llevar las opciones terapéuticas al mercado es costoso y requiere ensayos clínicos amplios y prolongados. A la complejidad se suma la alta tasa de exclusión provocada por el reclutamiento de los sujetos equivocados. Los riesgos financieros asociados con la realización de pruebas clínicas a un fármaco candidato siguen siendo altos. Es necesario reducir esto para permitir que los fármacos candidato avancen de manera más eficiente en la cartera de productos, aumenten las probabilidades de éxito y mantener la cartera de productos permitiendo que otros candidatos entren a las pruebas clínicas.

Utilizando la demencia y, en particular, la enfermedad de Alzheimer (EA) como ejemplo, actualmente no hay DMD disponible. Se han sometido a prueba muchos fármacos potenciales, pero hasta la fecha ninguno ha demostrado una eficacia significativa. Esto se debe en parte a que los ensayos se han centrado en sujetos en un estadio moderado de la enfermedad en lugar de en estadios más tempranos de la enfermedad. Para corregir esto, los estudios ahora tienen como objetivo reclutar sujetos con deterioro cognitivo leve (DCL) e identificar a aquellos que tengan probabilidades de convertirse en EA durante el transcurso del ensayo, ya que lo más probable es que muestren un efecto farmacológico. Para reclutar sujetos en este estadio, se requieren herramientas de selección y estratificación más sensibles.

En otras palabras, llevar los DMD al mercado para la enfermedad de Alzheimer es extremadamente complejo debido a la heterogeneidad de la enfermedad y a la necesidad de ensayos clínicos amplios y prolongados. Uno de los principales factores de confusión de la eficacia en los DMD es la selección de los sujetos apropiados. Se requieren herramientas de selección y estratificación más sensibles para identificar a los sujetos adecuados para una cohorte y un criterio de valoración dados.

Es posible mejorar la selección de los pacientes para los ensayos de EA con el uso de obtención de imágenes mediante tomografía por emisión de positrones (PET) de amiloides. Por ejemplo, VizamyTM (inyección de flutemetamol F-18, GE Healthcare) es un agente de diagnóstico radiactivo indicado para la obtención de imágenes (PET) del cerebro para estimar la densidad de la placa neurítica beta-amiloide (A β) en pacientes adultos con deterioro cognitivo que están siendo evaluados para detectar la EA u otras causas de deterioro cognitivo. Véase la información de prescripción de VizamyTM (inyección de flutemetamol F-18) de GE Healthcare. Otros agentes conocidos incluyen NeuraceqTM (inyección de florbetaben F18 de Piramal Imaging) y AmyvidTM (florbetapir, Eli Lilly and Company). Una gammagrafía negativa para A β indica que las placas neuríticas son escasas o inexistentes y no concuerda con el diagnóstico neuropatológico de la EA en el momento de la adquisición de imágenes; un resultado negativo de la gammagrafía reduce la probabilidad de que el deterioro cognitivo del paciente se deba a la EA. Una gammagrafía positiva para A β indica placas neuríticas amiloides de moderadas a frecuentes; el examen neuropatológico ha demostrado que esta

cantidad de placa neurítica está presente en pacientes con EA, pero también puede estar presente en pacientes con otros tipos de afecciones neurológicas, así como en personas mayores con cognición normal.

La obtención de imágenes mediante PET de A β puede considerarse un complemento de otras evaluaciones diagnósticas. Una gammagrafía positiva para A β no establece necesariamente por sí sola un diagnóstico de EA u otro trastorno cognitivo. Si bien es posible que se permita a un sujeto que dé positivo en una PET de amiloides participar en un ensayo clínico de terapias moduladoras de amiloides, esta prueba es insuficiente porque no todos los sujetos positivos para amiloides progresan a EA o no lo hacen dentro de un periodo de tiempo que sería relevante para el ensayo clínico. Tomando el ejemplo de la EA, esto significaría la identificación del sujeto con más probabilidades de tener una gammagrafía positiva para amiloides.

Como otro ejemplo, el diagnóstico de la enfermedad de Parkinson (EP) sigue siendo un desafío en los pacientes que tienen síntomas anómalos o muestran una falta de respuesta a la medicación. La EP forma parte de un grupo de enfermedades con características comunes denominadas síndrome parkinsoniano (PS), incluyendo la parálisis supranuclear progresiva (PNP) y la atrofia multisistémica (MSA). La obtención de imágenes del transportador de dopamina (DAT) puede utilizarse en un esfuerzo por obtener un diagnóstico preciso mediante la determinación de la pérdida de actividad dopaminérgica. Se conocen varios análogos de feniltropano radiomarcados para la visualización de los transportadores de dopamina, incluyendo el producto aprobado DaTscan™ (GE Healthcare), un agente marcado con ¹²³I para su uso en la obtención de imágenes mediante tomografía por emisión monofotónica (SPECT).

Además, las técnicas de obtención de imágenes *in vivo* que utilizan radiofármacos como la PET y la SPECT son procedimientos de diagnóstico relativamente costosos y requieren muchos recursos. Sería ventajoso, antes de llevar a cabo tales técnicas de obtención de imágenes *in vivo*, poder identificar a los sujetos que es probable que tengan un resultado que sugiera un estado patológico. Así, por ejemplo, para la EP sería ventajoso identificar a los sujetos con más probabilidades de tener deficiencias dopaminérgicas antes de llevar a cabo la obtención de imágenes moleculares, y para la EA sería ventajoso identificar a los sujetos con más probabilidades de tener placas amiloides antes de llevar a cabo la obtención de imágenes moleculares.

Proporcionar modelos predictivos basados en la combinación óptima de biomarcadores clínicos y de obtención de imágenes tiene un gran potencial para mejorar el proceso de selección. Los biomarcadores predictivos comunes recopilados en los ensayos clínicos pueden incluir el volumen del hipocampo adquirido mediante obtención imágenes mediante RM para predecir la atrofia cerebral y la obtención de imágenes mediante PET para evaluar la patología amiloide. Si bien estos biomarcadores funcionan mejor que la utilización de los criterios de inclusión típicos por sí solos, no son capaces de captar por completo la complejidad de la enfermedad.

Estudios recientes han demostrado la capacidad de incorporar la automatización en la cuantificación de los volúmenes cerebrales y el SUVr beta-amiloide, junto con la capacidad de utilizar estas medidas como sustitutos de las lecturas visuales humanas de las imágenes. Véase, por ejemplo, Thurfjell L, y col., Automated Quantification of 18F-Flutemetamol PET Activity for Categorizing Scans as Negative or Positive for Brain Amyloid: Concordance with Visual Image Reads. *Journal of Nuclear Medicine*, 55, 1623-1628 (2014).

Se han intentado correlacionar los datos clínicos con la EA. Por ejemplo, la patente US-9.687.199 concedida a Ithapu y col, titulada "Medical Imaging System Providing Disease Prognosis" describe un sistema de inteligencia artificial para analizar datos clínicos. El sistema utiliza múltiples intervalos de módulos de aprendizaje automático, cada uno de los cuales trata una parte independiente de los datos clínicos para abordar la alta dimensionalidad y el bajo tamaño de la muestra de los datos. Sin embargo, el sistema descrito en la patente de 199 carece de medios para identificar a los sujetos que están en riesgo de desarrollar EA dentro de un periodo de tiempo definido.

Pereira, T. y col., "Predicting progression of mild cognitive impairment to dementia using neuropsychological data: a supervised learning approach using time windows", *BMC Medical Informatics and Decision Making*, vol. 17, n.º 1, 19 de julio de 2017, páginas 1-15, describe un enfoque de ventanas de tiempo para predecir la conversión a demencia, aprendiendo con pacientes estratificados mediante ventanas de tiempo. Los pacientes se agruparon basándose en la información clínica sobre si se convirtieron (DCL conversor) o permanecieron con DCL (DCL estable) dentro de un periodo de tiempo específico. Se sometieron a prueba ventanas de tiempo de 2, 3, 4 y 5 años. Se desarrolló un modelo de pronóstico para cada ventana de tiempo utilizando datos clínicos y neuropsicológicos, y este enfoque se comparó con el que se usa comúnmente en la bibliografía, donde se utiliza a todos los pacientes para que los modelos aprendan.

Por tanto, es necesaria la innovación para ayudar a abordar las deficiencias de la metodología actual.

Resumen

A continuación se presenta un resumen simplificado de una o más realizaciones con el fin de proporcionar una comprensión básica del contenido descrito en el presente documento. Este resumen no es una visión general extensa de todas las realizaciones contempladas, y no pretende identificar los elementos clave o críticos de todas las realizaciones ni delinear el alcance de ninguna o todas las realizaciones. Su único propósito es presentar algunos

conceptos de una o más realizaciones de forma simplificada como preludio a la descripción más detallada que se presenta más adelante.

5 En un aspecto, se proporciona un sistema médico tal como se define en la reivindicación 1. Las características de las realizaciones se definen en las reivindicaciones dependientes.

10 Según un ejemplo, se describe un sistema médico para predecir una enfermedad, patología de la enfermedad o estado patológico en un sujeto que tiene un estado cognitivo incierto, comprendiendo el sistema un sistema informático dispuesto para recibir primeros datos médicos del sujeto y configurado, al menos en parte, como una máquina de aprendizaje entrenada que se ha entrenado en segundos datos médicos, adaptándose el sistema informático para usar la máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad y una pantalla para visualizar una indicación de la predicción, en donde los primeros datos médicos no comprende los datos obtenidos de un procedimiento de obtención de imágenes moleculares sobre el sujeto y los segundos datos médicos no comprenden los datos obtenidos de uno o más procedimientos de obtención de imágenes moleculares. La patología de la enfermedad puede ser la positividad para beta amiloide (A β) en el cerebro del sujeto. Los primeros datos médicos pueden incluir los resultados de las pruebas cognitivas del sujeto, los resultados de las pruebas cognitivas del sujeto, la edad del sujeto, el nivel educativo del sujeto o alguna combinación de estos datos.

