(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 065 630 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
09.09.2020 Bulletin 2020/37

(21) Application number: **14859375.9**

(22) Date of filing: **05.11.2014**

(51) Int Cl.:
G16H 50/30 *(2018.01)* ***G16H 50/20*** *(2018.01)*
G16H 10/40 *(2018.01)* ***G16H 10/60*** *(2018.01)*

(86) International application number:
PCT/IL2014/050960

(87) International publication number:
WO 2015/068157 (14.05.2015 Gazette 2015/19)

(54) **METHODS AND SYSTEMS OF EVALUATING A RISK OF LUNG CANCER**

VERFAHREN UND SYSTEME ZUR BESTIMMUNG EINES LUNGENKREBSRISIKOS

PROCÉDÉS ET SYSTÈMES POUR ÉVALUER UN RISQUE DE CANCER DU POUMON

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **07.11.2013 US 201361901059 P**

(43) Date of publication of application:
14.09.2016 Bulletin 2016/37

(73) Proprietor: **Medial Research Ltd.**
4592000 Kfar-Malal (IL)

(72) Inventors:
- **KALKSTEIN, Nir**
  4639924 Herzlia (IL)
- **KINAR, Yaron**
  6522812 Tel-Aviv (IL)
- **SHALEV, Varda**
  4356420 RaAnana (IL)
- **CHODICK, Gabriel**
  5360501 Givataim (IL)
- **GOLDSHTEIN, Inbal**
  4720416 Ramat-HaSharon (IL)

(74) Representative: **Luppi, Emanuele**
Brunacci & Partners S.r.l.
Via Scaglia Est, 19-31
41126 Modena (IT)

(56) References cited:
EP-A1- 2 518 507
EP-A2- 1 969 363
WO-A1-2009/065230
WO-A1-2010/030697
US-A- 6 059 724
US-A1- 2012 128 702

- **FRANCISCO JAVIER GONZALEZ BARCALA ET AL: "Platelet count: association with prognosis in lung cancer", MEDICAL ONCOLOGY, vol. 27, no. 2, 21 April 2009 (2009-04-21) , pages 357-362, XP055375583, GB ISSN: 1357-0560, DOI: 10.1007/s12032-009-9217-9**
- **SARRAF K M ET AL: "Neutrophil/lymphocyte ratio and its association with survival after complete resection in non-small cell lung cancer", JOURNAL OF THORACIC AND CARDIOVASCULAR SURGERY, MOSBY-YEAR BOOK, INC., ST. LOUIS, MO, US, vol. 137, no. 2, 28 February 2009 (2009-02-28), pages 425-428, XP025892189, ISSN: 0022-5223, DOI: 10.1016/J.JTCVS.2008.05.046 [retrieved on 2008-08-29]**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 3 065 630 B1

Description

FIELD AND BACKGROUND OF THE INVENTION

[0001] The present invention relates to methods and systems of evaluating a risk of lung cancer.

[0002] Lung cancer is the leading cause of cancer death worldwide. In addition, lung cancer has one of the lowest survival outcomes of any cancer since over two-thirds of patients are diagnosed at a late stage when curative treatment is not possible. An effective lung cancer screening will lead to earlier detection of the disease (before patients have symptoms and when treatment is more likely to be effective) and will decrease mortality. Currently, most of the lung cancer cases are diagnosed clinically when patients present with symptoms (such as cough, chest pain, weight loss); unfortunately, patients with these symptoms usually have advanced lung cancer.

[0003] Until very recently, lung cancer screening programs were rarely practiced worldwide, and early detection of lung cancer occurred sporadically through chest radiography. Recent studies indicated that Low Dose Computed Tomography (LDCT) can be used to screen patients who are at high risk for lung cancer. The National Lung Screening Trial (NLST) compared the use of LDCT and chest radiography for screening 53,454 persons at high risk for lung cancer. The study demonstrated a 20% reduction in mortality from lung cancer with LDCT comparing to chest radiography screening. Following the NLST and additional supporting studies, new guidelines for Lung cancer screening were issued recommending the practice of LDCT based lung cancer screening programs.

