

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 655 184**

51 Int. Cl.:

A61B 5/00 (2006.01)

G06F 19/00 (2011.01)

G01N 33/574 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.05.2013 PCT/IL2013/050368**

87 Fecha y número de publicación internacional: **07.11.2013 WO13164823**

96 Fecha de presentación y número de la solicitud europea: **02.05.2013 E 13784595 (4)**

97 Fecha y número de publicación de la concesión europea: **11.10.2017 EP 2844131**

54 Título: **Métodos y sistemas de evaluación de un riesgo de un cáncer gastrointestinal**

30 Prioridad:

03.05.2012 US 201261641979 P

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

19.02.2018

73 Titular/es:

**MEDIAL RESEARCH LTD. (100.0%)
11 HaZayit Street
4592000 Kfar-Malal, IL**

72 Inventor/es:

**KALKSTEIN, NIR;
KINAR, YARON;
SHALEV, VARDA;
CHODICK, GABRIEL y
GOLDSHTEIN, INBAL**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 655 184 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

Métodos y sistemas de evaluación de un riesgo de un cáncer gastrointestinal

5 Campo y antecedentes de la invención

La presente invención, en algunas realizaciones de la misma, se refiere al diagnóstico del cáncer y, más en particular, pero no exclusivamente, a métodos y sistemas de evaluación de un riesgo de cáncer.

10 Un diagnóstico de cáncer colorrectal incluye el diagnóstico basado en la reacción inmunológica de sangre oculta en heces, el diagnóstico por colonoscopia y similares. Sin embargo, el diagnóstico basado en un ensayo de sangre oculta en heces no sirve como diagnóstico definitivo y la mayoría de las personas con resultado positivo tienen falsos positivos. Además, con respecto al cáncer colorrectal temprano, existe la preocupación de que tanto la sensibilidad de detección como la especificidad de la detección disminuyen en el diagnóstico basado en un ensayo de sangre oculta en heces o el diagnóstico por colonoscopia. En particular, el cáncer temprano en el colon del lado derecho con frecuencia se pasa por alto cuando se diagnostica mediante un ensayo de sangre oculta en heces. La formación de imágenes de diagnóstico mediante TC (tomografía computarizada), RM (resonancia magnética), TEP (tomografía computarizada por emisión de positrones) o similares, no es adecuada para el diagnóstico del cáncer colorrectal.

20 Por otra parte, la biopsia colorrectal mediante colonoscopia sirve como diagnóstico definitivo, pero es un examen muy invasivo y la implementación del examen por colonoscopia en la etapa de detección no es práctica. Además, el diagnóstico invasivo tal como la colonoscopia es una carga para los individuos, tal como el dolor que la acompaña, y también puede haber un riesgo de sangrado en el examen, o similares.

25 Durante los últimos años, se han desarrollado algunos métodos nuevos para el diagnóstico del cáncer colorrectal. Por ejemplo, la Publicación de Patente de los EE.UU. N.º 2010/0009401 describe un método de evaluación del cáncer colorrectal, donde se miden datos de concentraciones de aminoácidos acerca del valor de la concentración de aminoácidos en sangre recogida de un sujeto que se ha de evaluar, y se evalúa un estado de cáncer colorrectal en el sujeto basándose en el valor de la concentración de al menos uno de Arg, Cys, Om, Trp, Glu, ABA, Val, Phe, Leu, Gln, Ile e His contenido en los datos de concentración de aminoácidos medidos del sujeto.

35 La Patente de los EE.UU. N.º 6.059.724 enseña un sistema informático para predecir la salud futura de individuos que comprende: (a) un ordenador que comprende un procesador que contiene una base de datos de valores de biomarcadores adquiridos longitudinalmente de miembros individuales de una población de ensayo, subpoblación D de miembros de ayuda que se identifica como que ha adquirido una afección biológica especificada en un periodo de tiempo o intervalo de edad especificados y una subpoblación D que se identifica como que no ha adquirido la afección biológica especificada en el periodo de tiempo o intervalo de edad especificados; y (b) un programa de ordenador que incluye etapas para: (1) seleccionar entre dichos biomarcadores un subconjunto de biomarcadores para discriminar entre los miembros pertenecientes a las subpoblaciones D y D, en el que el subconjunto de biomarcadores se selecciona basándose en las distribuciones de los valores de biomarcadores de los miembros individuales de la población de ensayo; y (2) usar las distribuciones de los biomarcadores seleccionados para desarrollar un procedimiento estadístico que es capaz de usarse para: (i) clasificar miembros de la población de ensayo como que pertenecen a una subpoblación PD que tiene una alta probabilidad prescrita de adquirir la afección biológica especificada en el periodo de tiempo o intervalo de edad especificados, o como que pertenecen a una subpoblación PD que tiene una baja probabilidad prescrita de adquirir la afección biológica especificada en el periodo de tiempo o intervalo de edad especificados; o (ii) estimar cuantitativamente, para cada miembro de la población de ensayo, la probabilidad de adquirir la afección biológica especificada en el periodo de tiempo o intervalo de edad especificados.

50 La Solicitud de Patente de los EE.UU. N.º 2006/173717 enseña que puede usarse información de la historia clínica personal y familiar para evaluar el riesgo familiar de enfermedad. Por ejemplo, puede recogerse información acerca de la historia de la enfermedad de una persona y los familiares de la persona de primer y segundo grado y, después, puede analizarse para determinar el riesgo familiar de enfermedades comunes, tales como las enfermedades coronarias, el ictus, la diabetes de tipo 2 y el cáncer colorrectal, de mama y de ovario. El riesgo familiar de enfermedad evaluado después puede usarse por los investigadores para estimar mejor la contribución de la historia personal y la historia familiar a la etiología y la historia natural de una enfermedad de interés, y por los consumidores y profesionales de la salud para determinar las recomendaciones para el tratamiento, la prevención y la detección de enfermedades que se personalizan y se dirigen al riesgo familiar.

60 La Publicación de Patente Internacional N.º WO2011/063389 enseña métodos útiles para la normalización de cualquier biomarcador en las plaquetas. Esto tiene una aplicación en cualquier método en el que se desee determinar o comparar el nivel de un biomarcador, por ejemplo, para métodos de diagnóstico o pronóstico relacionados con un biomarcador de interés. El uso de un enfoque de este tipo puede permitir la evaluación del estado de la enfermedad (por ejemplo, el estado angiogénico) de un individuo con menos error que un valor de expresión que no se normaliza o que se normaliza a los niveles totales de proteína. También se proporcionan

métodos para seleccionar una proteína de normalización para normalizar biomarcadores en una muestra, por ejemplo, una muestra de plaquetas.

5 KOULAOUZIDIS ANASTASIOS ET AL: "*Hemoglobin and red cell indices as predictors of gastrointestinal malignancy*", *AMERICAN JOURNAL OF GASTROENTEROLOGY*, ELSEVIER SCIENCE INC, EE.UU. vol. 101, n.º 9 Suplemento S2, 1 de septiembre de 2006 (01-09-2006), página S134, XP008177675, enseña que la hemoglobina y los índices de glóbulos rojos pueden ser factores predictivos de malignidad gastrointestinal.

10 LALITHA MAHADAVAN ET AL: "*Exfoliated colonocyte DNA levels and clinical features in the diagnosis of colorectal cancer: a cohort study in patients referred for investigation*", *COLORECTAL DISEASE*, vol. 14, n.º 3, 7 de febrero de 2012 (07-02-2012), páginas 306-313, enseña que en una selección de pacientes para la investigación de sospecha de cáncer colorrectal puede medirse ADN aislado de células exfoliadas recogidas del recto. El artículo concluye que la cuantificación de ADN exfoliado de material celular rectal es prometedora en el diagnóstico de cáncer colorrectal, pero esto requiere confirmación en un estudio más grande.

15 W CHUA ET AL: "*Neutrophil/lymphocyte ratio predicts chemotherapy outcomes in patients with advanced colorectal cancer*", *BRITISH JOURNAL OF CANCER*, vol. 104, n.º 8, 29 de Marzo de 2011 (29-03-2011), páginas 1288-1295, enseña que los avances en el tratamiento del cáncer colorrectal metastásico (CCRm) en la última década han mejorado significativamente la supervivencia; sin embargo, no se han identificado biomarcadores simples para predecir la respuesta o toxicidad, que sean aplicables a todos los escenarios de oncología en la comunidad en todo el mundo. El uso de marcadores inflamatorios basados en recuentos diferenciales de glóbulos blancos, tales como la relación de neutrófilos/linfocitos (RNL), puede ser de biomarcadores simples y fácilmente disponibles. El artículo destaca la RNL como biomarcador clínico potencialmente útil de la respuesta inflamatoria sistémica en la predicción de resultados clínicamente significativos en dos cohortes independientes. Los resultados de este estudio también han confirmado la importancia de una respuesta inflamatoria sistémica crónica que influye en los resultados clínicos en pacientes con CCRm.

20 La Publicación de Patente Internacional N.º WO 2006/102526 A2 enseña métodos de determinación del estado del cáncer colorrectal en un sujeto.

30 La Publicación de Patente Internacional N.º WO 2011/060098 A1 enseña métodos útiles en la predicción o determinación de un riesgo de que un paciente desarrolle displasia/cáncer y/o complicaciones inflamatorias después de un procedimiento quirúrgico en el que se retira el colon y se crea una bolsa interna.

35 Sumario de la invención

De acuerdo con algunas realizaciones de la presente invención, se proporciona un método informatizado de evaluación de riesgo de cáncer gastrointestinal. El método comprende la generación de un conjunto de características que comprende una pluralidad de resultados de ensayos de sangre actuales a partir de sangre
40 recogida de un individuo objetivo, proporcionando al menos un clasificador generado de acuerdo con un análisis de una pluralidad de resultados de ensayos de sangre históricos respectivos de cada uno de una pluralidad de individuos muestreados; y la evaluación, usando un procesador, un riesgo de cáncer gastrointestinal del individuo objetivo mediante la clasificación del conjunto de características usando el al menos un clasificador. Cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de al menos uno de los
45 siguientes ensayos de sangre: glóbulos rojos (GR), hemoglobina (HGB) y hematocrito (HCT) y al menos un resultado de los siguientes ensayos de sangre: hemoglobina (HCM) y concentración de hemoglobina corpuscular media (CHCM). Opcionalmente, los resultados de los ensayos de sangre se extraen de un ensayo de recuento sanguíneo completo (RSC).