20 Según otro ejemplo, se describe un sistema médico para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto, comprendiendo el sistema un sistema informático dispuesto para recibir primeros datos médicos del sujeto y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que se ha entrenado en segundos datos médicos y una segunda máquina de aprendizaje que se ha entrenado en terceros datos médicos, adaptándose el sistema informático para usar la primera máquina de aprendizaje entrenada y la segunda máquina de aprendizaje para proporcionar una predicción de la patología de la enfermedad, una pantalla para visualizar una indicación de la predicción, en donde los primeros datos médicos pueden comprender datos obtenidos de un procedimiento de obtención de imágenes moleculares sobre el sujeto y la primera máquina de aprendizaje entrenada se ha entrenado al menos en parte con datos obtenidos de uno o más procedimientos de obtención de imágenes moleculares.

30 Según otro ejemplo, se describe un método para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto, comprendiendo el método las etapas de recopilar datos médicos de la cohorte en una memoria electrónica para un primer conjunto de sujetos con resultados conocidos de la enfermedad, utilizando un sistema informático dispuesto para recibir los datos médicos del sujeto y que comprende una máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la cohorte, adaptándose el sistema informático para usar la máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad basada, al menos en parte, en los datos médicos del sujeto para proporcionar una predicción de la patología de la enfermedad; y usar una pantalla para indicar la predicción.

40 Según otro ejemplo, se describe un método para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto, comprendiendo el método las etapas de recopilar los datos médicos de la primera cohorte en una memoria electrónica para un primer conjunto de sujetos que tienen resultados conocidos para la enfermedad, recopilar los datos médicos de la segunda cohorte en la memoria electrónica para un segundo conjunto de sujetos que tienen resultados conocidos de la enfermedad, utilizando un sistema informático dispuesto para recibir los datos médicos del sujeto y que comprende una primera máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la primera cohorte y una segunda máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la segunda cohorte, adaptándose el sistema informático para usar la primera y la segunda máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad basándose al menos en parte en los datos médicos del sujeto, y visualizando una indicación de la predicción. Los datos médicos de la primera cohorte pueden comprender tipos de datos diferentes de los tipos de datos de los datos médicos de la segunda cohorte. Los datos médicos de la primera cohorte pueden comprender tipos de datos al menos parcialmente iguales a los tipos de datos de los datos médicos de la segunda cohorte.

55 Los sistemas médicos anteriores pueden usarse para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto y/o para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto. La capacidad de predecir una patología de la enfermedad particular sin necesidad de obtener imágenes moleculares como herramienta de preselección permite el enriquecimiento de una cohorte de sujetos que luego se someten al procedimiento de obtención de imágenes moleculares. También hay beneficios prácticos en cuanto a la economía de la sanidad y de reducción de la exposición de los sujetos a la radiación. Llevar a cabo la obtención de imágenes moleculares en los sujetos seleccionados a partir de la preselección permite la estratificación en cuanto a la tasa de progresión de la enfermedad.

65 Según otro ejemplo, se describe un sistema médico para identificar a los sujetos que están en riesgo de desarrollar la enfermedad de Alzheimer (EA), comprendiendo el sistema un sistema informático dispuesto para recibir datos médicos de uno o más sujetos con resultados desconocidos para la EA y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para

proporcionar una identificación de qué sujetos que tienen resultados desconocidos para la EA están en riesgo de desarrollar EA dentro de un periodo de tiempo definido y una pantalla para visualizar la identificación. La primera indicación puede comprender la probabilidad de que un sujeto sea positivo para A β . La segunda indicación puede comprender una probabilidad de que el deterioro cognitivo leve de un sujeto progrese rápidamente. La primera máquina de aprendizaje entrenada y la segunda máquina de aprendizaje entrenada pueden implementarse en el mismo hardware físico. El sistema médico puede incluir una memoria electrónica para almacenar los datos médicos y estar dispuesta para proporcionar los datos médicos al sistema informático. Los datos médicos pueden incluir datos de imágenes *in vivo*, datos de memoria cognitiva y funcional y datos genéticos. Los datos de imágenes *in vivo* pueden incluir razones de valores estandarizados de captación (SUVR) o volumétricas de resonancia magnética (RM) ponderadas en T1. Los datos médicos pueden incluir datos demográficos y/o la historia clínica electrónica de cada sujeto. Los sujetos pueden ser pacientes con deterioro cognitivo leve y el periodo de tiempo definido puede ser de 3 años o menos o 2 años o menos. La primera indicación puede ser una probabilidad de que un sujeto sea positivo para A β , la segunda indicación puede ser una probabilidad de que el deterioro cognitivo leve de un sujeto progrese rápidamente, y el sistema informático puede configurarse para usar la primera probabilidad y la segunda probabilidad para identificar a los sujetos que están en riesgo de desarrollar EA dentro de un periodo de tiempo definido.

Según otro ejemplo, se describe un método para identificar sujetos que están en riesgo de desarrollar la enfermedad de Alzheimer (EA), comprendiendo el método las etapas de recopilar datos médicos en una memoria electrónica almacenada para un primer conjunto de sujetos que tienen resultados conocidos para la enfermedad, usar una primera máquina de aprendizaje entrenada y una segunda máquina de aprendizaje entrenada para identificar a los sujetos que están en riesgo de desarrollar la EA dentro de un periodo de tiempo definido, y seleccionar los sujetos que están en riesgo de desarrollar la EA dentro de un periodo de tiempo definido para un estudio sobre la EA. Los datos médicos pueden comprender datos de imágenes *in vivo*, datos de memoria cognitiva y funcional y datos genéticos. Los datos de imágenes *in vivo* pueden comprender razones de valores estandarizados de captación (SUVR) o volumétricas de resonancia magnética (RM) ponderadas en T1. Los datos médicos pueden comprender datos demográficos y/o la historia clínica electrónica de cada sujeto. Los sujetos pueden ser pacientes con deterioro cognitivo leve y el periodo de tiempo definido puede ser de 3 años o menos o 2 años o menos.

Según otro ejemplo, se describe un método para determinar el pronóstico de un paciente con la enfermedad de Alzheimer (EA), comprendiendo el método las etapas de administrar un agente de contraste de proteína amiloide a un paciente que lo necesita, obtener imágenes de los depósitos de proteína amiloide en el paciente y correlacionar la obtención de imágenes de dichos depósitos de amiloides en el paciente junto con otras variables con un conjunto de entrenamiento de pacientes con pronósticos conocidos de EA.

Según otro ejemplo, se describe un método para tratar a un paciente con la enfermedad de Alzheimer (EA), comprendiendo el método las etapas de comparar, en un ordenador, los datos de obtención de imágenes y otros datos obtenidos del paciente con un conjunto de entrenamiento que comprende datos de obtención de imágenes y otros datos con pacientes para un pronóstico conocido de la EA con el fin de clasificar al paciente en una cohorte de enfermedades, identificar un tratamiento farmacológico conocido por mejorar los resultados de los pacientes para la EA dentro de la cohorte de la enfermedad y tratar al paciente con una cantidad eficaz de la terapia farmacológica identificada.

Según otro ejemplo, se describe el uso de datos médicos de uno o más sujetos con resultados desconocidos para la enfermedad de Alzheimer (EA) para determinar cuáles de los sujetos están en riesgo de desarrollar EA, comprendiendo el uso suministrar los datos médicos a un sistema informático configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación, y combinar la primera y la segunda indicación para proporcionar una identificación de qué sujetos que tienen resultados desconocidos para EA están en riesgo de desarrollar EA dentro de un periodo de tiempo definido.

Según otro ejemplo, se describe un sistema médico para clasificar a los sujetos que tienen deterioro cognitivo leve (DCL) o enfermedad de Alzheimer (EA), comprendiendo el sistema un sistema informático dispuesto para recibir datos médicos de uno o más sujetos con una clasificación desconocida para DCL o EA y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para proporcionar una identificación de qué sujetos tienen DCL y qué sujetos tienen EA, una pantalla para visualizar la clasificación. El sistema informático puede estar dispuesto para proporcionar a los sujetos clasificados como DCL una clasificación adicional de si el DCL es DCL temprano o DCL avanzado.

Según otro ejemplo, se describe un sistema médico para clasificar a los sujetos como enfermedad de Alzheimer (EA) o alguna otra forma de demencia, comprendiendo el sistema un sistema informático dispuesto para recibir datos médicos de uno o más sujetos con una clasificación desconocida para DCL o EA y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para proporcionar una identificación de qué sujetos tienen EA y qué sujetos tienen otra forma de demencia, y una pantalla para visualizar la clasificación.

Según otro ejemplo, se describe un sistema médico para identificar indicaciones adicionales para un fármaco, comprendiendo el sistema un sistema informático dispuesto para recibir datos médicos de uno o más sujetos que toman el fármaco y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para determinar si al fármaco se le puede asignar una indicación además de una indicación existente y una pantalla para visualizar la indicación adicional.

Otras características y ventajas de la presente invención, así como la estructura y el funcionamiento de diversas realizaciones de la presente invención, se describen en detalle a continuación con referencia a los dibujos adjuntos. Se observa que la presente invención no se limita a las realizaciones específicas descritas en el presente documento. Tales realizaciones se presentan en el presente documento únicamente con fines ilustrativos. Las realizaciones adicionales resultarán evidentes para los expertos en la(s) técnica(s) relevante(s) basándose en las enseñanzas contenidas en el presente documento.

Breve descripción de los dibujos

Los dibujos adjuntos, que se incorporan en el presente documento y forman parte de la memoria descriptiva, ilustran el presente contenido y, junto con la descripción verbal, sirven además para explicar los principios del presente contenido y para permitir a un experto en la(s) técnica(s) relevante(s) crear y usar el presente contenido.