[0004] Although recommended, lung cancer screening with LDCT has inherited risks: (A) High false positive results, leading to unnecessary testing and invasive procedures, increased costs, and decreased quality of life because of mental anguish. (B) False negative results, which may delay or prevent diagnosis and treatment. (C) Inability to detect small aggressive tumors. (D) Over-diagnosis. (E) Radiation exposure. Therefore, there is a great need to accurately identify the high risk individuals and prevent potential harm from individuals at lower risk. To this end, lung cancer screening guidelines suggest criteria for determining which patients are at high risk. These criteria are based on a combination of age, smoking history, and additional risk factors. Unfortunately, these criteria fail to accurately identify patients at a treatable cancer stage. In fact, the NLST results indicate that in order to prevent one death from lung cancer (in the US), 320 high risk individuals must be screened with LDCT. The implications of such a relatively low rate should be studied to determine if the benefits are greater than the harms of this screening process. Overall, there is a great need to develop a model that identifies patients at high risk for lung cancer (at the pre-screening stage), and enables an efficient, minimal risk screening program by screening only those individuals with high chance of having cancer.

[0005] See also European Patent No EP1969363B1, EP 2 518 507 A1, International Application WO 2010/030697 A1, all teaching certain biomarkers and methods for aiding in the diagnosis of lung cancer in a subject and kits for performing said methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Some embodiments of the invention are herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of embodiments of the invention. In this regard, the description taken with the drawings makes apparent to those skilled in the art how embodiments of the invention may be practiced.

[0007] In the drawings:

FIG. 1 is a flowchart of a method of generating one or more classifiers for estimating a lung cancer risk score according to an analysis of a plurality of individual records, according to some embodiments of the present invention; and

FIG. 2 is a schematic illustration of a system for generating one or more classifiers, for example by implementing the method depicted in FIG. 1, according to some embodiments of the present invention.

DESCRIPTION

[0008] The invention is set out in the appended set of claims. The embodiments of the following description which are not covered by the appended claims are considered as not being part of the present invention.

[0009] According to some embodiments of the present invention, there is provided a lung cancer evaluating system. The system comprises a processor and a memory unit which stores at least one classifier generated according to an analysis of a plurality of historical blood test results of each of another of a plurality of sampled individuals. The system further includes an input unit which receives a plurality of current blood test results taken from a blood of a target individual and a lung cancer evaluating module which evaluates, using the processor, a lung cancer risk of the target individual

by classifying, using the at least one classifier, a set of features extracted from the plurality of current blood test results. Each of the plurality of historical and current blood test results comprises at least the following 18 blood test results: red blood cells (RBC); white blood cell count - WBC (CBC); mean platelet volume (MPV); hemoglobin (HGB); hematocrit (HCT); mean cell volume (MCV); mean cell hemoglobin (MCH); mean corpuscular hemoglobin concentration (MCHC); red cell distribution width (RDW); platelet count (CBC); eosinophils count; neutrophils percentage; monocytes percentage; eosinophils percentage; basophils percentage; neutrophils count; monocytes count; and Platelets hematocrit (PCT). The at least one classifier comprises a member of a group consisting of: a weighted linear regression classifier, a K-Nearest neighbors (KNN) classifier, and a random forest classifier. Optionally, the each of the plurality of historical and current blood test results comprises Biochemistry results which is selected from a group consisting of Erythrocyte Sedimentation Rate (ESR), Glucose, Urea, Blood Urea Nitrogen (BUN), Creatinine, Sodium, Potassium, Chloride, Calcium, Phosphorus, Uric Acid, Bilirubin Total, Lactate Dehydrogenase (LDH), glutamic oxaloacetic transaminase (GOT), Serum glutamic oxaloacetic transaminase (SGOT), and Glutamate Oxaloacetate, Aspartate transaminase (AST), Aspartate Aminotransferase, glutamate pirovate transaminase (GPT) Serum glutamate pirovate transaminase (SGPT), alanine aminotransferase (ALT), Alkaline Phosphatase (Alk Phos/ALP), gamma glutamyl transpeptidase (GGT), Albumin, CK (Creatine Kinase), Iron, HbA1, B12, Vitamin D, G-6-PD, Lithium, Folic Acid, CRP (C reactive protein), low-density lipoprotein (LDL), high-density lipoprotein (HDL), Triglycerides, Total cholesterol, Amylase, PT (Prothrombin Time), Partial Thromboplastin Time (PTT), Activated Partial Thromboplastin Time (APPT), (International Normalized Ratio (INR), Fibrinogen, Cytidine triphosphate (CPT), Ferritin, glomerular filtration rate (GFR), transferrin, Total iron-binding capacity (TIBC), Unsaturated iron-binding capacity (UIBC).