50 Opcionalmente, el conjunto de características comprende una edad del individuo objetivo; en el que el al menos un clasificador se genera de acuerdo con un análisis de la edad de cada uno de una pluralidad de individuos muestreados.

55 Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de amplitud de distribución de glóbulos rojos (ADR).

Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de hematocrito de plaquetas (PCT).

60 Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de volumen celular medio (VCM).

Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende al menos uno de los siguientes ensayos de sangre: recuento de glóbulos blancos sanguíneos - GBS (RSC); volumen de plaquetas medio (VPM); células medias; recuento de plaquetas (RSC); recuento de eosinófilos; porcentaje

neutrófilos; porcentaje monocitos; porcentaje de eosinófilos; recuento de basófilos; y recuento de neutrófilos; recuento de monocitos.

5 Opcionalmente, el al menos un clasificador comprende un miembro de un grupo que consiste en: un clasificador ponderado de regresión lineal, un clasificador de vecinos más cercanos a K (VCK) y un clasificador de bosque aleatorio.

10 Opcionalmente, el conjunto de características comprende al menos una característica demográfica del individuo objetivo y el al menos un clasificador generado de acuerdo con un análisis de la al menos una característica demográfica respectiva de cada uno de la pluralidad de individuos muestreados.

15 Opcionalmente, el método comprende adicionalmente seleccionar el al menos un clasificador de acuerdo con al menos una característica demográfica del individuo entre una pluralidad de clasificadores cada uno generado de acuerdo con una pluralidad de resultados de ensayos de sangre históricos respectivos de individuos muestreados que tienen al menos una característica demográfica diferente.

20 Opcionalmente, la pluralidad de resultados de ensayos de sangre comprende al menos un resultado de la siguiente pluralidad de ensayos de sangre: los resultados de ensayos de sangre bioquímicos pueden incluir cualquiera de los siguientes resultados de ensayos de sangre: albúmina, calcio, cloruro, colesterol, creatinina, lipoproteína de alta densidad (HDL), lipoproteína de baja densidad (LDL), potasio, sodio, triglicéridos, urea y/o ácido úrico.

25 De acuerdo con algunas realizaciones de la presente divulgación, se proporciona un sistema de evaluación de cáncer gastrointestinal. El sistema comprende un procesador, una unidad de memoria que almacena al menos un clasificador generado de acuerdo con un análisis de una pluralidad de resultados de ensayos de sangre históricos de cada uno de una pluralidad de individuos muestreados, y una unidad de entrada que recibe una pluralidad de resultados de ensayos de sangre actuales tomados de una sangre de un individuo objetivo, y un módulo de evaluación de cáncer gastrointestinal que evalúa, usando el procesador, un riesgo de cáncer gastrointestinal del individuo objetivo mediante la clasificación, usando el al menos un clasificador, un conjunto de características extraídas de la pluralidad de resultados de ensayos de sangre actuales. La pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de al menos uno de la siguiente pluralidad de ensayos de sangre: glóbulos rojos sanguíneos (GRS), hemoglobina (HGB) y hematocrito (HCT) y al menos un resultado de los siguientes ensayos de sangre: hemoglobina (HCM) y concentración de hemoglobina corpuscular media (CHCM).

30

35 Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de amplitud de distribución de glóbulos rojos (ADR).

Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de hematocrito de plaquetas (PCT).

40 Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de volumen celular medio (VCM).

45 De acuerdo con algunas realizaciones de la presente divulgación, se proporciona un método de generación de un clasificador para una evaluación de riesgo de CCR. El método comprende proporcionar una pluralidad de resultados de ensayos de sangre históricos de cada uno de una pluralidad de individuos muestreados, generar un conjunto de datos que tiene una pluralidad de conjuntos de características, cada conjunto generado de acuerdo con la pluralidad respectiva de resultados de ensayos de sangre históricos de otro de la pluralidad de individuos muestreados, generar al menos un clasificador de acuerdo con un análisis del conjunto de datos y emitir al menos un clasificador.

50 Opcionalmente, la generación comprende calcular y añadir al menos una versión manipulada de un resultado de ensayo de sangre histórico tomado de una pluralidad respectiva de resultados de ensayos de sangre históricos como una característica del conjunto respectivo de características.

55 Opcionalmente, la generación comprende ponderar cada una del conjunto de características de acuerdo con una fecha de la pluralidad respectiva de resultados de ensayos de sangre históricos.

Opcionalmente, la generación comprende filtrar la pluralidad de conjuntos de características para eliminar valores atípicos de acuerdo con un umbral máximo de desviación típica.

60 Opcionalmente, la pluralidad de conjuntos de características se pondera de acuerdo con una fecha de la pluralidad respectiva de resultados de ensayos de sangre históricos.

65 Opcionalmente, la pluralidad de resultados de ensayos de sangre de al menos uno de los siguientes ensayos de sangre: glóbulos rojos sanguíneos (GRS), hemoglobina (HGB) y hematocrito (HCT), y al menos un resultado de los siguientes ensayos sanguíneos: hemoglobina (HCM) y concentración de hemoglobina corpuscular media (CHCM).

Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de amplitud de distribución de glóbulos rojos (ADR).

5 Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de hematocrito de plaquetas (PCT).

Opcionalmente, cada uno de la pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de volumen celular medio (VCM).

10 Más opcionalmente, el método comprende adicionalmente la suma de al menos un parámetro demográfico de cada uno de la pluralidad de individuos muestreados con respecto a un conjunto de características.

Más opcionalmente, el parámetro demográfico al menos una es un miembro de un grupo que consiste en género, edad, zona residencial, la raza y característica socio-económico.

15 Más opcionalmente, la generación comprende calcular y añadir al menos una versión manipulada del al menos un parámetro demográfico como una característica con respecto al conjunto de características.

20 A menos que se defina lo contrario, todos los términos técnicos y/o científicos utilizados en el presente documento tienen el mismo significado que el entendido comúnmente por un experto habitual en la materia a la que pertenece la invención. Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o el ensayo de realizaciones de la invención, se describen a continuación métodos y/o materiales de ejemplo. En caso de conflicto, la memoria descriptiva de la patente, incluyendo las definiciones, prevalecerá. Además, los materiales, métodos y ejemplos son solamente ilustrativos y no pretenden ser limitantes necesariamente.

25 La implementación del método de la invención y/o el sistema de realizaciones de la divulgación puede implicar la realización o la terminación de tareas seleccionadas manualmente, automáticamente o una combinación de las mismas. Además, de acuerdo con la instrumentación y el equipo reales de realizaciones del método de la invención y/o el sistema de la divulgación podrían implementarse varias tareas seleccionadas mediante hardware, mediante software o mediante firmware o mediante una combinación de los mismos usando un sistema operativo.

30 Por ejemplo, podría implementarse hardware para la realización de tareas seleccionadas de acuerdo con realizaciones de la invención en forma de un chip o un circuito. Como software, podrían implementarse tareas seleccionadas de acuerdo con realizaciones de la invención como una pluralidad de instrucciones de software ejecutadas por un ordenador usando cualquier sistema operativo adecuado. En una realización de ejemplo de la invención, se realiza una o más tareas de acuerdo con realizaciones ejemplares del método y/o sistema como se describe en el presente documento mediante un procesador de datos, tal como una plataforma informática para la ejecución de una pluralidad de instrucciones. Opcionalmente, el procesador de datos incluye una memoria volátil para el almacenamiento de instrucciones y/o datos y/o un almacenamiento no volátil, por ejemplo, un disco duro magnético y/o medios extraíbles, para el almacenamiento de instrucciones y/o datos. Opcionalmente, también se proporciona una conexión de red. También se proporcionan opcionalmente una pantalla y/o un dispositivo de entrada de usuario tal como un teclado o un ratón.

45 Breve descripción de los dibujos

En el presente documento se describen algunas realizaciones de la invención, a modo de ejemplo solamente, con referencia a los dibujos adjuntos. Ahora, con referencia específica a los dibujos en detalle, se hace hincapié en que los detalles que se muestran son a modo de ejemplo y con fines de análisis ilustrativo de realizaciones de la invención. En este sentido, la descripción tomada con los dibujos hace evidente para los expertos en la materia cómo pueden ponerse en práctica las realizaciones de la invención.

En los dibujos:

55 la FIG. 1 es un diagrama de flujo de un método de generación de uno o más clasificadores para la estimación de una puntuación de riesgo de cáncer gastrointestinal de acuerdo con un análisis de una pluralidad de registros de individuos, de acuerdo con algunas realizaciones de la presente invención;

60 la FIG. 2 es una ilustración esquemática de un sistema para la generación de uno o más clasificadores, por ejemplo, mediante la implementación del método representado en la FIG. 1, de acuerdo con algunas realizaciones de la presente invención;

la FIG. 3 es un gráfico de una curva de una característica operativa receptora (COR), de acuerdo con algunas realizaciones de la presente invención;

las FIG. 4A-4C son tablas que resumen los rendimientos de los diferentes clasificadores de ejemplo, de acuerdo con algunas realizaciones de la presente invención;

65 la FIG. 5A es una imagen de una tabla de un conjunto ampliado de características que se enumeran de acuerdo con su importancia en un clasificador de bosque aleatorio para hombres;

la FIG. 5B es una tabla que indica la correlación entre pares de resultados de ensayos de sangre;
 la FIG. 6 es una imagen de una tabla que muestra rendimientos para varias ventanas temporales.
 la FIG. 7 es una imagen de una tabla que muestra rendimientos de un clasificador de bosque aleatorio;
 la FIG. 8 es un diagrama de flujo de un método de uso de un clasificador o clasificadores para la estimación de
 5 una puntuación de riesgo gastrointestinal para un individuo objetivo, de acuerdo con algunas realizaciones de la
 presente invención;
 la FIG. 9 es una tabla que indica los rendimientos de los clasificadores para cada uno de cáncer de colon,
 estómago, recto y esófago en diferentes sensibilidades para grupos diferentes de poblaciones, de acuerdo con
 algunas realizaciones de la presente invención; y
 10 la FIG. 10 es un conjunto de tablas que resumen un análisis de los resultados del uso de los clasificadores
 descritos anteriormente para la clasificación de individuos anémicos y no anémicos (estadounidenses de raza
 blanca), de acuerdo con algunas realizaciones de la presente invención.