La figura 1 es un diagrama de bloques de un sistema informático tal como el que podría usarse para implementar las enseñanzas del presente documento.

La figura 2 es un diagrama de flujo que representa un procedimiento para seleccionar sujetos para un estudio de acuerdo con una realización.

La figura 3 es un diagrama de flujo que representa un procedimiento para seleccionar sujetos para un estudio de acuerdo con otra realización.

La figura 4 es un diagrama de bloques de una implementación basada en la web de las enseñanzas del presente documento.

Descripción detallada

Ahora se describen diversas realizaciones con referencia a los dibujos, en donde se utilizan números de referencia similares para referirse a elementos similares en todas partes. En la siguiente descripción, con fines explicativos, se exponen numerosos detalles específicos para promover una comprensión completa de una o más realizaciones. Sin embargo, puede resultar evidente en algunos o en todos los casos que cualquier realización descrita a continuación puede ponerse en práctica sin adoptar los detalles de diseño específicos que se describen a continuación. En otros casos, las estructuras y dispositivos bien conocidos se muestran en forma de diagrama de bloques para facilitar la descripción de una o más realizaciones. A continuación se presenta un resumen simplificado de una o más realizaciones con el fin de proporcionar una comprensión básica de las realizaciones. Este resumen no es una visión general extensa de todas las realizaciones contempladas, y no pretende identificar los elementos clave o críticos de todas las realizaciones ni delinear el alcance de ninguna o todas las realizaciones.

La realización/las realizaciones descrita(s) y las referencias en la memoria descriptiva a “una realización”, “una realización de ejemplo”, etc., indican que la realización/las realizaciones descrita(s) puede(n) incluir un rasgo distintivo, una estructura o característica particulares, pero cada realización puede no incluir necesariamente el rasgo distintivo, la estructura o característica particulares. Además, tales expresiones no se refieren necesariamente a la misma realización. Además, cuando se describe un rasgo distintivo, una estructura o característica particulares en relación con una realización, se entiende que está dentro del conocimiento de un experto en la técnica efectuar tal rasgo distintivo, estructura o característica en relación con otras realizaciones, estén o no descritas explícitamente.

Las realizaciones de la presente invención pueden implementarse en hardware, firmware, software o cualquier combinación de los mismos. Las realizaciones de la presente invención también pueden implementarse como instrucciones almacenadas en un medio legible por máquina, que pueden ser leídas y ejecutadas por uno o más procesadores. Un medio legible por máquina puede incluir cualquier mecanismo para almacenar o transmitir información en una forma legible por una máquina (por ejemplo, un dispositivo informático). Por ejemplo, un medio legible por máquina puede incluir memoria de solo lectura (ROM); memoria de acceso aleatorio (RAM); medios de almacenamiento en disco magnético; medios de almacenamiento óptico; dispositivos de memoria flash; medios eléctricos, ópticos, acústicos o de otro tipo de medios legibles por máquina no transitorios. Además, el firmware, el software, las rutinas y las instrucciones pueden describirse en el presente documento como la realización de determinadas acciones. Sin embargo, debe apreciarse que tales descripciones son simplemente por conveniencia y que tales acciones, de hecho, son el resultado de dispositivos informáticos, procesadores, controladores u otros dispositivos que ejecutan el firmware, el software, las rutinas, las instrucciones, etc.

Para describir y apuntar de manera más clara y concisa el objeto de la invención reivindicada, se proporcionan definiciones a continuación para términos específicos usados en toda la presente memoria descriptiva y reivindicaciones. Cualquier ejemplificación de términos específicos en el presente documento debe considerarse como un ejemplo no limitante.

El término “estado cognitivo incierto” se usa en el presente documento para aplicarse a un sujeto que presenta síntomas que sugieren una enfermedad o afección asociada con el deterioro cognitivo. Los síntomas pueden incluir confusión, mala coordinación motora, confusión de identidad, alteración de sentido de la realidad, pérdida de memoria subjetiva, falta de concentración o enfoque e incapacidad para comunicarse bien. Los ejemplos no limitativos de enfermedades y afecciones conocidas asociadas con el deterioro cognitivo incluyen la EA, la EP, DCL, TBI (traumatismo craneoencefálico) y la encefalopatía crónica por traumatismo (CTE).

El término “patología de la enfermedad” se usa en el presente documento para referirse a una característica patológica normalmente asociada con una enfermedad o afección asociada con el deterioro cognitivo. Los ejemplos no limitativos de patologías de la enfermedad contempladas por la presente invención incluyen la positividad para beta amiloide (es decir, una presencia anómala de amiloide), la deficiencia dopaminérgica y la presencia de ovillos neurofibrilares (NFT) (ovillos tau).

El término “sujeto” se usa en el presente documento para referirse a cualquier sujeto humano o animal. En un aspecto de las realizaciones, el objeto de la invención es un mamífero. En otro aspecto de las realizaciones, el sujeto es un cuerpo de mamífero intacto *in vivo*. En otro aspecto de las realizaciones, el sujeto es un ser humano. El sujeto puede tener problemas subjetivos de memoria (SMC) o padecer un deterioro cognitivo leve (DCL) y se está investigando la posible aparición de la enfermedad de Alzheimer (EA).

Una “máquina de aprendizaje entrenada” es un sistema informático que se ha entrenado usando aprendizaje automático en un conjunto de datos de entrenamiento para proporcionarle la capacidad de proporcionar predicciones, conclusiones y clasificaciones, por ejemplo, cuando se le proporcionan datos nuevos. El sistema informático aprende por sí mismo sin estar programado explícitamente con las relaciones entre los datos y las predicciones, conclusiones y clasificaciones.

Una “pantalla” puede significar cualquier dispositivo capaz de visualizar información en forma alfanumérica o pictórica y, normalmente, incluye una pantalla, un conjunto de circuitos, una carcasa y una fuente de alimentación. Los ejemplos no limitativos de pantallas incluyen monitores de ordenador, pantallas de tabletas y pantallas de teléfonos inteligentes. Una pantalla puede ser local, es decir, conectada directamente a un ordenador local que aloja el sistema, o remota, es decir, parte de un sistema de usuario con el que un ordenador que aloja el sistema se comunica a través de una red o, por ejemplo, Internet con los servicios del sistema proporcionados como una utilidad basada en la nube.

El término “indicación de la predicción”, tal como se usa en el presente documento, pretende referirse a la información que se visualiza en la pantalla en forma alfanumérica o pictórica. Los ejemplos no limitativos de indicaciones incluyen gráficos, tablas, gráficos de barras

El término “procedimiento de obtención de imágenes moleculares” utilizado en el presente documento se refiere a un procedimiento de obtención de imágenes *in vivo* que permite la visualización (es decir, la producción de una imagen molecular) de las funciones celulares o los procesos moleculares en un sujeto. Un “procedimiento de obtención de imágenes *in vivo*” es una técnica que produce de manera no invasiva una imagen de todo o parte del aspecto interno de un sujeto. Los ejemplos no limitativos de procedimientos de obtención de imágenes moleculares contemplados por la presente invención incluyen la obtención de imágenes de beta amiloide, la obtención de imágenes de transportadores de dopamina y la obtención de imágenes de tau.

El término “positividad para beta amiloide (A β)” indica placas neuríticas amiloides de moderadas a frecuentes, por ejemplo, tal como se observa en una imagen molecular de A β obtenida mediante un agente de obtención de imágenes moleculares de A β . La predicción de la probabilidad de positividad para amiloides puede lograrse mediante aspectos de la presente invención usando datos que no incluyen datos de obtención de imágenes moleculares. Los ejemplos no limitativos de tales datos incluyen la referencia de las actividades de la vida diaria (ADL), memorización de dígitos hacia atrás, la memoria lógica II 30 minutos después de la referencia de la historia, prueba del trazo, parte A (segundos), la educación, el sexo masculino, el volumen del núcleo caudado izquierdo, el volumen del complejo amigdalino derecho y el volumen del núcleo caudado derecho. En una realización tales datos comprenden la edad, el sexo, la minipuntuación mental (MMSE), la clasificación de la demencia clínica (CDR), la suma de casillas de la calificación de la demencia clínica (CDR-SB), el estado de las pruebas genéticas de ApoE, los volúmenes cerebrales regionales izquierdo y derecho del hipocampo, el parahipocampo, el complejo amigdalino, la corteza entorrinal, el lóbulo temporal medial, la circunvolución recta, los ventrículos y la circunvolución angular.

El término “pruebas cognitivas” se refiere a las pruebas realizadas en sujetos para ayudar a determinar el estado cognitivo. Los expertos en la técnica conocen bien las pruebas típicas. Los ejemplos no limitativos de pruebas incluyen diversas formas de pruebas de coeficiente intelectual (IQ), memoria, atención, dibujar con atención y concentración.

Los ejemplos no limitativos de pruebas que normalmente se llevan a cabo en sujetos animales no humanos incluyen la prueba del espejo y la prueba del laberinto en T.

El término “datos médicos de cohorte”, tal como se usa en el presente documento, se entiende como una recopilación de datos de una cohorte definida de sujetos, por ejemplo, aquellos que se considera que tienen potencialmente una enfermedad o afección en particular y/o los sujetos que se está considerando su inclusión en un estudio clínico.

El término “deterioro cognitivo leve” (DCL) se usa para referirse a una afección que implica problemas con la función cognitiva, incluyendo la memoria, el lenguaje, la actividad mental y el sentido de la realidad a menudo mayores que los cambios normales relacionados con la edad.

El término “que progresa rápidamente” se refiere a una transición relativa desde un sujeto que muestra los síntomas iniciales de una enfermedad o afección hasta que se le diagnostica la enfermedad o afección. Un ejemplo no limitativo es la progresión de un sujeto del DCL a la demencia en un periodo de tiempo relativamente corto, por ejemplo, en un plazo de aproximadamente tres años. Algunos datos pueden ser útiles en la predicción de una progresión rápida. Los ejemplos no limitativos de dichos datos incluyen la referencia de las ADL, la referencia del miniexamen del estado mental (MMSE), la prueba de fluidez de categoría: categoría animal, la razón de valores estandarizados de captación de amiloide compuesto (SUVR) (pons) y el volumen del hipocampo.