[0010] Optionally, at least one of the plurality of sets of features further comprising at least one demographic parameter of each of the plurality of sampled individuals such that the at least one classifier is further trained using the at least one demographic parameter; wherein the at least one demographic parameter is a member of a group consisting of gender, age, residential zone, race and socio-economic characteristic. The each of the plurality of historical and current blood test results comprises results of white blood cells blood test results including at least one of neutrophils count, basophils count, eosinophils count, lymphocytes count, monocytes count, WBC count, neutrophils percentage, basophils percentage, eosinophils percentage, lymphocytes percentage, monocytes percentage.

Optionally, the each of the plurality of historical and current blood test results comprises platelets blood test results including at least one of platelets count and MPV. Optionally, the each of the plurality of historical and current blood test results comprises a combination of smoking history and red cell test results including at least one of RBC, RDW, MCH, MCV, MCHC, Hematocrit, and Hemoglobin. Optionally, each of the plurality of historical and current blood test results comprises results of both neutrophils percentage/count and lymphocytes percentage/count. Optionally, each of the plurality of historical and current blood test results comprises results of Platelets hematocrit (PCT). Optionally, each of the plurality of historical and current blood test results comprises results of both HGB and HCT.

[0011] According to some embodiments of the present invention, there is provided a method of generating a classifier for a lung cancer risk evaluation. The method comprises providing a processor and a memory unit which stores at least one classifier generated according to an analysis of a plurality of historical blood test results of each of another of a plurality of sampled individuals, receiving a plurality of current blood test results taken from a blood of a target individual, and evaluating, using the processor, a lung cancer risk of the target individual by classifying, using the at least one classifier, a set of features extracted from the plurality of current blood test results. Each of the plurality of historical and current blood test results comprises at least the following 18 blood test results: red blood cells (RBC); white blood cell count - WBC (CBC); mean platelet volume (MPV); hemoglobin (HGB); hematocrit (HCT); mean cell volume (MCV); mean cell hemoglobin (MCH); mean corpuscular hemoglobin concentration (MCHC); red cell distribution width (RDW); platelet count (CBC); eosinophils count; neutrophils percentage; monocytes percentage; eosinophils percentage; basophils percentage; neutrophils count; monocytes count; and Platelets hematocrit (PCT). The at least one classifier comprises a member of a group consisting of: a weighted linear regression classifier, a K-Nearest neighbors (KNN) classifier, and a random forest classifier.

[0012] Optionally, at least one of the plurality of sets of features further comprising at least one demographic parameter of each of the plurality of sampled individuals such that the at least one classifier is further trained using the at least one demographic parameter; wherein the at least one demographic parameter is a member of a group consisting of gender, age, residential zone, race and socio-economic characteristic; wherein the each of the plurality of historical and current blood test results comprises results of white blood cells blood test results including at least one of neutrophils count, basophils count, eosinophils count, lymphocytes count, monocytes count, WBC count, neutrophils percentage, basophils percentage, eosinophils percentage, lymphocytes percentage, monocytes percentage.

[0013] Unless otherwise defined, all technical and/or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, exemplary methods and/or materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be necessarily

limiting.

[0014] Implementation of the method and/or system of embodiments of the invention can involve performing or completing selected tasks manually, automatically, or a combination thereof. Moreover, according to actual instrumentation and equipment of embodiments of the method and/or system of the invention, several selected tasks could be implemented by hardware, by software or by firmware or by a combination thereof using an operating system.

[0015] For example, hardware for performing selected tasks according to embodiments of the invention could be implemented as a chip or a circuit. As software, selected tasks according to embodiments of the invention could be implemented as a plurality of software instructions being executed by a computer using any suitable operating system. In an exemplary embodiment of the invention, one or more tasks according to exemplary embodiments of method and/or system as described herein are performed by a data processor, such as a computing platform for executing a plurality of instructions. Optionally, the data processor includes a volatile memory for storing instructions and/or data and/or a non-volatile storage, for example, a magnetic hard-disk and/or removable media, for storing instructions and/or data. Optionally, a network connection is provided as well. A display and/or a user input device such as a keyboard or mouse are optionally provided as well.

[0016] According to some embodiments of the present invention, there are provided methods and systems of evaluating lung cancer risk by classifying a set of current blood test results of a target individual using one or more classifiers which are generated according to an analysis of historical blood test results of a plurality of individuals.

[0017] Reference is now made to FIG. 1, which is a flowchart of a method 100 of generating one or more classifiers for estimating a lung cancer risk score according to an analysis of a plurality of historical test results of each of a plurality of diagnosed individuals, according to some embodiments of the present invention.

[0018] Reference is also made to FIG. 2, which is a schematic illustration of a system 200 for generating classifier(s) for estimating lung cancer risk scores, for example by implementing the method depicted in FIG. 1, according to some embodiments of the present invention.