Descripción de realizaciones de la invención

15 La presente invención, en algunas realizaciones de la misma, se refiere al diagnóstico del cáncer y, más en
 particular, pero no exclusivamente, a métodos y sistemas de evaluación de un riesgo de cáncer.

20 De acuerdo con algunas realizaciones de la presente invención, se proporcionan métodos de evaluación del riesgo
 de cáncer gastrointestinal mediante la clasificación de un conjunto de resultados de ensayos de sangre actuales de
 un individuo objetivo usando uno o más clasificadores que se generan de acuerdo con un análisis de resultados de
 ensayos de sangre históricos de una pluralidad de individuos. El conjunto de resultados de ensayos de sangre
 actuales incluye al menos un resultado de los siguientes ensayos de sangre: hemoglobina (HGB), hematocrito (HCT)
 25 y glóbulos rojos sanguíneos (GRS) y al menos un resultado de los siguientes ensayos de sangre: hemoglobina
 corpuscular media (HCM) y concentración de hemoglobina corpuscular media (CHCM) y la edad del individuo
 objetivo. Opcionalmente, el conjunto de resultados de ensayos de sangre actuales incluye adicionalmente uno o más
 de los siguientes ensayos de sangre: recuento de glóbulos blancos sanguíneos - GBS (RSC); volumen de plaquetas
 medio (VPM); volumen celular medio (VCM); amplitud de distribución de glóbulos rojos (ADR); recuento de
 30 plaquetas (RSC); recuento de eosinófilos; porcentaje de neutrófilos; porcentaje de monocitos; porcentaje de
 eosinófilos; porcentaje de basófilos; recuento de neutrófilos; recuento de monocitos; y hematocrito de plaquetas
 (PCT).

Opcionalmente, el riesgo de cáncer gastrointestinal se evalúa mediante la clasificación de resultados de ensayos de
 35 sangre bioquímicos del individuo objetivo. En dichas realizaciones, los clasificadores se generan de acuerdo con un
 análisis de resultados de ensayos de sangre bioquímicos históricos de la pluralidad de individuos. Los resultados de
 ensayos de sangre bioquímicos pueden incluir resultados de cualquiera de los siguientes ensayos de sangre:
 albúmina, calcio, cloruro, colesterol, creatinina, lipoproteína de alta densidad (HDL), lipoproteína de baja densidad
 (LDL), potasio, sodio, triglicéridos, urea y/o ácido úrico.

Opcionalmente, el riesgo de cáncer gastrointestinal se evalúa mediante la clasificación de características
 40 demográficas del individuo objetivo. En dichas realizaciones, los clasificadores se generan de acuerdo con un
 análisis de características demográficas de la pluralidad de individuos.

Opcionalmente, se usan tanto los resultados de ensayos de sangre actuales de individuos objetivo como los
 45 resultados de ensayos de sangre históricos de individuos muestreados para la generación de conjuntos ampliados
 de características que incluyen valores manipulados y/o ponderados. Opcionalmente, cada conjunto ampliado de
 características se basa en las características demográficas de un individuo respectivo, por ejemplo, como se
 describe a continuación.

Opcionalmente, el uno o más clasificadores se adaptan a una o más características demográficas del individuo
 50 objetivo. Opcionalmente, los clasificadores se seleccionan para que coincidan con una o más características
 demográficas del individuo objetivo. En dichas realizaciones, pueden usarse clasificadores diferentes para las
 mujeres y los hombres.

De acuerdo con algunas realizaciones de la presente invención, se proporcionan métodos de generación de uno o
 55 más clasificadores para la evaluación del riesgo gastrointestinal de acuerdo con la reivindicación 15. Los métodos y
 sistemas se basan en el análisis de una pluralidad de resultados de ensayos de sangre históricos de cada uno de
 una pluralidad de individuos muestreados y en la generación en consecuencia de un conjunto de datos que tiene
 una pluralidad de conjuntos de características cada una generada de acuerdo con resultados respectivos de
 60 ensayos de sangre históricos en los que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos
 comprende resultados de al menos uno de los siguientes ensayos de sangre: glóbulos rojos sanguíneos (GRS),
 hemoglobina (HGB) y hematocrito (HCT) y al menos un resultado de los siguientes ensayos de sangre: hemoglobina
 (HCM) y concentración de hemoglobina corpuscular media (CHCM). Después, el conjunto de datos se usa para la
 65 generación y emisión de uno o más clasificadores, tales como clasificadores de vecinos más cercanos a K (VCK),
 clasificadores de bosque aleatorios y clasificadores de regresión lineal ponderados. Los clasificadores pueden
 proporcionarse como módulos para su ejecución en terminales clientes o se usan como un servicio en línea para la

evaluación del riesgo de cáncer gastrointestinal de individuos objetivo basándose en sus resultados de ensayos de sangre actuales.

5 Antes de explicar al menos una realización de la invención en detalle, ha de entenderse que la invención no se limita necesariamente en su aplicación a los detalles de construcción y la disposición de los componentes y/o métodos establecidos en la siguiente descripción y/o ilustrados en los dibujos y/o en los Ejemplos. La invención es susceptible de otras realizaciones o de ser puesta en práctica o realizada de diversas maneras.

10 Ahora se hace referencia a la FIG. 1, que es un diagrama de flujo de un método 100 de generación de uno o más clasificadores para la estimación de una puntuación de riesgo de cáncer gastrointestinal de acuerdo con un análisis de una pluralidad de resultados de ensayos históricos de cada uno de una pluralidad de individuos diagnosticados, de acuerdo con algunas realizaciones de la presente invención. Como se usa en el presente documento, un cáncer gastrointestinal puede ser cáncer de colon, de estómago, de recto o de esófago. También se hace referencia a la
 15 FIG. 2, que es una ilustración esquemática de un sistema 200 para la generación de un clasificador o clasificadores para la estimación de puntuaciones de riesgo de cáncer gastrointestinal, por ejemplo, mediante la implementación del método representado en la FIG. 1, de acuerdo con algunas realizaciones de la presente invención.

El sistema 200 incluye una o más bases de datos 201 de historiales clínicos y/o está conectado a una interfaz de bases de datos de historiales clínicos. La base o bases de datos 201 incluyen una pluralidad de registros de individuos, también denominada una pluralidad de muestras de individuos, que describen, para cada uno de una pluralidad de individuos muestreados, uno o más conjuntos de una pluralidad de conjuntos de resultados de ensayos históricos de cada individuo y, opcionalmente, uno o más parámetros demográficos y un pronóstico de cáncer gastrointestinal. El conjunto de una pluralidad de resultados de ensayos históricos, un parámetro o parámetros demográficos, tales como la edad, y/o un pronóstico de cáncer gastrointestinal, pueden almacenarse en un registro
 20 de muestras común y/o puede reunirse a partir de una serie de bases de datos independientes y/o conectados.

Opcionalmente, el pronóstico de cáncer gastrointestinal es una indicación binaria establecida de acuerdo con un archivo de registros de cáncer. Los diferentes resultados de ensayos pueden ser de ensayos de sangre realizados habitualmente y/o ensayos de sangre realizados durante el mismo periodo. Opcionalmente, algunos conjuntos de una pluralidad de resultados de ensayos históricos tienen resultados de ensayos de sangre faltantes. Estos resultados se completan opcionalmente mediante el promedio ponderado de los resultados de ensayos de sangre disponibles de otros individuos. El método incluye adicionalmente un procesador 204, un módulo 205 de generación de clasificadores y una unidad de interfaz 206, tal como una interfaz de red.
 30

35 Como se usa en el presente documento, un parámetro demográfico incluye la edad, el género, la raza, el peso, el origen nacional, la región geográfica de residencia y/o similares.

En primer lugar, como se muestra en 101, se proporciona uno o más conjuntos de datos de una pluralidad de muestras de individuos.
 40

Opcionalmente, como se muestra en 102, la pluralidad de muestras de individuos se criba y/o se selecciona de acuerdo con criterios de correspondencia. Por ejemplo, los archivos de muestras son de individuos de 40 años de edad o más, que o bien aparecen en un registro de cáncer con cáncer de colon, y, opcionalmente, sin otros tipos de cáncer, o no aparecen en el registro de cáncer. Opcionalmente, solo se toman archivos de muestras de individuos que aparecen en el registro de cáncer solo si el último conjunto de una pluralidad de resultados de ensayos históricos que documentan se tomó durante un periodo de tiempo determinado antes del registro de individuo respectivo en el registro de cáncer, por ejemplo, durante un periodo de tiempo de al menos 30 días antes de una fecha actual y un máximo de 2 años. Opcionalmente, solo se toman archivos de muestras de individuos que no aparecen en el registro de cáncer si incluyen un conjunto de una pluralidad de resultados de ensayos históricos que crea una distribución de tiempo igual (temporización de ensayos de sangre) para las poblaciones de cánceres gastrointestinales positivos y negativos. El proceso de equiparar la distribución de tiempo de las muestras positivas y negativas también conduce a omitir al menos algunas muestras negativas (no registradas) y a un cambio en la prevalencia gastrointestinal en el conjunto de datos.
 45
 50

55 Ahora, como se muestra en 103, un conjunto de datos de evaluación, tal como una matriz, se genera de acuerdo con los datos de muestras extraídos de los registros de muestras, por ejemplo, mediante el módulo 205 de generación de clasificadores. El conjunto de datos de evaluación incluye una pluralidad de conjuntos de características, opcionalmente ampliados. Cada conjunto de características se genera a partir de cada uno de los registros de muestras cribado y/o seleccionado. El conjunto de características, opcionalmente, está formado por
 60 características sin procesar que incluyen un ensayo de sangre real y/o valores característicos demográficos.