Las compañías farmacéuticas pueden utilizar la invención en diversos puntos del flujo de trabajo de los ensayos clínicos, para ayudar a reducir el tiempo y el coste de lanzar nuevos fármacos al mercado. Más específicamente, mediante el enriquecimiento de los criterios de inclusión (herramientas de selección y estratificación) o la identificación de la cohorte adecuada que tiene más probabilidades de responder al tratamiento (herramienta de acceso al mercado).

Por ejemplo, después de una evaluación diagnóstica inicial, puede considerarse a un paciente para un ensayo evaluando su idoneidad en función de un criterio de inclusión. Esto se produce normalmente mediante un procedimiento de selección, que puede implicar el uso de varias pruebas de diagnóstico económicas y, a veces, también más costosas, como la obtención de imágenes mediante tomografía por emisión de positrones (PET). Un ensayo de fase 3 recluta a varios cientos de sujetos. En los ensayos, en los que hay una alta ineficacia de selección, se seleccionan muchos más sujetos de los que entran al ensayo. Por tanto, los métodos de selección tradicionales son una fuente de carga financiera para la industria farmacéutica en cuanto al mayor coste asociado a algunas pruebas; y el gran número de sujetos que pasan por esta etapa del procedimiento. La presente invención usa datos clínicos recopilados mediante pruebas más baratas para predecir el resultado de las pruebas más costosas. De esta manera, se realiza el triaje de los pacientes a través del procedimiento de selección. Cuando sea necesario, la funcionalidad adicional de la herramienta estratificará adicionalmente a los sujetos basándose en la tasa de progresión de la enfermedad o de los que tienen más probabilidades de responder a la terapia. De esta manera, hay una mejora de los criterios de inclusión mediante la predicción del resultado.

Una vez que el ensayo tiene éxito, los equipos de acceso al mercado requieren datos como prueba de la eficacia del fármaco para garantizar que los pagadores reembolsen el fármaco. En este caso, la presente invención puede ayudar a identificar qué pacientes se beneficiarán más del fármaco, analizar la evidencia real existente de múltiples fuentes y desarrollar los datos de resultados que los pagadores y los gerentes de ventas requieren para su adopción.

La aplicación de la presente invención a los datos de la fase de reclutamiento puede reducir el número de pacientes necesarios para demostrar la eficacia de un fármaco. Además, estas técnicas pueden predecir los sujetos con más probabilidades de mostrar un efecto farmacológico y, por tanto, refinar el número y la idoneidad de los sujetos necesarios para lograr un tamaño de efecto estadísticamente relevante. Estas eficiencias afectan al coste total del ensayo al acortar la fase de reclutamiento y reducir el coste general asociado con las pruebas de selección. Este es solo un ejemplo de un contexto en el que las enseñanzas de la presente descripción pueden aplicarse de manera beneficiosa.

En la figura 1 se muestra un sistema informático generalizado 100. El sistema 100 de la figura 1 incluye un procesador 110 conectado a un bus 120. También se conectan al bus una memoria (tal como firmware, RAM o ROM) 130, un disco duro 140, una interfaz de red 150 y una interfaz de entrada/salida (I/O) 160. La interfaz de red 150 interactúa con una red 170. La interfaz I/O 160 puede interactuar con uno o más dispositivos de entrada/salida, tales como un dispositivo de entrada 180 y una pantalla 190. El sistema 100 puede tener componentes adicionales tales como puertos, impresoras, lectores y grabadoras de CD/DVD-ROM, etc.

Un sistema informático generalizado 100, tal como el mostrado en la figura 1, puede configurarse para el aprendizaje automático, es decir, para poder realizar determinadas tareas sin una programación específica. Un aspecto de esto es proporcionar al sistema informático 100 un conjunto de datos de entrenamiento (“rasgos distintivos”) sobre los que pueda aprender a hacer predicciones. Luego, el sistema informático puede hacer predicciones basadas en nuevos datos.

Un aspecto del contenido descrito implica el uso de un enfoque analítico basado en el aprendizaje automático que facilita la identificación de los sujetos apropiados basándose en sus datos específicos de obtención de imágenes,

genéticos, psicométricos y demográficos para determinar la estratificación óptima de los sujetos de ensayos clínicos con deterioro cognitivo leve (DCL). El enfoque abarcaba la aplicación de modelos de aprendizaje automático en serie o individualmente para excluir a los pacientes que carecían de biomarcadores, una tasa de progresión apropiados o ambos.

Específicamente, en la figura 2 se muestra un procedimiento para seleccionar sujetos para un estudio. En una etapa S20, los sujetos con deterioro cognitivo leve (DCL) se identifican mediante una selección de ensayo convencional mediante pruebas psicométricas, la recopilación de datos demográficos de los sujetos y la adquisición de una IRM. En la etapa S30, se usa un modelo de aprendizaje automático para predecir la probabilidad de que un sujeto identificado sea positivo para beta-amiloide (A β). En la etapa S40, se adquiere una imagen de PET de amiloides para sujetos para los que la probabilidad de ser positivos para A β está por encima de un determinado umbral. En la etapa S50, se usa un modelo de aprendizaje automático para predecir la probabilidad de la conversión del sujeto en EA en el periodo de tiempo del ensayo clínico. En la etapa S60, los sujetos para los que la probabilidad de conversión está por encima de un determinado umbral se seleccionan para su inclusión en el estudio.

La figura 2 muestra un procedimiento en el que dos modelos entrenados, es decir, modelos de aprendizaje automático, uno para predecir la positividad para A β y otro para predecir la probabilidad de conversión, se utilizan en serie, es decir, uno tras otro, para seleccionar sujetos. Estos modelos también se denominan máquinas de aprendizaje entrenadas en el presente documento, y se entenderá que estas máquinas de aprendizaje entrenadas pueden implementarse en hardware independiente o pueden implementarse en el mismo hardware. Alternativamente, cada modelo puede usarse por sí solo para personalizar el procedimiento de selección según sea necesario. De esta manera, los sujetos que no se beneficiarían del DMD que se está investigando no tienen que someterse a pruebas innecesarias, los gastos de los ensayos clínicos pueden reducirse drásticamente y la eficiencia general de los ensayos puede mejorarse. Otra posibilidad es utilizar los modelos en paralelo, usándose los resultados de cada uno de ellos como entrada para la selección de sujetos. Esto se muestra en la figura 3, en la que la etapa S20 es nuevamente una etapa para identificar a los sujetos con un deterioro cognitivo leve mediante una selección convencional de ensayo mediante pruebas psicométricas, la recopilación de datos demográficos de los sujetos y la adquisición de una IRM. En la etapa S70, se recopilan los marcadores de positividad para A β . En la etapa S80, se usa un modelo de aprendizaje automático para predecir la probabilidad de que un sujeto identificado sea positivo para A β . Esta probabilidad se usa como entrada para la selección del sujeto en la etapa S90. En la etapa S100, se recopilan los marcadores de positividad para determinar si el DCL progresa de hecho. En la etapa 110, se usa un modelo de aprendizaje automático para predecir la probabilidad de que un sujeto identificado tenga un DCL que, de hecho, está progresando. Esta probabilidad se usa como otra entrada para la selección del sujeto en la etapa S90.

Como ejemplo, un modelo que predice el estado de A β y un modelo que predice a los sujetos con DCL que progresan rápidamente, en donde los sujetos que progresan rápidamente se definen como aquellos que se convirtieron a un estado probable de EA (pEA) en un periodo de tiempo de 36 meses, se aprendieron a partir de un conjunto de datos de un ensayo clínico de fase III que contenía 232 sujetos con DCL con 87 sujetos positivos para A β y 81 sujetos que se convirtieron a EA en el plazo de los 36 meses posteriores al examen de obtención de imágenes de referencia. El objetivo principal de este ensayo clínico de fase III fue comparar la conversión de pEA en sujetos con DCL con la captación normal y anómala de [18 F]flutemetamol. El seguimiento longitudinal del estado de conversión de los sujetos a pEA lo convierte en un buen conjunto de datos para fines de aprendizaje modelo.

El objetivo del modelo de positividad para amiloides es seleccionar sujetos que tengan probabilidades de tener una gammagrafía de obtención de imágenes mediante PET positiva para amiloides antes de asumir los gastos y las molestias de someter a los posibles participantes del ensayo a una gammagrafía mediante PET. Los sujetos del ensayo de fase III que carecían de una gammagrafía mediante PET de amiloides fueron retirados del conjunto de entrenamiento, dejando 227 sujetos y una tasa de positividad para amiloides del 38 %. El objetivo del modelo de DCL que progresa rápidamente es seleccionar sujetos que tengan probabilidades de convertirse en pEA dentro del alcance de un ensayo clínico de tres años. Los sujetos del ensayo de fase III que carecían de una etiqueta de conversión o los sujetos con una etiqueta de conversión de "no convertidos" en un tiempo inferior a 3 años fueron excluidos del conjunto de entrenamiento, dejando 182 sujetos y una tasa de conversión del 45 %.

La generación de modelos se implementó utilizando los rasgos distintivos disponibles en el estudio de fase III. Los datos de la obtención de imágenes se cuantificaron utilizando un software de cuantificación automatizado que produjo volúmenes cerebrales cuantificados a partir de las imágenes de RM en T1 y valores de SUVr en relación con la región de referencia de Pons a partir de las imágenes mediante PET de amiloides. Estas regiones cuantificadas, en combinación con los datos demográficos, neuropsicométricos y genéticos recopilados, se seleccionaron a la baja mediante algoritmos de selección de rasgos distintivos para su uso en los modelos. La mediana de imputación se utilizó para abordar a los sujetos a los que les faltaban datos distintos de la etiqueta objetivo. La figura 2 representa gráficamente un procedimiento generalizado para usar los modelos.