[0019] The system 200 includes to one or more medical record database(s) 201 and/or connected to a medical record database interface. The database(s) 201 include a plurality of individual records, also referred to as a plurality of individual samples, which describe, for each of another of a plurality of sampled individuals, one or more sets of a plurality of historical test results each set of another individual, and optionally one or more demographic parameter(s) and a lung cancer diagnosis The set of a plurality of historical test results, demographic parameter(s), such as age, and/or lung cancer prognosis may be stored in a common sample record and/or gathered from a number of independent and/or connected databases. Optionally, the lung cancer prognosis is a binary indication set according to a cancer registry record. The different test results may be of commonly performed blood tests biochemistry tests and/or blood tests held during the same period. Optionally, some sets of a plurality of historical test results have missing blood test results. These results are optionally completed by weighted averaging of the available blood test results of other individuals. The method further includes a processor 204, a classifier generation module 205, and an interface unit 206, such as a network interface.

[0020] As used herein, a demographic parameter includes age, gender, race, weight, national origin, geographical region of residence and/or the like.

[0021] First, as shown at 101, one or more dataset(s) of a plurality of individual samples are provided.

[0022] Optionally, as shown at 102, the plurality of individual samples are screened and/or selected according to matching criteria. For example, the sample records are of individuals in the age of 30 or older who either appear in a cancer registry with lung cancer, and optionally without other types of cancer, or do not appear in the cancer registry. Optionally, sample records of individuals that appear in the cancer registry are taken only if the latest set of a plurality of historical test results they document was taken during a certain period before the registration of a respective individual in the cancer registry, for example during a period of at least 30 days before a current date and at most 2 years. Optionally, sample records of individuals that do not appear in the cancer registry are taken only if they include a set of a plurality of historical test results that creates an equal time-distribution (blood tests timing) for the positive and negative lung cancer populations. The process of equating the time-distribution of the positive and negative samples also leads to omit at least some negative (non-registered) samples and to a change in the lung cancer prevalence in the data set.

[0023] Now, as shown at 103, a derivation dataset, such as a matrix, is generated according to the sample data extracted from the sample records, for example by the classifier generation module 205. The derivation dataset includes a plurality of sets of features, optionally expended. Each set of features is generated from each one of the screened and/or selected sample records. The set of features are optionally unprocessed features which includes actual blood test and/or demographic characteristic values.

[0024] Each sample record includes one or more sets of a plurality of historical test results of a individual, each includes a combination blood test results and/pr biochemistry test results, for example a combination of more than 10, 15, 20 and/or any intermediate number of blood test results or less. In one example, each extracted set of unprocessed features includes at least the following 18 blood test results: red blood cells (RBC); white blood cell count - WBC (CBC); mean platelet volume (MPV); hemoglobin (HGB); hematocrit (HCT); mean cell volume (MCV); mean cell hemoglobin (MCH);

mean corpuscular hemoglobin concentration (MCHC); red cell distribution width (RDW); platelet count (CBC); eosinophils count; neutrophils percentage; monocytes percentage; eosinophils percentage; basophils percentage; neutrophils count; monocytes count; and Platelets hematocrit (PCT). Optionally, this extracted set of unprocessed features further includes one or more of the following blood tests RDW, Platelets, and MCV. Additionally, this extracted set of unprocessed features may further includes one or more of the following blood tests WBC, eosinophils count, neutrophils percentage and/or count, basophils percentage and/or count, and monocytes percentage and/or count.

[0025] The set of current blood test results includes some or more of the following Biochemistry test results: Erythrocyte Sedimentation Rate (ESR), Glucose, Urea, Blood Urea Nitrogen (BUN), Creatinine, Sodium, Potassium, Chloride, Calcium, Phosphorus, Uric Acid, Bilirubin Total, Lactate Dehydrogenase (LDH), glutamic oxaloacetic transaminase (GOT), Serum glutamic oxaloacetic transaminase (SGOT), and Glutamate Oxaloacetate, Aspartate transaminase (AST), Aspartate Aminotransferase, glutamate pirovate transaminase (GPT) Serum glutamate pirovate transaminase (SGPT), alanine aminotransferase (ALT), Alkaline Phosphatase (Alk Phos/ALP), gamma glutamyl transpeptidase (GGT), Albumin, CK (Creatine Kinase), Iron, HbA1, B12, Vitamin D, G-6-PD, Lithium, Folic Acid, CRP (C reactive protein), low-density lipoprotein (LDL), high-density lipoprotein (HDL), Triglycerides, Total cholesterol, Amylase, PT (Prothrombin Time), Partial Thromboplastin Time (PTT), Activated Partial Thromboplastin Time (APPT), (International Normalized Ratio (INR), Fibrinogen, Cytidine triphosphate (CPT), Ferritin, glomerular filtration rate (GFR), transferrin, Total iron-binding capacity (TIBC), Unsaturated iron-binding capacity (UIBC).