Como se ha descrito anteriormente, cada registro de ejemplo incluye uno o más conjuntos de una pluralidad de resultados de ensayos históricos de un individuo, cada uno incluye resultados de ensayos de sangre de combinación, por ejemplo, una combinación de más de 10, 15, 20 y/o cualquier número intermedio de resultados de ensayos de sangre. En un ejemplo, cada conjunto extraído de características sin procesar incluye al menos los siguientes 18 resultados de ensayos de sangre: glóbulos rojos sanguíneos (GRS); recuento de glóbulos blancos
 65

5 sanguíneos - GBS (RSC); volumen de plaquetas medio (VPM); hemoglobina (HGB); hematocrito (HCT); volumen celular medio (VCM); hemoglobina celular media (HCM); concentración de hemoglobina corpuscular media (CHCM); amplitud de distribución de glóbulos rojos (ADR); recuento de plaquetas (RSC); recuento de eosinófilos; porcentaje de neutrófilos; porcentaje de monocitos; porcentaje de eosinófilos; porcentaje de basófilos; recuento de neutrófilos; recuento de monocitos; y hematocrito de plaquetas (PCT). En otro ejemplo, cada conjunto extraído de características sin procesar incluye al menos resultados de los siguientes ensayos de sangre: HGB, HCT y GRS, al menos un resultado de los siguientes ensayos de sangre: HCM y CHCM y datos adicionales que reflejan la edad del individuo objetivo. Opcionalmente, este conjunto extraído de características sin procesar incluye adicionalmente uno o más de los siguientes ensayos de sangre: ADR, plaquetas y VCM. Adicionalmente, este conjunto extraído de características sin procesar además puede incluir uno o más de los siguientes ensayos de sangre: GBS, recuento de eosinófilos, porcentaje y/o recuento de neutrófilos, porcentaje y/o recuento de basófilos y porcentaje y/o recuento de monocitos.

15 Opcionalmente, el conjunto de características sin procesar está ampliado. El conjunto ampliado de características contiene características como los resultados de ensayos de sangre sin procesar anteriores y/o uno o más parámetros demográficos y resultados de ensayos de sangre opcionalmente manipulados y/o una combinación de resultados de ensayos de sangre, por ejemplo, como se describe a continuación. Por ejemplo, cada característica en el conjunto de características ampliadas se basa en un resultado de ensayo de sangre, una característica demográfica, una combinación de resultado o resultados de ensayos de sangre y/o característica o características demográficas, y/o una manipulación de resultado o resultados de ensayos de sangre y/o característica o características demográficas. Por ejemplo, cuando el conjunto de características sin procesar incluye 18 resultados de ensayos, se genera un conjunto ampliado de 114 características basándose en lo siguiente:

- 25 1. 18 características, cada una incluye otro de los 18 resultados ensayos de sangre.
- 30 2. 18 características, cada una incluye una diferencia (por ejemplo, una relación) entre uno de los 18 resultados de ensayos de sangre y un primer resultado virtual. El primer resultado virtual se calcula opcionalmente mediante el promedio ponderado de resultados disponibles respectivos de los registros de muestras. Opcionalmente, cada ensayo disponible se pondera de acuerdo con un periodo transcurrido desde la realización del mismo y la fecha objetivo, opcionalmente una fecha de un conjunto de una pluralidad de resultados de ensayos históricos de un individuo objetivo, denominada en el presente documento una fecha objetivo. Opcionalmente, los ensayos disponibles son ensayos tomados durante un primer periodo, por ejemplo, 540 días antes de la fecha objetivo. Por ejemplo, un peso puede calcularse como un valor absoluto derivado del tiempo transcurrido desde el registro (por ejemplo, cuando se tomó el ensayo) del mismo. El peso puede calcularse como una función cuadrática o cualquier otra función que sea monótona con respecto al valor absoluto.
- 35 3. 18 características, cada una incluye una diferencia (por ejemplo, una relación) entre uno de los 18 resultados de ensayos de sangre y un segundo resultado virtual, que se calcula opcionalmente como el primer resultado virtual descrito anteriormente, basada en ensayos disponibles tomados durante un segundo periodo, por ejemplo, durante 1080 días antes de la fecha objetivo.
- 40 4. 1 característica - el número de conjuntos de una pluralidad de resultados de ensayos históricos que el usuario realizó durante un periodo de un año antes de la fecha objetivo.
- 45 5. 1 característica - el número de conjuntos de una pluralidad de resultados de ensayos históricos que el usuario realizó durante un periodo de entre 180×6 y 180×10 días antes de la fecha objetivo.
6. 1 característica - la edad del individuo, por ejemplo, el año de nacimiento del individuo.
7. 57 características que son valores cuadrados de todas las características anteriores (detalladas en los puntos 1-6).

50 Opcionalmente, pueden documentarse uno o más resultados de ensayos de sangre bioquímicos por individuo y pueden añadirse opcionalmente como característica al conjunto de características. Estas características pueden tratarse como los resultados de ensayos de sangre anteriores. Los resultados de ensayos de sangre bioquímicos pueden incluir cualquiera de los siguientes resultados de ensayos de sangre: albúmina, calcio, cloruro, colesterol, creatinina, lipoproteína de alta densidad (HDL), lipoproteína de baja densidad (LDL), potasio, sodio, triglicéridos, urea y/o ácido úrico.

55 Opcionalmente, se generan diferentes conjuntos de datos de evaluación, por ejemplo, matrices, que tienen diferentes conjuntos de características ampliadas para crear diferentes clasificadores que clasifican individuos objetivo que tienen diferente característica o características demográficas, por ejemplo, género.

60 Opcionalmente, el conjunto de datos de evaluación, por ejemplo, la matriz, se filtra, para eliminar valores atípicos iterativamente. Opcionalmente, se calcula una desviación promedio y/o una desviación típica de cada característica y características que tienen valores excepcionales, por ejemplo, más de un umbral máximo de desviación típica, por ejemplo 10, se truncan hasta el umbral máximo de desviación típica. Por ejemplo, el proceso se repite iterativamente 10 veces (o menos, si no se realiza ningún truncamiento). Un pseudocódigo ejemplar que describe el proceso de limpieza es como se indica a continuación:

Repetir 10 veces
 Para cada característica –
 Calcular promedio y desviación típica (dt)
 Para cada muestra –
 5 Si $(\text{Valor}\{\text{muestra, característica}\} > \text{media}\{\text{característica}\} + 10 \cdot \text{dt}\{\text{característica}\})$
 $\text{Valor}\{\text{muestra, característica}\} = \text{media}\{\text{característica}\} + 9 \cdot \text{dt}\{\text{característica}\}$
 10 Si $(\text{Valor}\{\text{muestra, característica}\} < \text{media}\{\text{característica}\} - 10 \cdot \text{dt}\{\text{característica}\})$
 $\text{Valor}\{\text{muestra, característica}\} = \text{media}\{\text{característica}\} - 9 \cdot \text{dt}\{\text{característica}\}$
 Finalizar bucle de muestras
 Finalizar bucle de características
 15 Salir si no se realizó ningún cambio
 Finalizar repeticiones

Ahora, como se muestra en 104, el conjunto de datos de evaluación se usa para la generación de un clasificador o clasificadores clasificando cada uno un riesgo de cáncer gastrointestinal de un individuo objetivo basándose en una o más características demográficas del mismo y un conjunto actual de una pluralidad de resultados de ensayos, por ejemplo, mediante el módulo 205 de generación de clasificadores. Opcionalmente, puede generarse uno o más de los siguientes clasificadores basándose en el conjunto de datos de evaluación:

un clasificador de regresión lineal ponderado donde los registros de muestras positivos reciben una puntuación que es aproximadamente 100 veces la puntuación de los registros de muestras negativos;
 un clasificador de vecinos más cercanos a K (VCK), por ejemplo 100 veces el muestreo descendente de un registro de muestras negativo; y un clasificador de bosque aleatorio, por ejemplo, donde cada árbol se construye usando una relación 2:1 de registros de muestras de negativas a positivas.

Opcionalmente, el rendimiento de cada uno de los clasificadores se estima usando un proceso de validación cruzada de 10 veces en el que el conjunto de datos de evaluación, denominado en el presente documento una población, se divide aleatoriamente en diez partes de igual tamaño. Para cada parte, puede realizarse lo siguiente:

seleccionar conjuntos aceptables de resultados de ensayos de sangre del 90 % de la población no en la parte respectiva;
 entrenar un clasificador de acuerdo con los conjuntos seleccionados de resultados de ensayos de sangre;
 seleccionar conjuntos de resultados de ensayos de sangre de un 10 % de la población en la parte respectiva; y
 usar el clasificador en los conjuntos seleccionados de resultados de ensayos de sangre del 10 % de la población.

Ahora, se recogen clasificaciones, también denominadas predicciones, para medir el rendimiento de cada clasificador. Por ejemplo, las mediciones de rendimiento se seleccionan de acuerdo con una curva de una característica operativa receptora (COR), por ejemplo, como se representa en la FIG. 3. Opcionalmente, se usa la especificidad en diferentes valores de sensibilidad (exhaustividad) (5 %, 10 %, 20 %, 50 % y 70 %) para la identificación de las mediciones. Los rendimientos de los diferentes clasificadores de ejemplo se resumen en las tablas proporcionadas en las FIG. 4A-4C, que tienen, respectivamente, un área bajo la curva (ABC) de $0,840 \pm 0,001$, $0,820 \pm 0,001$ y $0,833 \pm 0,001$. Cabe señalar que la FIG. 3 y las FIG. 4A-4C se calculan basándose en una población total de 217.246 hombres de más de 40 años, se han identificado 1.415 como que tienen cáncer colorrectal positivo (CCR). Opcionalmente, cada uno de los números de la tabla (Mejora de la confianza, Precisión Est. y Especificidad) representa la media \pm una desviación típica (dt) calculada sobre diferentes esquemas de validación cruzada, por ejemplo 10.