El modelo de positividad para A β mostró el potencial de mejorar la eficiencia del ensayo al seleccionar sujetos positivos para A β en un 43 %, con una precisión del modelo del 79 % y una especificidad del 85 %, donde la mejora en la eficiencia del ensayo se mide como el valor predictivo positivo (PPV) del modelo en comparación con la inclusión original de sujetos positivos para A β . El modelo de DCL que progresa rápidamente mostró una mejora del 24 % en la

selección de sujetos con DCL que progresan rápidamente, con una precisión del 86 % y una especificidad del 92 %, donde la mejora en la eficacia del ensayo se mide como el PPV del modelo en comparación con los sujetos que se identificaron como positivos para A β únicamente mediante obtención de imágenes mediante PET de amiloides.

- 5 Si bien las métricas de rendimiento del modelo de informes producidas mediante la validación cruzada son un enfoque válido para presentar los resultados del aprendizaje automático, es importante saber que pueden aplicarse a conjuntos de datos independientes y no solo son específicos del conjunto de datos de aprendizaje.

10 Para evaluar esta capacidad, se reconstruyó el modelo de positividad para A β utilizando rasgos distintivos comunes tanto al conjunto de datos de fase III como al conjunto de datos del Estudio insignia sobre el envejecimiento de la obtención de imágenes, biomarcadores y estilo de vida de Australia (AIBL). La validación del AIBL para determinar la positividad para A β con n=551 proporcionó una puntuación de precisión del 75 % con una especificidad del 87 %, lo que demuestra la extrapolación a otras poblaciones y estados patológicos debido a la inclusión de sujetos sanos y con EA en el conjunto de datos del AIBL.

15 El uso del aprendizaje automático para la selección y estratificación de los sujetos en un ensayo clínico puede aumentar la probabilidad de éxito a la hora de demostrar la eficacia de los fármacos y la velocidad a la que están disponibles terapias eficaces con DMD. El rendimiento del modelo demostrado sugiere que la eficiencia de inclusión puede mejorarse en un 50 % o más. Además, la validación del modelo de positividad para A β con el conjunto de datos del AIBL demuestra la viabilidad de extender un modelo aprendido en una población a una población diferente, así como la capacidad de extender las puntuaciones de riesgo de positividad aprendidas en sujetos con DCL solo a sujetos sanos y con EA.

20 Ahora se describirá un procedimiento para construir modelos de acuerdo con un aspecto de la invención. Puede considerarse que el procedimiento implica dos fases principales. La primera es preparar, es decir, “limpiar” los datos. La segunda consiste en evaluar los datos una vez preparados.

25 Los rasgos distintivos que deben incluirse para su consideración pueden provenir de una o más categorías, como la psicometría, la demografía, la genética, PET de A β e IRM en T1. Para la psicometría, los rasgos distintivos pueden incluir datos de CDR, CDR-SB, ADL, MMSE, prueba de fluidez de categoría: animal, prueba de fluidez de categoría: vegetal, ADAS-cog, memorización de dígitos hacia atrás, prueba de sustitución de símbolos de dígitos, memoria lógica II - 30 minutos después de la historia, memoria lógica II - historia posterior inmediata, prueba del trazo, parte A y/o prueba del trazo, parte B. Los rasgos distintivos demográficos pueden incluir educación y/o edad y/o sexo. Los rasgos distintivos genéticos pueden incluir ApoE. Los rasgos distintivos del PET de A β pueden incluir SUVR-PONS prefrontal de A β , SUVR-PONS de A β del cíngulo anterior, SUVR-PONS de A β del cíngulo posterior del precúneo, SUVR-PONS de A β parietal, SUVR-PONS de A β lateral temporal, SUVR-PONS de A β mesial temporal, SUVR-PONS de A β occipital, SUVR-PONS de A β sensitivomotor y/o SUVR-PONS de A β compuesto. La RM en T1 puede incluir el volumen del hipocampo, el volumen del tálamo, el volumen del complejo amigdalino, el volumen del putamen, el volumen del núcleo caudado, el volumen del parahipocampo, el volumen de la corteza entorrinal, el volumen del lóbulo temporal medial, el volumen de los ventrículos, el volumen de la circunvolución recta, el volumen de la circunvolución angular, la materia gris completa y/o la materia blanca completa. En general, se utilizará un subconjunto de estos rasgos distintivos, como ADL, MMSE, prueba de fluencia de categoría: animal, SUVR-PONS de A β compuesto y volumen del hipocampo.

40 En lo que respecta a la preparación de los datos, también puede considerarse como una recopilación de procedimientos, por ejemplo, la selección de rasgos distintivos (estrechando desde un primer conjunto de rasgos distintivos potenciales hasta un subconjunto de rasgos distintivos potenciales) y la ingeniería de rasgos distintivos, la selección de sujetos y la imputación de los rasgos distintivos faltantes. La selección de rasgos distintivos y la ingeniería de rasgos distintivos implican elegir un subconjunto de rasgos distintivos para eliminarlos en primera instancia basándose en el conocimiento a priori sobre estos rasgos distintivos. Por ejemplo, algunos rasgos distintivos pueden formar parte de los criterios de inclusión del ensayo y, por tanto, no proporcionan suficiente variabilidad para su uso en la construcción de modelos. Se sabe que algunos rasgos distintivos se correlacionan fuertemente con otros rasgos distintivos y, por tanto, se eliminan para evitar que los rasgos distintivos altamente correlacionados sesguen el modelo. Por último, algunos rasgos distintivos, mediante la experimentación, pueden resultar inadecuados para su uso como rasgo distintivo en el modelo.

45 Las fases de experimentación de la construcción de modelos también pueden implicar la utilización de la eliminación recursiva de rasgos distintivos, la selección aleatoria de rasgos distintivos del bosque u otros métodos similares conocidos por un experto en la técnica, antes de la fase de construcción del modelo. De esta manera, pueden eliminarse rasgos distintivos de baja importancia para simplificar el modelo. Este procedimiento es útil para construir el modelo de DCL que progresa rápidamente que se describe a continuación. Tras la fase de experimentación, los rasgos distintivos del modelo pueden establecerse en el modelo sin que se realice ninguna selección adicional de rasgos distintivos. La selección de rasgos distintivos puede realizarse, por ejemplo, utilizando las herramientas disponibles en la biblioteca de aprendizaje automático de Scikit-learn disponible en scikit-learn.org.

65

La ingeniería de rasgos distintivos adicionales puede implicar promediar los rasgos distintivos de las exploraciones de RM de diferentes hemisferios cerebrales. Este procedimiento también es útil para construir el modelo de DCL que progresa rápidamente que se describe a continuación. Por ejemplo, el volumen del hipocampo derecho y el volumen del hipocampo izquierdo pueden promediarse para crear un rasgo distintivo de volumen del hipocampo, o los volúmenes de los ventrículos izquierdo y derecho pueden sumarse para crear un rasgo distintivo de volumen de los ventrículos.

En el caso de que se utilicen múltiples conjuntos de datos, es útil utilizar la intersección entre los conjuntos de rasgos distintivos para utilizar solo los rasgos distintivos que son comunes a todos los conjuntos de datos.

Con respecto a la selección de sujetos, para el modelo de DCL que progresa rápidamente, los sujetos que tenían un estado de conversión de “No” antes del punto de tiempo de los 3 años fueron excluidos de la fase de construcción del modelo. Estos sujetos no fueron seguidos durante todo el estudio por un motivo u otro y, por tanto, su estado de conversión no es fiable. Para el modelo de positividad para amiloides, a los sujetos se les asignó un estado negativo (negativo para amiloides) o positivo (positivo para amiloides) en función de su SUVR compuesto en relación con el umbral de 0,62. Los sujetos que carecían de una etiqueta de conversión o una gammagrafía mediante PET que permitiera generar una etiqueta de conversión fueron eliminados de la población de construcción del modelo porque se desconocía su verdadero estado.

Con respecto a la imputación de los rasgos distintivos faltantes, para tener en cuenta los sujetos que carecen de valores para los rasgos distintivos utilizados en el modelo, puede realizarse una mediana de la imputación o puede eliminarse el sujeto. Alternativamente, los sujetos pueden asignarse al valor más común del rasgo distintivo. Por ejemplo, en los casos en los que el estado de ApoE es un rasgo distintivo, a los sujetos que carecen de un estado de ApoE se les puede dar el estado de modo, es decir, la combinación de alelos más común, en lugar de realizar la imputación en la mediana de los datos. Como alternativa, si falta un rasgo distintivo, la variable no se rellena y el modelo se aplica sin el rasgo distintivo en particular. Basándose en el conocimiento de los rasgos distintivos importantes, es deseable que el software de modelado destaque al usuario si determinados rasgos distintivos no están disponibles para sujetos individuales o cohortes de sujetos. En algunos casos, esto puede permitir realizar dichas pruebas (por ejemplo, determinar el estado de ApoE) y añadir los valores correspondientes si se sabe que la precisión de la predicción puede mejorarse.

Una vez preparados los datos, se evalúan. Para construir el modelo de DCL que progresa rápidamente, puede usarse Stratified K-Folds (5 pliegues). Si tanto la positividad para amiloides como la etiqueta de conversión están presentes en el conjunto de datos utilizado para la construcción del modelo, las poblaciones pueden estratificarse de tal manera que tengan distribuciones uniformes de convertidores positivos para amiloides, no convertidores positivos para amiloides, convertidores negativos para amiloides y no convertidores negativos para amiloides. De esta manera, los pliegues pueden estratificarse para que coincidan de manera más similar con la población real de sujetos en lugar de estratificarlos únicamente en función del estado de conversión. La implementación estratificada de k pliegues de Scikit-learn puede utilizarse para este propósito.