Como se usa en el presente documento, una exhaustividad indica una tasa de positivos verdaderos (PV) de individuos clasificados como PV que equivale a un porcentaje de individuos con CCR (tanto PV como negativos falsos (NF)), por ejemplo, $\text{PV}/(\text{PV} + \text{NF})$. Como se usa en el presente documento, la precisión indica un porcentaje de población que se ha clasificado como que tiene CCR, por ejemplo, $\text{PV}/(\text{PV} + \text{PF})$. Como se usa en el presente documento, una mejora de la confianza indica una relación de precisión con respecto a la prevalencia global de CCR en la población. Por ejemplo, entre una población global de 217.246 hombres de más de 40 años, se han identificado 1.415 como que tienen CCR positivo. En este ejemplo, la prevalencia es del 0,65 %. La selección de muestras negativas para crear la distribución de tiempo correcta (véase anteriormente), crea un sesgo en las poblaciones de aprendizaje y de ensayo, conduciendo a una prevalencia de CCR del 1,2 %. Por tanto, la mejora de la confianza puede encontrarse directamente, pero solo puede usarse indirectamente para llegar a la precisión estimada mediante el ajuste de una precisión medida.

Opcionalmente, el conjunto de características que son utilizadas por un determinado clasificador para la clasificación se pondera de acuerdo con una importancia de clasificación. La importancia de una característica puede determinarse como un valor promedio, sobre árboles de datos, de una disminución de impurezas de nodo según se

mide por el coeficiente de Gini (dispersión estadística) debido a divisiones. Como ejemplo, la FIG. 5A representa una tabla de un conjunto ampliado de características que se enumeran de acuerdo con su importancia en un clasificador de bosque aleatorio para hombres.

5 Cabe señalar que los resultados de ensayos de sangre históricos de los clasificadores y el ensayo de sangre actual pueden incluir pares de resultados de ensayos de sangre de ensayos de sangre considerados de naturaleza similar. Estos pares incluyen uno o más de los siguientes pares: hemoglobina (HGB) y hematocrito (HCT), porcentaje/recuento de neutrófilos y porcentaje/recuento de linfocitos (es decir, extraído de un ensayo de RSC), VCM y HCM, GRS y hematocrito (HCT), GRS y VCM. La comprobación de ambos resultados de sangre de un par de este tipo no es trivial ya que para el experto en la materia estos resultados de sangre tienen indicaciones comunes y ningún valor acumulativo y, por tanto, el experto en la materia usaría solamente los resultados de uno de los miembros del par de un ensayo de sangre similar y no ambos miembros del par de un ensayo de sangre similar. Los inventores sorprendentemente descubrieron que la correlación entre miembros de un par de este tipo no es absoluta y que la contribución al rendimiento del clasificador o clasificadores descritos anteriormente es sustancial. Por ejemplo, véase la tabla en la FIG. 5B.

20 Debe tenerse en cuenta que el rendimiento de un clasificador depende del tiempo entre la fecha del último conjunto de resultados de ensayos de sangre y la fecha del descubrimiento del cáncer, por ejemplo, según se registra en el registro de cáncer. Esta dependencia se captura considerando clasificadores en los que los ensayos de sangre aceptables para muestras positivas están dentro de ventanas temporales limitadas (por ejemplo, 30-90 días antes del registro del cáncer, 90-180 días y/o similares). Por ejemplo, la FIG. 6 representa una tabla que muestra rendimientos de varias ventanas temporales. La tabla muestra datos de RSC ajustados a la edad tomados 30-720 días antes del diagnóstico de casos de CCR comparados con datos de control sanos que indican variaciones a largo plazo. Además, para cada caso, el valor/resultado del parámetro seleccionado se comparó con los resultados del mismo parámetro 1,5 años (delta 1,5) y 3 años (delta 3) antes de la evaluación valor/resultado del parámetro seleccionado. Como se indica mediante la tabla de la FIG. 6, la especificidad se reduce cuando los datos son más antiguos.

30 Cabe señalar que la anemia en la sangre, puede estar provocada por varias afecciones gastrointestinales y otras, siendo el cáncer GI el menos habitual. La anemia inexplicada es un factor predictivo de máxima importancia para el CCR en ancianos y, junto con las hemorroides, es una causa habitual de retraso en el diagnóstico del CCR. La pérdida de sangre está presente en el 60 % de los casos de CCR y una pérdida diaria de tan poco como 3 ml en las heces puede provocar anemia ferropénica. Aproximadamente el 18 % de los casos de CCR tenía anemia más de un año antes del diagnóstico 14, sin embargo, una proporción significativa no eran anémicos 1. Puede estar presente un ensayo de sangre oculta positivo. Sin embargo, la sangre en heces, actualmente utilizada para la detección de CCR, solamente detecta el sangrado actual, mientras que en el CCR, la pérdida de sangre es crónica.

40 En Spell DW, Jones DV, Jr., Harper WF, David Bessman J. *The value of a complete blood count in predicting cancer of the colon. Cancer Detect Prev* 2004; 28 (1): 37-42 se informa que el 88 % de los pacientes con CCR tenía al menos una anomalía en la sangre. De este modo, los intentos de predecir el CCR a partir de recuentos sanguíneos completos (RSC) se encuentran en investigación activa. En un estudio retrospectivo en pacientes con CCR recién diagnosticado de los que había parámetros de RSC disponibles 0-122 días antes del diagnóstico, se demostró que la amplitud de distribución de glóbulos rojos (ADR) estaba aumentada por encima del intervalo normal y tenía una sensibilidad del 84 % y una especificidad del 88 %, principalmente para los casos de CCR del lado derecho. No se documentó ninguna sensibilidad mejorada en combinación con ADR, hemoglobina y volumen corpuscular medio (VCM).

50 De acuerdo con algunas realizaciones de la presente invención, el rendimiento de un clasificador generado como se describe en el presente documento puede usarse para clasificar tanto los individuos con afección anémica como los individuos sin afección anémica. Por ejemplo, la FIG. 11 es un conjunto de tablas que resume un análisis de los resultados del uso de los clasificadores descritos anteriormente para la clasificación de individuos anémicos y no anémicos (estadounidenses de raza blanca). El conjunto incluye una pluralidad de tablas, resumiendo cada una la probabilidad de un individuo anémico de un grupo determinado de individuos con respecto a individuos no anémicos de otro grupo de individuos. Los grupos se dividen opcionalmente basándose en una combinación entre la antigüedad de los resultados de sangre y la edad del individuo. La FIG. 10 muestra pruebas de que mediciones independientes de parámetros de recuentos sanguíneos se relacionan con el CCR y de que pueden usarse cambios combinados en parámetros de RSC, incluso sutiles, dentro del intervalo normal, como parte del proceso de detección del CCR, por ejemplo, mediante puntuación, para un individuo con o sin anemia.

60 Ahora, como se muestra en 105, el clasificador o clasificadores se emiten, opcionalmente, en forma de un módulo que permite la clasificación de individuos objetivo, por ejemplo, mediante la unidad de interfaz 206. Opcionalmente, se definen diferentes clasificadores para individuos que tienen diferentes características demográficas, por ejemplo, un clasificador para hombres y otro para mujeres. Por ejemplo, aunque se establece un clasificador que se basa en un grupo de características de las 18 características anteriores para hombres, se usa un clasificador de Bosque Aleatorio sin ensayos bioquímicos para mujeres. El clasificador de Bosque Aleatorio tiene un ABC de $0,833 \pm 0,001$

y rendimientos como se representan en la FIG. 7 donde la precisión se estima de acuerdo con una prevalencia total del 0,45 %.

5 Se hace referencia ahora a la FIG. 8, que es un diagrama de flujo 400 de un método de uso de un clasificador o clasificadores, tales como el clasificador o clasificadores anteriores, para estimar una puntuación de riesgo de cáncer gastrointestinal para un individuo objetivo, de acuerdo con algunas realizaciones de la presente invención. En su uso, el clasificador o clasificadores pueden alojarse en un servidor web que recibe los datos del individuo objeto y evalúa, usando un módulo de evaluación de cáncer gastrointestinal que usa el clasificador o clasificadores, una puntuación de riesgo de cáncer gastrointestinal en un sujeto que se ha de evaluar. Los datos del individuo objetivo pueden recibirse a través de una red de comunicación, tal como Internet, desde un terminal cliente, tal como un ordenador portátil, un ordenador de sobremesa, un teléfono inteligente, una tableta y/o similar, que proporciona el conjunto de resultados de ensayos de sangre y características demográficas del sujeto o una referencia a estos datos objetivo.

15 En primer lugar, como se muestra en 401 y 402, se proporciona un clasificador o clasificadores y datos de un individuo objetivo. Los datos del individuo objetivo incluyen uno o más parámetros demográficos y un conjunto de una pluralidad de resultados de ensayos de sangre actuales realizados en la fecha objetivo, que incluye una serie de resultados de ensayos actuales de un individuo objetivo. Los datos del individuo objetivo pueden introducirse manualmente por un usuario, por ejemplo, utilizando una interfaz gráfica de usuario (IGU), seleccionada por un usuario, usando opcionalmente la IGU, y/o pueden proporcionarse automáticamente, por ejemplo, mediante un módulo y/o sistema de diagnóstico asistido por ordenador (DAO). Opcionalmente, los datos del individuo objetivo incluyen el número de conjuntos de resultados de ensayos de sangre del usuario realizados durante el último año, la última década y/o cualquier periodo intermedio. Cada uno de los conjuntos de resultados de ensayos de sangre incluye resultados de ensayos de sangre, por ejemplo, un grupo seleccionado entre los 18 resultados de ensayos de sangre diferentes anteriores.

Ahora, como se muestra en 403, se extrae un conjunto de características del individuo objetivo de los datos del individuo objetivo y, opcionalmente, se amplían como se ha descrito anteriormente.

30 Después, como se muestra en 404, el clasificador o clasificadores se usan para calcular una puntuación de riesgo de cáncer gastrointestinal para el individuo objetivo mediante la ponderación de cada característica en el conjunto de características del individuo objetivo. Ahora, como se muestra en 405, se emite la puntuación de riesgo de cáncer gastrointestinal.

35 Cabe señalar que los clasificadores descritos anteriormente pueden usarse para la estimación de una puntuación de riesgo de cáncer para un cáncer gastrointestinal que puede ser cáncer de colon, estómago, recto o esófago. Por ejemplo, la FIG. 9 es una tabla que indica los rendimientos de los clasificadores para cada uno de los cánceres de colon, estómago, recto y esófago en diferentes sensibilidades para diferentes grupos de poblaciones. Se definen grupos de poblaciones de acuerdo con una combinación entre una antigüedad de los respectivos resultados de ensayos de sangre (expresada en días, por ejemplo, 90-540, 90-540, 30-270 y 360-720 días) y el intervalo de edades, por ejemplo, 40-100 y 50-75. Cabe señalar que la tabla de la FIG. 9 indica los rendimientos de los clasificadores en una población diferente de la utilizada para los clasificadores documentados con referencia a las FIG. 4A y 4B. En la FIG. 9 los datos incluyen los resultados de ensayos de sangre de un total de 81.641 individuos ingleses de más de 40 años de edad de los cuales a 3.099 se les diagnosticó cáncer de colon, a 1.286 cáncer de recto, a 578 cáncer gástrico y a 1061 cáncer de esófago.

Se espera que durante la vida de una patente madurada a partir de la presente solicitud se desarrollaran muchos sistemas y métodos relevantes y el alcance del término un procesador, una pantalla e interfaz de usuario tiene por objeto incluir todas estas nuevas tecnologías a priori.

50 Como se usa en el presente documento el término "aproximadamente" se refiere a \pm el 10 %.

Las expresiones "comprende", "que comprende", "incluye", "que incluye", "que tiene" y sus conjugados significan "incluyendo pero no limitado a". Esta expresión abarca las expresiones "que consiste en" y "que consiste esencialmente en".

La frase "que consiste esencialmente en" significa que la composición o método puede incluir ingredientes y/o etapas adicionales, pero solo si los ingredientes y/o etapas adicionales no alteran materialmente las características básicas y novedosas de la composición o el método reivindicados.

60 Como se usa en el presente documento, las formas singulares "un", "una", "el" y "la" incluyen referencias plurales a menos que el contexto dicte claramente lo contrario. Por ejemplo, la expresión "un compuesto" o "al menos un compuesto" puede incluir una pluralidad de compuestos, incluyendo mezclas de los mismos.

La expresión "de ejemplo" se usa en el presente documento con el significado de "que sirve como ejemplo, caso o ilustración". Ninguna realización descrita como "de ejemplo" ha de interpretarse necesariamente como preferida o ventajosa sobre otras realizaciones ni ha de excluirse la incorporación de características de otras realizaciones.

5 La palabra "opcionalmente" se usa en el presente documento con el significado de "se proporciona en algunas realizaciones y no se proporciona en otras realizaciones". Cualquier realización particular de la invención puede incluir una pluralidad de características "opcionales" a menos que dichas características entren en conflicto.

10 En toda la presente solicitud, diversas realizaciones de la presente invención pueden presentarse en un formato de intervalo. Ha de entenderse que la descripción en formato de intervalo es meramente por conveniencia y brevedad y no ha de interpretarse como una limitación inflexible sobre el alcance de la invención. En consecuencia, ha de considerarse que la descripción de un intervalo tiene todos los subintervalos posibles desvelados específicamente, así como valores numéricos individuales dentro de ese intervalo. Por ejemplo, ha de considerarse que la descripción
15 de un intervalo tal como de 1 a 6 tiene subintervalos desvelados específicamente tales como de 1 a 3, de 1 a 4, de 1 a 5, de 2 a 4, de 2 a 6, de 3 a 6 etc., así como números individuales dentro de ese intervalo, por ejemplo, 1, 2, 3, 4, 5 y 6. Esto se aplica independientemente de la amplitud del intervalo.

20 Siempre que se indique un intervalo numérico en el presente documento, se entiende que incluye cualquier numeral (fraccionario o integral) citado dentro del intervalo indicado. Las frases "que varía/varía entre" un primer número indicado y un segundo número indicado y "que varía/varía de" un primer número indicado "a" un segundo número indicado el número se usan indistintamente en el presente documento y tienen por objeto incluir los números indicados primero y segundo y todos los números fraccionarios e integrales entre los mismos.

25 Se aprecia que ciertas características de la invención, que, para mayor claridad, se describen en el contexto de realizaciones separadas, también pueden proporcionarse en combinación en una única realización. Por el contrario, diversas características de la invención, que, por brevedad, se describen en el contexto de una única realización, también pueden proporcionarse por separado o en cualquier subcombinación adecuada o como sea adecuado en cualquier otra realización descrita de la invención. Ciertas características descritas en el contexto de diversas realizaciones no han de considerarse características esenciales de esas realizaciones, a menos que la realización
30 sea inoperante sin esos elementos.

REIVINDICACIONES

1. Un método informatizado de evaluación del riesgo de cáncer gastrointestinal, que comprende:

- 5 generar un conjunto de características que comprende una pluralidad de resultados de ensayos de sangre actuales de una sangre recogida de un individuo objetivo;
- 10 seleccionar al menos un clasificador de acuerdo con al menos una característica demográfica de dicho individuo objetivo entre una pluralidad de clasificadores cada uno generado de acuerdo con una pluralidad de resultados de ensayos de sangre históricos respectivos de una pluralidad de individuos muestreados que tienen al menos una característica demográfica diferente; dicho al menos un clasificador se genera de acuerdo con un análisis de dicha pluralidad de resultados de ensayos de sangre históricos respectivos de cada uno de dicha pluralidad de individuos muestreados; y
- 15 evaluar, usando un procesador, un riesgo de cáncer gastrointestinal de dicho individuo objetivo mediante la clasificación de dicho conjunto de características usando dicho al menos un clasificador; y
- 20 en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de al menos uno de los siguientes ensayos de sangre: glóbulos rojos sanguíneos (GRS), hemoglobina (HGB) y hematocrito (HCT), y al menos un resultado de los siguientes ensayos de sangre: hemoglobina (HCM) y concentración de hemoglobina corpuscular media (CHCM).
- 25 2. El método de la reivindicación 1, en el que dicho riesgo de cáncer gastrointestinal es para un cáncer seleccionado entre un grupo que consiste en cáncer de colon, cáncer de estómago, cáncer de recto y cáncer de esófago.
- 30 3. El método de la reivindicación 1, en el que dicho conjunto de características comprende una edad de dicho individuo objetivo; en el que dicho al menos un clasificador se genera de acuerdo con un análisis de la edad de cada uno de una pluralidad de individuos muestreados.
- 35 4. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de amplitud de distribución de glóbulos rojos (ADR).
- 40 5. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de hematocrito.
- 45 6. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de volumen celular medio (VCM).
- 50 7. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados tanto de HGB como de HCT.
- 55 8. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados tanto de porcentaje/recuento de neutrófilos y porcentaje/recuento de linfocitos.
- 60 9. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende al menos uno de los resultados tanto de VCM como de HCM, de los resultados tanto de GRS como de HCT y de los resultados tanto de GRS como de VCM.
- 65 10. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende al menos uno de los siguientes ensayos de sangre: recuento de glóbulos blancos sanguíneos - GBS (RSC); volumen de plaquetas medio (VPM); volumen celular medio; recuento de plaquetas (RSC); recuento de eosinófilos; porcentaje de neutrófilos; porcentaje de monocitos; porcentaje de eosinófilos; porcentaje de basófilos; y recuento de neutrófilos; recuento de monocitos.
- 70 11. El método de la reivindicación 1, en el que dicho al menos un clasificador comprende un miembro de un grupo que consiste en: un clasificador ponderado de regresión lineal, un clasificador de vecinos más cercanos a K (VCK) y un clasificador de bosque aleatorio.
- 75 12. El método de la reivindicación 1, en el que dicho conjunto de características comprende al menos una característica demográfica de dicho individuo objetivo y dicho al menos un clasificador generado de acuerdo con un análisis de dicha respectiva al menos una característica demográfica de cada uno de dicha pluralidad de individuos muestreados.
- 80 13. El método de la reivindicación 1, en el que dicha pluralidad de resultados de ensayos de sangre comprende al menos un resultado de la siguiente pluralidad de ensayos de sangre: resultados de ensayos de sangre bioquímicos que pueden incluir cualquiera de los siguientes resultados de ensayos de sangre: albúmina, calcio, cloruro, colesterol, creatinina, lipoproteína de alta densidad (HDL), lipoproteína de baja densidad (LDL), potasio, sodio, triglicéridos, urea y/o ácido úrico.

14. El método de la reivindicación 1, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos y actuales comprende resultados de al menos uno de amplitud de distribución de glóbulos rojos (ADR), hematocrito y volumen celular medio (VCM).