La regresión logística también puede utilizarse para crear el modelo de DCL que progresa rápidamente. Por ejemplo, la implementación de Scikit-learn de la regresión logística puede utilizarse para la construcción de modelos, utilizando la implementación de CV de búsqueda en cuadrícula de Scikit-learn para determinar los hiperparámetros óptimos. El modelo puede reconstruirse para cada pliegue con el fin de determinar las estadísticas promedio y, a continuación, el modelo final utilizado puede reconstruirse utilizando todos los datos disponibles.

Las estadísticas del modelo para construir el modelo de DCL que progresa rápidamente pueden calcularse para cada pliegue y, a continuación, promediarse para las estadísticas informadas y mostrarse con intervalos de confianza del 95 %. Las estadísticas calculadas pueden incluir la precisión, la puntuación f1, la especificidad, la coincidencia, pr-AUC, NPV y la precisión.

Para construir el modelo de positividad para amiloides, puede usarse nuevamente Stratified K-Folds (5 pliegues). Si tanto la positividad para amiloides como la etiqueta de conversión están presentes en el conjunto de datos utilizado para la construcción del modelo, las poblaciones pueden estratificarse de tal manera que tengan distribuciones uniformes de convertidores positivos para amiloides, no convertidores positivos para amiloides, convertidores negativos para amiloides y no convertidores negativos para amiloides. De esta manera, los pliegues pueden estratificarse para que coincidan de manera más similar con la población real de sujetos en lugar de estratificarlos únicamente en función de la positividad para amiloides. Si falta una etiqueta de conversión en los conjuntos de datos utilizados para la construcción, entonces los sujetos pueden estratificarse utilizando únicamente la positividad para amiloides para garantizar que cada pliegue tenga una distribución igual de sujetos positivos para amiloides y negativos para amiloides que en la población real del conjunto de datos. La implementación estratificada de k pliegues de Scikit-learn puede utilizarse para este propósito.

La selección secuencial de rasgos característicos de retroalimentación también puede usarse para construir el modelo de positividad para amiloides. La selección de rasgos característicos de retroalimentación puede implementarse dentro de cada pliegue durante la fase de experimentación para determinar los rasgos característicos óptimos para cada

pliegue. Una vez determinados los rasgos característicos óptimos para cada pliegue, los rasgos característicos que se han seleccionado en 3 o más pliegues pueden seleccionarse para su uso en la fase de construcción del modelo. Después de la fase de experimentación, la selección de rasgos característicos de retroalimentación puede eliminarse de las etapas de procesamiento y los rasgos característicos utilizados no pueden seleccionarse de manera algorítmica cada vez. La biblioteca de Python mlxtend (extensión de aprendizaje automático) puede usarse para la implementación de la selección de rasgos característicos de retroalimentación.

También puede usarse un clasificador gaussiano de Naive Bayes para construir el modelo de positividad para amiloides. La implementación de clasificadores gaussianos de Naive Bayes en Scikit-learn puede usarse para este propósito manteniendo las opciones predeterminadas. El modelo puede reconstruirse para cada pliegue con el fin de determinar las estadísticas promedio y, a continuación, el modelo final utilizado puede reconstruirse utilizando todos los datos disponibles.

Las estadísticas del modelo pueden calcularse para cada pliegue y, a continuación, promediarse para las estadísticas informadas y mostrarse con intervalos de confianza del 95 %. Las estadísticas calculadas pueden incluir la precisión, la puntuación f1, la especificidad, la coincidencia, pr-AUC, NPV y la precisión.

Los modelos de aprendizaje automático desarrollados con el propósito de identificar y estratificar los sujetos ideales para los ensayos clínicos mejoran significativamente la capacidad de un procedimiento automatizado para ayudar en los ensayos clínicos, en beneficio tanto de los sujetos como de la agencia que respalda el ensayo. El modelo de positividad para amiloides ha demostrado la capacidad de mejorar la eficiencia de los ensayos hasta en un 43 %, mientras que el modelo de DCL que progresa rápidamente, ha demostrado la capacidad de estratificar a los sujetos de tal manera que se aumente la eficiencia de conversión en un 17 % a los tres años.

Estos modelos permiten un ajuste preciso para satisfacer las necesidades específicas de un ensayo clínico; los modelos pueden reconstruirse para priorizar diversas medidas estadísticas, como la sensibilidad, la F1 o la precisión, y el umbral de riesgo del sujeto puede modificarse de tal manera que se dirija a los sujetos que presentan la patología ideal para un ensayo en particular. Además, los modelos pueden usarse individualmente o en serie para presentar un ensayo clínico con un enfoque escalonado para incluir sujetos en un ensayo antes de gastar recursos en sujetos que no son adecuados para un ensayo. Por ejemplo, el modelo de positividad para amiloides podría usarse para seleccionar menos a los sujetos a los que se les realizará una gammagrafía de PET de amiloides y, a continuación, el modelo de DCL que progresa rápidamente podría determinar si es probable que el sujeto progrese a una velocidad ideal durante la duración del ensayo antes de incluir al sujeto en un protocolo prolongado y costoso. Si bien los modelos tienen la capacidad de combinarse de esta manera, también pueden emplearse individualmente sin sacrificar el rendimiento del modelo de ninguna manera.

Pueden incluirse más conjuntos de datos en la fase de construcción del modelo. Además, la construcción de modelos que utilice datos de obtención de imágenes de RM y PET sin procesar podría usarse como complemento o para reemplazar los algoritmos de cuantificación automatizados descritos en el presente documento. El término “datos sin procesar” pretende comprender datos directos de píxeles o vóxeles.

Las enseñanzas de la presente descripción pueden implementarse como una aplicación basada en la web. Esto se muestra en la figura 4. Un ordenador cliente 400 que incluye un navegador web o una aplicación dedicada 405 se comunica con una nube privada virtual 410 a través de un sistema de información de contenido distribuido 420, tal como Internet. La nube privada virtual 410 puede incluir una capa de aplicación 430 para proporcionar una interfaz a un marco de ejecución 440. El marco de ejecución 440 usa modelos entrenados de un almacén modelo 450 para operar con los datos suministrados a través de la capa de aplicación 430 y almacenados en la base de datos 460 para proporcionar, por ejemplo, la probabilidad de que un paciente relacionado con la información proporcionada a la capa de aplicación 430 a través del navegador web 400 desarrolle EA.

El ejemplo anterior se refiere al uso del sistema para seleccionar sujetos para estudios clínicos basándose en su pronóstico, pero también puede usarse en otras aplicaciones. Como otro ejemplo, el sistema puede usarse para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto en el que un sistema informático está adaptado para usar una máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad y visualizar una indicación de la predicción, en donde no se usan los datos obtenidos de un procedimiento de obtención de imágenes moleculares sobre el sujeto. La patología de la enfermedad puede ser la positividad para beta amiloide (A β) en el cerebro del sujeto. Los primeros datos médicos pueden incluir los resultados de las pruebas cognitivas del sujeto, los resultados de las pruebas cognitivas del sujeto, la edad del sujeto, el nivel educativo del sujeto o alguna combinación de estos datos.

Como otro ejemplo, el sistema médico podría usarse para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto. El sistema incluiría un sistema informático dispuesto para recibir los primeros datos médicos del sujeto y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada en segundos datos médicos y una segunda máquina de aprendizaje entrenada en terceros datos médicos, adaptándose el sistema informático para usar la primera máquina de aprendizaje entrenada y la segunda máquina de aprendizaje para proporcionar una predicción de la patología de la enfermedad. Los primeros datos médicos pueden comprender datos

obtenidos de un procedimiento de obtención de imágenes moleculares sobre el sujeto y la primera máquina de aprendizaje entrenada ha sido entrenada, al menos en parte, con datos obtenidos de uno o más procedimientos de obtención de imágenes moleculares.

Por ejemplo, como se indicó anteriormente, el diagnóstico de la enfermedad de Parkinson (EP) en pacientes que tienen síntomas anómalos o muestran una falta de respuesta a la medicación puede ser difícil. La obtención de imágenes con DaTscan™ puede utilizarse en un esfuerzo por obtener un diagnóstico preciso mediante la determinación de la pérdida de actividad dopaminérgica. Además, las técnicas de obtención de imágenes *in vivo* que utilizan radiofármacos como la PET y la tomografía de emisión monofotónica (SPECT) son procedimientos de diagnóstico relativamente costosos y requieren muchos recursos. El uso de un aprendizaje entrenado, tal como se describió anteriormente, utilizando los resultados de la gammagrafía mediante la obtención de imágenes *in vivo* como una de las técnicas de rasgos distintivo mejoraría la capacidad de identificar a los sujetos que es probable que tengan un resultado que sugiera un estado patológico. Para la EP, sería ventajoso identificar a los sujetos con más probabilidades de tener deficiencias dopaminérgicas antes de llevar a cabo la obtención de imágenes moleculares.

Como otro ejemplo, las enseñanzas del presente documento podrían usarse para implementar un método para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto, en donde los datos médicos de la cohorte se recopilan y almacenan en una memoria electrónica para un primer conjunto de sujetos que tienen resultados conocidos para la enfermedad y un sistema informático está dispuesto para recibir los datos médicos del sujeto e incluye una máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la cohorte, adaptándose el sistema informático para usar la máquina de aprendizaje entrenada, proporcionar una predicción de la patología de la enfermedad basándose, al menos en parte, en los datos médicos del sujeto para proporcionar una predicción de la patología de la enfermedad.