5 15. Un método de generación de un clasificador para una evaluación de riesgo de cáncer gastrointestinal, que comprende:

proporcionar una pluralidad de resultados de ensayos de sangre históricos de cada uno de una pluralidad de individuos muestreados;

10 generar un conjunto de datos que tiene una pluralidad de conjuntos de características, cada conjunto generado de acuerdo con una pluralidad respectiva de resultados de ensayos de sangre históricos de cada uno de dicha pluralidad de individuos muestreados;

añadir al menos un parámetro demográfico de cada uno de dicha pluralidad de individuos muestreados a dicho conjunto respectivo de características;

15 generar al menos un clasificador de acuerdo con un análisis de dicho conjunto de datos;

emitir dicho al menos un clasificador en el que dicha generación comprende calcular y añadir al menos una versión manipulada de dicho al menos un parámetro demográfico como una característica a dicho conjunto respectivo de características, en el que cada uno de dicha pluralidad de resultados de ensayos de sangre históricos comprende resultados de al menos uno de los siguientes ensayos de sangre: glóbulos rojos sanguíneos (GRS), hemoglobina (HGB) y hematocrito (HCT), y al menos un resultado de los siguientes ensayos de sangre: hemoglobina (HCM) y concentración de hemoglobina corpuscular media (CHCM).

20

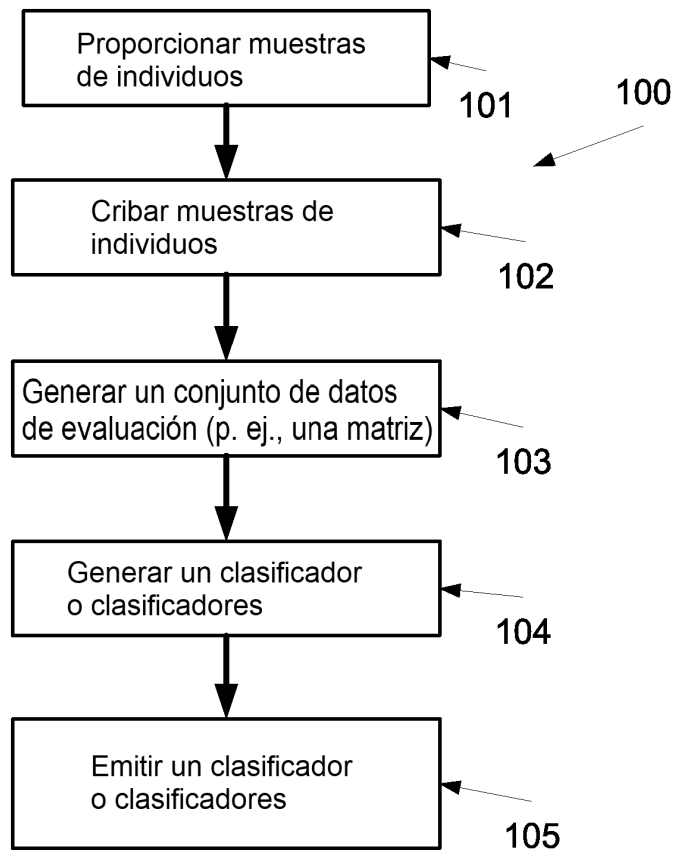


FIG. 1

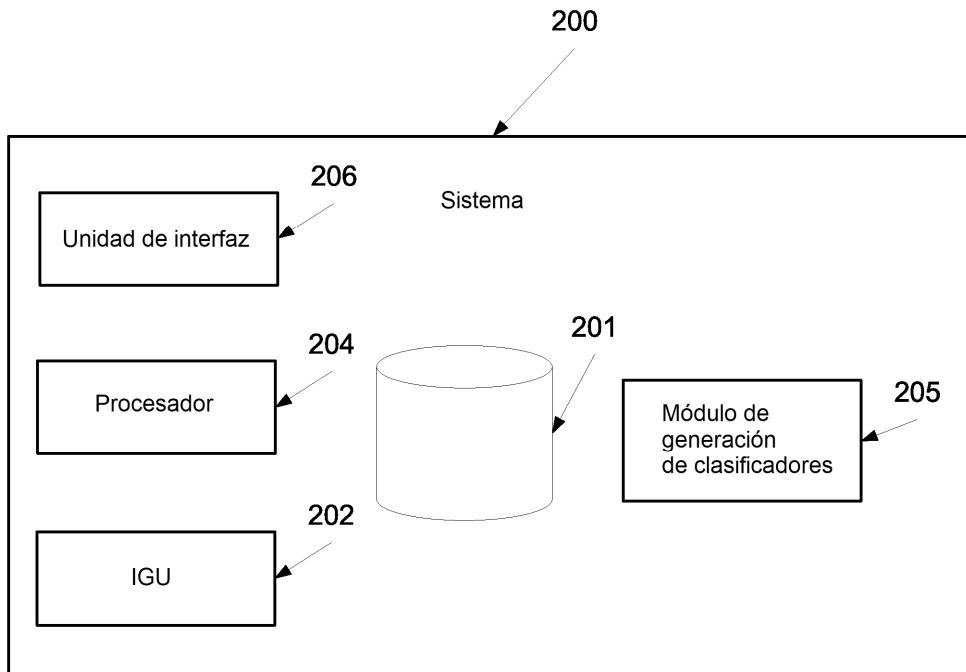


FIG. 2

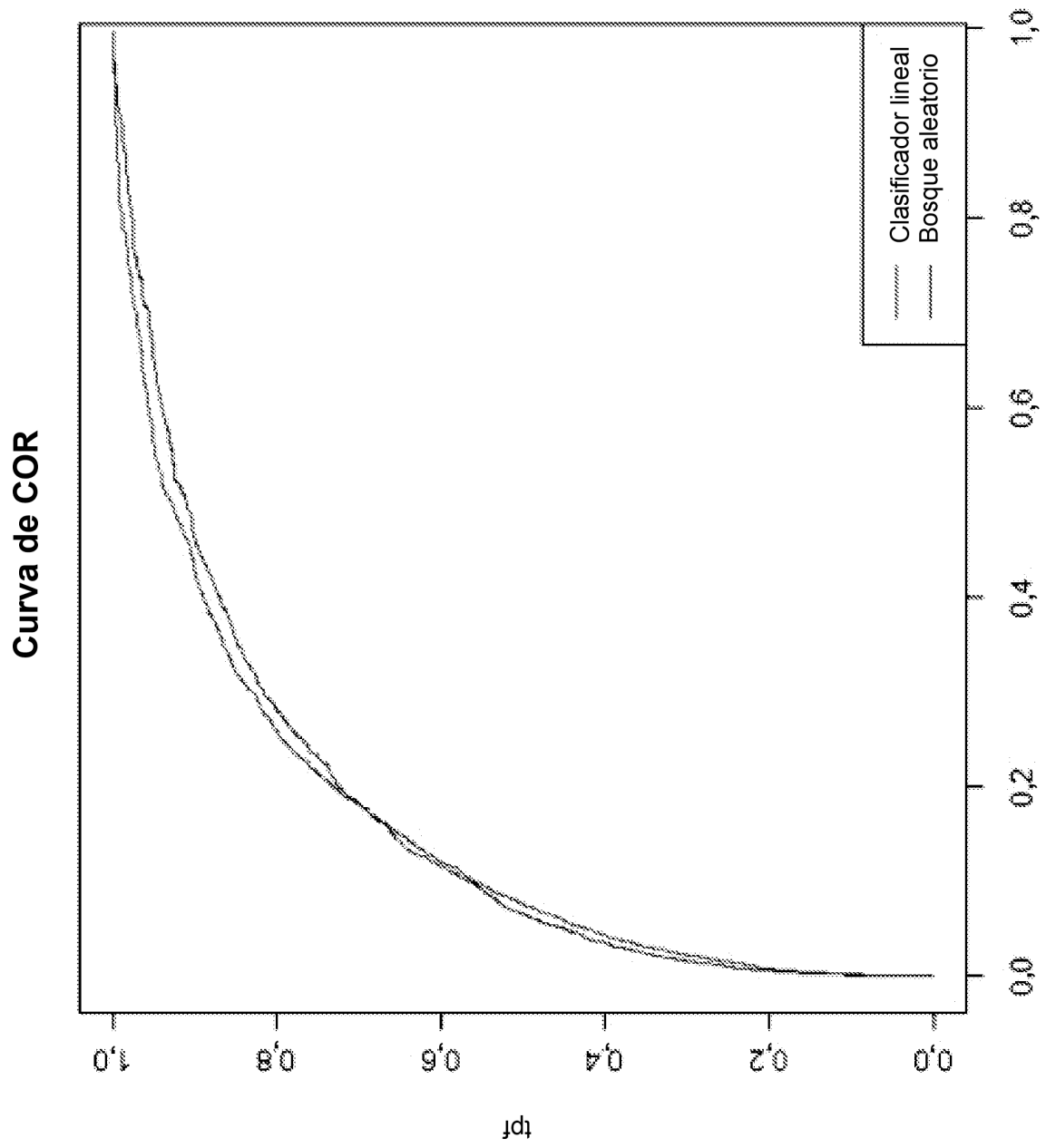


FIG. 3 tpv

Exhaustividad (%)	Mejora de la confianza	Precisión est. (%)	Especificidad (%)
5	38 ± 1	24,9 ± 0,7	99,930 ± 0,004
10	31 ± 1	20,1 ± 0,7	99,79 ± 0,01
20	19,0 ± 0,5	12,2 ± 0,3	99,16 ± 0,03
50	6,0 ± 0,1	3,90 ± 0,05	92,9 ± 0,1
70	3,70 ± 0,05	2,40 ± 0,03	81,6 ± 0,2

FIG. 4A

Exhaustividad (%)	Mejora de la confianza	Precisión est. (%)	Especificidad (%)
5	43 ± 2	27,5 ± 1,5	99,940 ± 0,007
10	33 ± 2	21,5 ± 1,0	99,81 ± 0,02
20	19,6 ± 0,5	12,70 ± 0,15	99,21 ± 0,01
50	5,5 ± 0,2	3,60 ± 0,15	91,3 ± 0,4
70	3,27 ± 0,08	2,10 ± 0,05	79,2 ± 0,5

FIG. 4B

Exhaustividad (%)	Mejora de la confianza	Precisión est. (%)	Especificidad (%)
5	50,5 ± 1,5	32,5 ± 1	99,960 ± 0,003
10	41,5 ± 1,5	27 ± 1	99,87 ± 0,01
20	23,6 ± 0,5	15,4 ± 0,3	99,38 ± 0,02
50	7,0 ± 0,2	4,5 ± 0,1	93,3 ± 0,2
70	3,66 ± 0,06	2,40 ± 0,05	81,5 ± 0,3

FIG. 4C

Característica	Importancia	N.º de Monocitos historia de 1,5 años	17,35154
Edad	206,06278	N.º de Eosinófilos historia de 1,5 años	17,14117
Hemoglobina actual	97,94094	N.º de Basófilos actual	14,47268
Hematocrito actual	75,68337	Recuento de RSC de 1000 días	13,73909
CHCM-M actual	65,28961	N.º de Basófilos historia de 1,5 años	12,74062
HCM historia de 3 años	61,24915	N.º de Basófilos historia de 3 años	12,58885
ADE actual	59,87631		
HCM historia de 1,5 años	59,41473		
Hemoglobina historia de 3 años	47,9564		
GR S actual	42,17517		
HCM actual	41,98025		
Hemoglobina historia de 1,5 años	41,33024		
Plaquetas actuales	39,70198		
Hematocrito historia de 3 años	39,57657		
% de Neutrófilos actual	34,09554		
VCM historia de 1,5 años	33,75738		
Hematocrito historia de 1,5 años	32,66408		
VCM historia de 3 años	32,25354		
ADE historia de 1,5 años	31,58052		
N.º de Neutrófilos actual	31,49156		
ADR historia de 3 años	31,433 %		
% de Eosinófilos actual	29,21783		
N.º de Monocitos actual	29,07631		
VPM actual	29,039		
% de Monocitos actual	28,70935		
CMB actual	27,77065		
VCM actual	27,16896		
N.º de Eosinófilos actual	26,03917		
Plaquetas historia de 3 años	25,98912		
Plaquetas historia de 1,5 años	24,74618		
CHCM-M historia de 3 años	23,69543		
CHCM-M historia de 1,5 años	22,48561		
VPM historia de 3 años	21,91487		
GR S historia de 3 años	21,77685		
Recuento de RSC de 500 días	21,66158		
GR S historia de 1,5 años	20,76823		
VPM historia de 1,5 años	20,40031		
% de Neutrófilos historia de 1,5 años	20,29873		
% de Neutrófilos historia de 3 años	19,53775		
% de Eosinófilos historia de 3 años	19,23888		
% de Monocitos historia de 3 años	19,03372		
% de Eosinófilos historia de 1,5 años	18,81205		
% de Monocitos historia de 1,5 años	18,41941		
N.º de neutrófilos historia de 1,5 años	18,38092		
% de Basófilos historia de 3 años	18,17924		
N.º de neutrófilos historia de 3 años	18,08296		
% de Basófilos historia de 1,5 años	17,88699		
GBS historia de 3 años	17,87557		
% de Basófilos actual	17,7247		
N.º de Monocitos historia de 3 años	17,72236		
N.º de Eosinófilos historia de 3 años	17,70108		
Recuento de RSC de 1500 días	17,62082		
CMB historia de 1,5 años	17,4697		

FIG. 5A

característica1 ▼	característica2 ▼	Correlación ▼	abc caract. 1 ▼	abc caract. 2 ▼	auc ambas ▼
[Hemoglobina]	[Hematocrito]	0,938058	0,7416	0,719	0,7411
[Neutrófilos]	[Linfocitos]	-0,930858	0,5424	0,5737	0,59
[VCM]	[HCM]	0,873668	0,5915	0,6453	0,689
[GRS]	[Hematocrito]	0,774977	0,6582	0,719	0,7249
[GRS]	[Hemoglobina]	0,724278	0,6582	0,7416	0,7461
[GRS]	[VCM]	-0,534006	0,6582	0,5915	0,725

FIG. 5B

Periodo (Días)	Sens. del 5 %		Sens. del 10 %		Sens. del 20 %		Sens. del 50 %		Sens. del 70 %	
	Mejora de la confianza	Espec.	Mejora de la confianza	Espec.	Mejora de la confianza	Espec.	Mejora de la confianza	Espec.	Mejora de la confianza	Espec.
30-90	99	99,98	87	99,95	48	99,71	12	96,23	5,6	87,90
90-180	65	99,95	53	99,86	32	98,97	8,3	93,77	4,6	84,87
180-360	26	99,84	23	99,64	14	98,68	5,6	91,27	3,5	80,47
360-540	18	99,75	13	99,23	8	97,62	4,5	89,06	3,2	78,31
540-720	10	99,54	8,6	98,89	6,1	96,76	3,9	87,28	3,1	77,31

FIG. 6

Exhaustividad (%)	Mejora de la confianza	Precisión est. (%)	Especificidad (%)
5	58,0 ± 3,5	27 ± 2	99,950 ± 0,004
10	44,5 ± 1,5	20,5 ± 0,8	99,850 ± 0,008
20	28,3 ± 0,7	13,0 ± 0,3	99,44 ± 0,02
50	7,6 ± 0,2	3,5 ± 0,1	93,8 ± 0,2
70	3,52 ± 0,06	1,6 ± 0,03	80,5 ± 0,4

FIG. 7

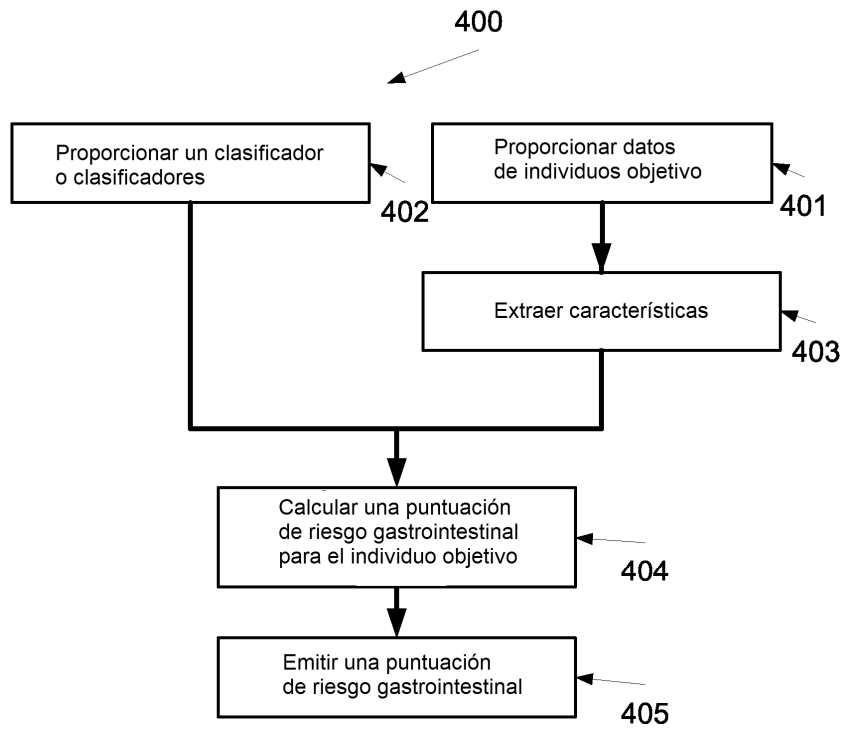


FIG. 8

Resultados finos		
Sens al 90%	ABC	Colon
55,80%	0,85	90-540/40-100:
48,00%	0,79	90-540/50-75:
67,90%	0,89	30-270/40-100:
60,20%	0,84	30-270/50-75:
40,90%	0,8	360-720/40-100:
30,90%	0,73	360-720/50-75:
Sens al 90%	ABC	Estómago
48,30%	0,83	90-540/40-100:
41,90%	0,76	90-540/50-75:
61,80%	0,87	30-270/40-100:
56,70%	0,8	30-270/50-75:
36,40%	0,79	360-720/40-100:
27,50%	0,71	360-720/50-75:
Sens al 90%	ABC	Recto
40,50%	0,79	90-540/40-100:
31,30%	0,72	90-540/50-75:
48,40%	0,81	30-270/40-100:
40,00%	0,74	30-270/50-75:
32,90%	0,78	360-720/40-100:
28,40%	0,73	360-720/50-75:
Sens al 90%	ABC	Esófago
37,80%	0,79	90-540/40-100:
31,40%	0,74	90-540/50-75:
45,00%	0,81	30-270/40-100:
41,30%	0,76	30-270/50-75:
34,00%	0,79	360-720/40-100:
24,30%	0,72	360-720/50-75:

FIG. 9

RESULTADOS DE ANEMIA EN HOMBRES						
edad	Hgb					
20-59	13,7					
>=60	13,2					
			anemia	sin anemia	sens	espec
ventana temporal	90-365	cancer -pos	256	328	0,44	0,86
edad	40-100	cancer_neg	29.651	181.670		
		prob	0,86%	0,18%		
			anemia	sin anemia	sens	espec
ventana temporal	30-720	cancer -pos	389	400	0,49	0,86
edad	40-100	cancer_neg	30.172	181.541		
		prob	1,27%	0,22%		
			anemia	sin anemia	sens	espec
ventana temporal	365-720	cancer -pos	157	365	0,30	0,87
edad	40-100	cancer_neg	28.289	181.326		
		prob	0,55%	0,20%		
			anemia	sin anemia	sens	espec
ventana temporal	30-720	cancer -pos	231	307	0,43	0,87
edad	50-75	cancer_neg	15.186	100.178		
		prob	1,50%	0,31%		
			anemia	sin anemia	sens	espec
ventana temporal	365-720	cancer -pos	91	280	0,25	0,87
edad	50-75	cancer_neg	14.467	100.079		
		prob	0,63%	0,28%		

FIG. 10