Como otro ejemplo, las enseñanzas del presente documento podrían usarse para implementar un método para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto en donde los datos médicos de la primera cohorte se recopilan y almacenan en una memoria electrónica para un primer conjunto de sujetos que tienen resultados conocidos para la enfermedad y los datos médicos de la segunda cohorte se recopilan y almacenan en una memoria electrónica para un segundo conjunto de sujetos que tienen resultados conocidos para la enfermedad, y el sistema informático usa los datos médicos del sujeto para el sujeto e incluye un primera máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la primera cohorte y una segunda máquina de aprendizaje entrenada que se ha entrenado en los datos médicos de la segunda cohorte. El sistema informático usa la primera y la segunda máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad basándose, al menos en parte, en los datos médicos del sujeto, y visualizar una indicación de la predicción. Los datos médicos de la primera cohorte pueden comprender tipos de datos diferentes de los tipos de datos de los datos médicos de la segunda cohorte. Los datos médicos de la primera cohorte pueden comprender tipos de datos al menos parcialmente iguales a los tipos de datos de los datos médicos de la segunda cohorte.

Como otro ejemplo las enseñanzas en el presente documento podrían usarse para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto y/o para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto.

Las enseñanzas de la presente descripción también pueden aplicarse para determinar un pronóstico para un paciente con EA. Puede administrarse un agente de contraste de proteína amiloide a un paciente y, a continuación, pueden obtenerse imágenes de los depósitos de proteína amiloide en el paciente para obtener datos de imagen. Los datos de imagen pueden correlacionarse, junto con otros datos, con los datos de un conjunto de entrenamiento de pacientes con pronósticos conocidos para la EA.

Las enseñanzas de la presente descripción también pueden aplicarse al tratamiento de un paciente con EA. Puede usarse un ordenador para comparar los datos de obtención de imágenes y otros datos obtenidos del paciente con un conjunto de entrenamiento que comprende datos de obtención de imágenes y otros datos con pacientes para pronósticos conocidos de EA con el fin de clasificar al paciente en una cohorte de enfermedades. Los resultados de la comparación pueden usarse para identificar un tratamiento farmacológico conocido por mejorar los resultados de los pacientes con la EA dentro de la cohorte de la enfermedad, y luego puede tratarse al paciente con una cantidad eficaz del tratamiento farmacológico identificado.

Las enseñanzas de la presente descripción también pueden aplicarse al uso de datos médicos de uno o más sujetos que tienen resultados desconocidos para la EA para determinar cuáles de los sujetos están en riesgo de desarrollar EA. Los datos médicos se suministran a un sistema informático configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación. Las indicaciones se combinan para proporcionar una identificación de qué sujetos con resultados desconocidos para la EA están en riesgo de desarrollar EA dentro de un periodo de tiempo definido.

Las enseñanzas de la presente descripción también pueden aplicarse para clasificar a los sujetos como personas con deterioro cognitivo leve (DCL) o enfermedad de Alzheimer (EA). Puede estar dispuesto un ordenador para recibir datos

médicos de uno o más sujetos con una clasificación desconocida para DCL o EA y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para proporcionar una identificación de qué sujetos tienen DCL y qué sujetos tienen EA. El sistema informático puede estar dispuesto además para proporcionar, a los sujetos clasificados como DCL una clasificación adicional de si el DCL es DCL temprano o DCL avanzado.

Las enseñanzas de la presente descripción también pueden aplicarse a un sistema médico para clasificar a los sujetos que tienen enfermedad de Alzheimer (EA) o alguna otra forma de demencia, en donde el sistema informático está dispuesto para recibir datos médicos de uno o más sujetos con una clasificación desconocida para DCL o EA y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para proporcionar una identificación de qué sujetos tienen EA y qué sujetos tienen otra forma de demencia.

Las enseñanzas de la presente descripción también pueden aplicarse a un sistema médico para identificar indicaciones adicionales para un fármaco, en donde el sistema informático está dispuesto para recibir datos médicos de uno o más sujetos que toman el fármaco y configurado, al menos en parte, como una primera máquina de aprendizaje entrenada que proporciona una primera indicación y una segunda máquina de aprendizaje entrenada que proporciona una segunda indicación y para combinar la primera y la segunda indicación para determinar si al fármaco se le puede asignar una indicación además de una indicación existente.

Esta descripción escrita usa ejemplos para describir la invención, que incluye el mejor modo, y para permitir que cualquier experto en la técnica ponga en práctica la invención, incluyendo la fabricación y el uso de cualquier dispositivo o sistema y realizar cualquier método incorporado. El alcance que puede patentarse de la invención está definido por las reivindicaciones.

REIVINDICACIONES

1. Un sistema médico para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto, comprendiendo el sistema:
5 un sistema informático (100) dispuesto para recibir primeros datos médicos del sujeto y configurado, al menos en parte, como una máquina de aprendizaje entrenada que se ha entrenado en segundos datos médicos, adaptándose el sistema informático (100) para usar la máquina de aprendizaje entrenada para proporcionar una predicción de la patología de la enfermedad, en donde los
10 segundos datos médicos utilizados para entrenar la máquina de aprendizaje entrenada comprenden una gammagrafía mediante tomografía por emisión de positrones, PET, de amiloides para cada uno de una pluralidad de sujetos, en donde un subconjunto de la pluralidad de sujetos son positivos para beta amiloides, A β ; y
15 una pantalla (190) para visualizar una indicación de la predicción, en donde los primeros datos médicos no comprenden datos obtenidos de un procedimiento de obtención de imágenes moleculares sobre el sujeto; en donde la patología de la enfermedad es la positividad para A β en el cerebro del sujeto.
2. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos comprenden la referencia de las actividades de la vida diaria (ADL), memorización de dígitos hacia atrás, memoria lógica II 30 minutos después de la referencia de la historia, prueba del trazo, parte A (segundos), la educación, el sexo masculino, el volumen del núcleo caudado izquierdo, el volumen del complejo amigdalino derecho y el volumen del núcleo caudado derecho.
- 25 3. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos comprenden la edad, el sexo, la minipuntuación mental (MMSE), la clasificación de la demencia clínica (CDR), la suma de casillas de la calificación de la demencia clínica (CDR-SB), el estado de las pruebas genéticas de ApoE, los volúmenes cerebrales regionales izquierdo y derecho del hipocampo, el parahipocampo, el complejo amigdalino, la corteza entorrinal, el lóbulo temporal medial, la circunvolución recta, los ventrículos y la circunvolución angular.
- 30 4. Un sistema médico según la reivindicación 1, en donde el resultado clínico es una progresión rápida del DCL.
5. Un sistema médico según la reivindicación 4, en donde los primeros datos médicos comprenden la referencia de las ADL, la referencia del miniexamen del estado mental (MMSE), la prueba de fluidez de categoría: categoría animal, la razón de valores estandarizados de captación de amiloide compuesto (SUVR) (pons) y el volumen del hipocampo.
- 35 6. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos incluyen los resultados de las pruebas cognitivas del sujeto.
7. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos incluyen la edad del sujeto.
- 45 8. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos incluyen el número de años de educación del sujeto.
9. Un sistema médico según la reivindicación 1, en donde los primeros datos médicos incluyen los resultados de las pruebas genéticas de ApoE
- 50 10. Un sistema médico según la reivindicación 1 para su uso en un método para predecir una patología de la enfermedad en un sujeto que tiene un estado cognitivo incierto.
11. Un sistema médico según la reivindicación 1 para su uso en un método para predecir un resultado clínico en un sujeto que tiene un estado cognitivo incierto.
- 55 12. Un sistema médico según la reivindicación 1, en donde el sujeto presenta problemas subjetivos de memoria (SMC) o padece un deterioro cognitivo leve (DCL) y está siendo investigado para determinar la posible aparición de la enfermedad de Alzheimer (EA).
- 60 13. Un sistema médico según la reivindicación 1, en donde el sujeto es cognitivamente normal.

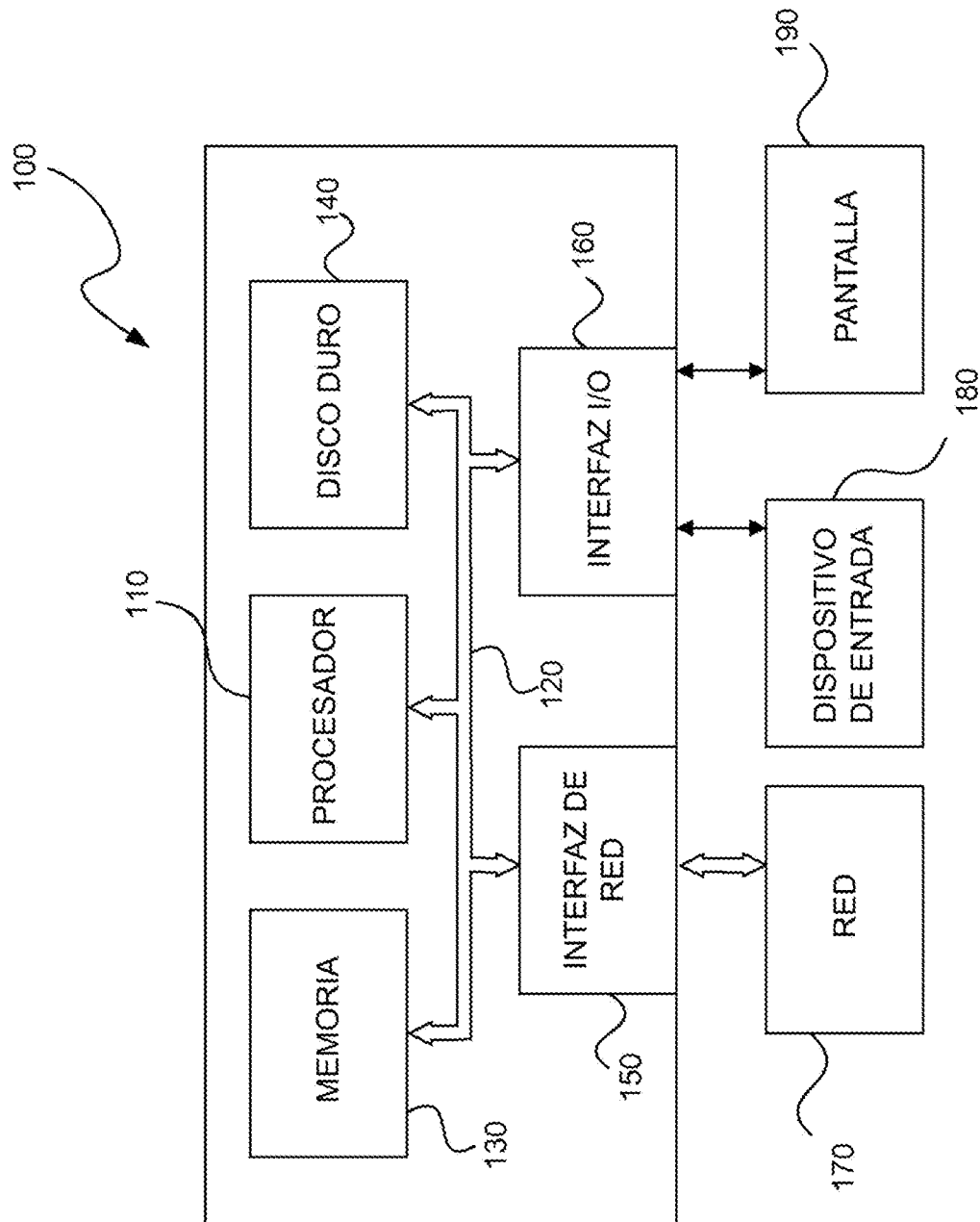


Figura 1

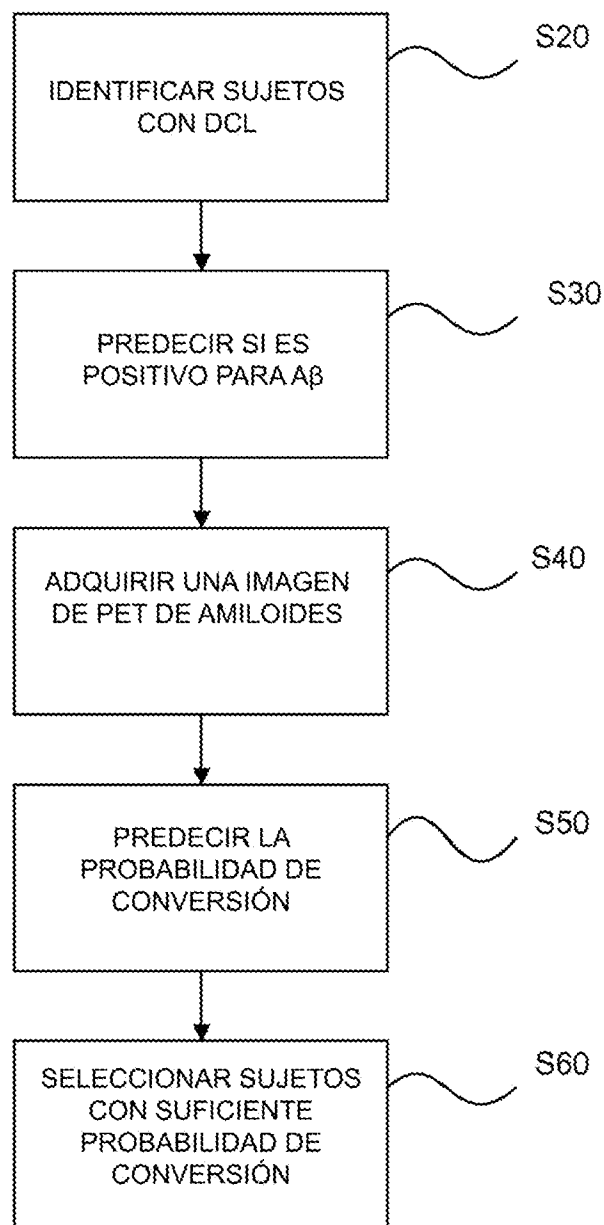


Figura 2

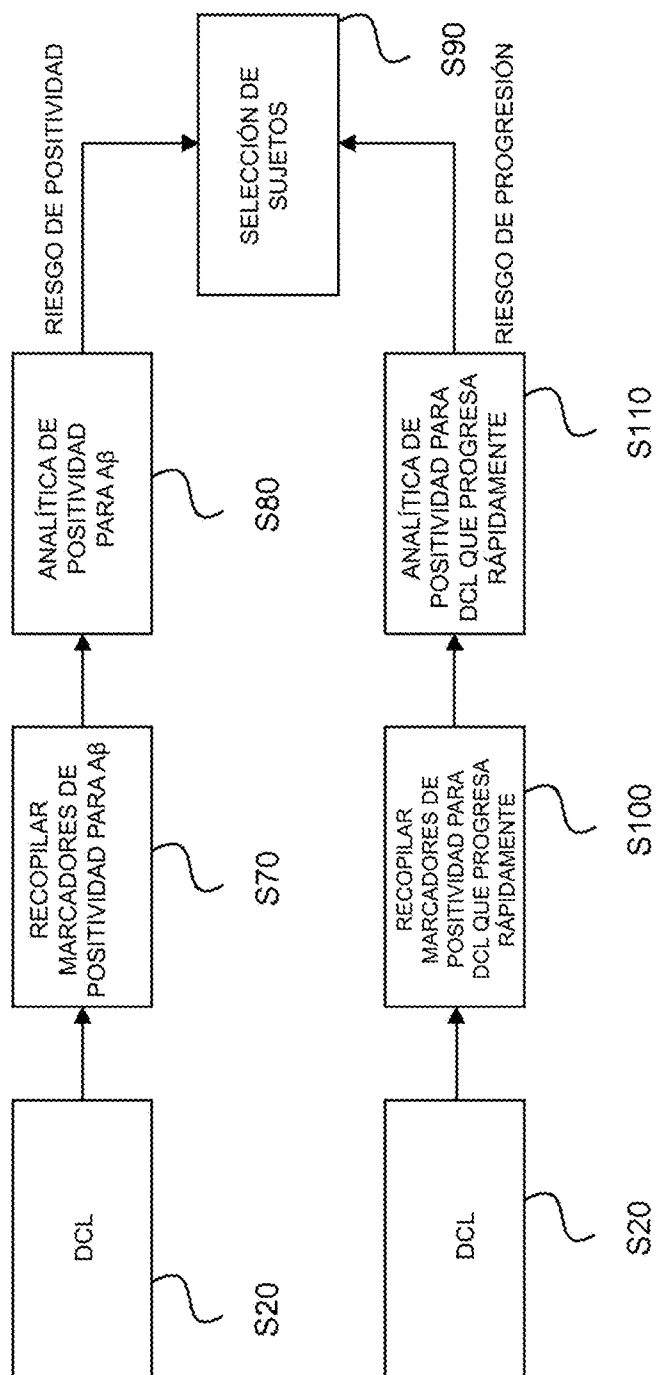


Figura 3

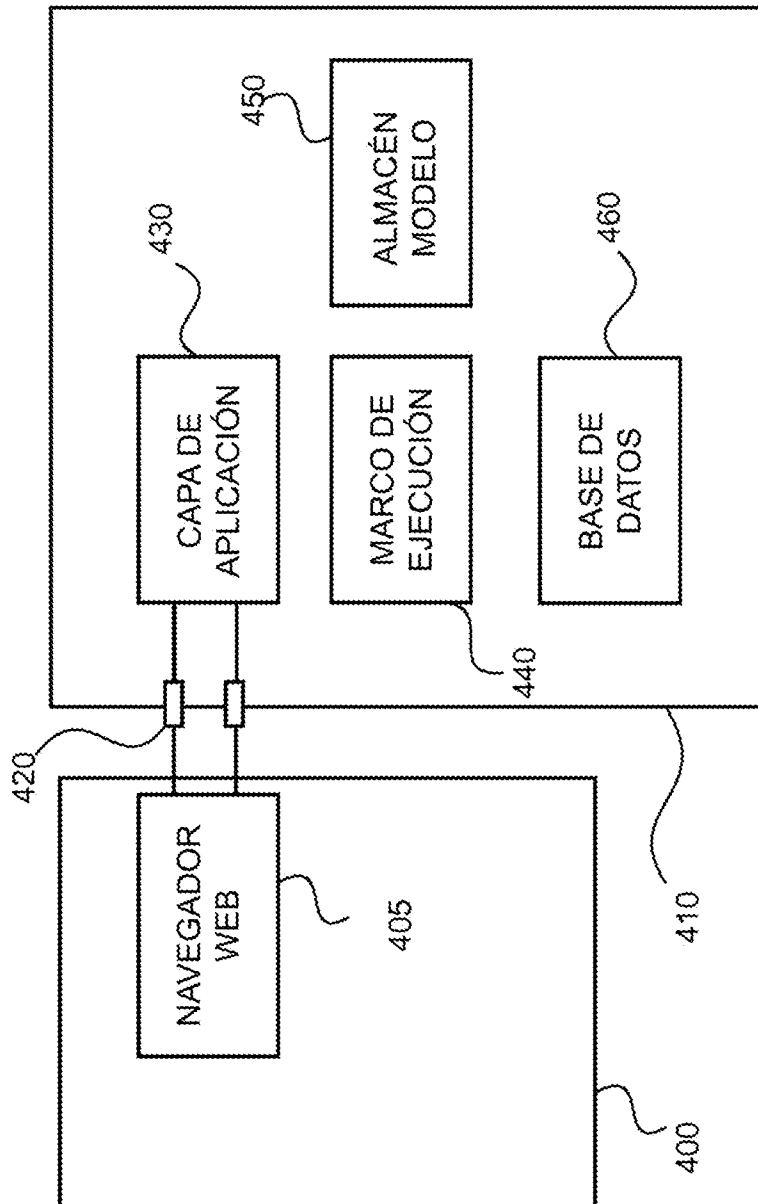


Figura 4