[0026] Optionally, the set of unprocessed features is expended. The expended set of features contains features as the above unprocessed blood test results and/or biochemistry test results and/or smoking data and/or one or more demographic parameter(s) and optionally manipulated blood test results and/or combination of blood test results, for instance as described below. For example, each feature in the set of expended features is based on a blood test result, a demographic characteristic, a smoking history, biochemistry test results, a combination of blood test result(s) and/or demographic characteristic(s), and/or a manipulation of blood test result(s) and/or demographic characteristic(s). For example, when the set of unprocessed features includes 18 test results, an expended set of 95 to 190, for example 114 features may be generated.

[0027] Optionally, different derivation datasets, for example matrixes, having different sets of expended features are generated to create different classifiers which classify target individuals having different demographic characteristic(s), for example gender, different smoking history and/or different current data.

[0028] Optionally, the derivation dataset, for example the matrix, is filtered, to remove iteratively outliers. Optionally, an average deviation and/or a standard deviation of each feature is calculated and features having exceptional values, for example more than a standard deviation maximum threshold, for example 10, are truncated to the standard deviation maximum threshold. For example, the process is iteratively repeated 10 times (or less if no truncations are performed). Now, as shown at 104, the derivation dataset is used for generating classifier(s) each classifying a lung cancer risk of a target individual based on one or more demographic characteristics thereof and a current set of a plurality of test results, for example by the classifier generation module 205. Optionally, one or more of the following classifiers may be generated based on the derivation dataset:

a weighted linear regression classifier where positive sample records receive a score that is about 100 times the score of negative sample records;
a K-Nearest neighbors (KNN) classifier, for example 100 times down-sampling of a negative sample record;
a random forest classifier, for example where each tree is built using a 2:1 ratio of negative to positive sample records; and
a gradient boosting machine (GBM) classifier.

[0029] Optionally, the performance of each one of the classifiers is estimated using a 10-fold cross validation process where the derivation dataset, referred to herein as a population, is randomly split to ten equal-sized parts. For each part, the following may be performed:

selecting acceptable sets of blood test results from 90% of population not in the respective part;
training a classifier according to the selected sets of blood test results;
selecting sets of blood test results from a 10% of population in the respective part; and
using the classifier on the selected sets of blood test results from the 10% of population.

[0030] Now, as shown at 105, the classifier(s) are outputted, optionally as a module that allows classifying target individuals, for example by the interface unit 206. Optionally, different classifiers are defined for individuals having different demographic characteristics, for example one classifier for men and another for women. In another example one classifier is used for smokers and another for non smokers.

[0031] Optionally, the classifiers allow evaluating lung cancer by combining the set of current blood test results with

smoking history. The set of current blood test results includes some or more of the following blood tests: eosinophil count; eosinophil percentage; neutrophil count; neutrophil percentage; monocyte count; monocyte percentage; basophil count; basophil percentage; lymphocyte counts; lymphocyte percentage; and white blood cell count (WBC); or at least one result of the following red blood cell count (RBC); red blood cell distribution width (RDW); mean cell volume (MCV); mean cell hemoglobin concentration (MCHC); Hematocrit (HCT); Hemoglobin (HGB); and mean cell hemoglobin (MCH); or at least one result of Platelets count; and mean platelet volume (MPV).

[0032] Optionally, the lung cancer risk is evaluated by result of at least 2 of the above specified blood test groups.

[0033] Optionally, the lung cancer risk is evaluated by classifying biochemistry blood test results of the target individual. In such embodiments, the classifiers are generated according to an analysis of historical biochemistry blood test results of the plurality of individuals, for example as described above. The biochemistry blood test results may include results of any of the above biochemistry tests, includes for example the following blood tests: Albumin, Calcium, Chloride, Cholesterol, Creatinine, high density lipoprotein

[0034] (HDL), low density lipoprotein (LDL), Potassium, Sodium, Triglycerides, Urea, and/or Uric Acid.

[0035] Optionally, the lung cancer risk is evaluated by classifying demographic characteristics of the target individual. In such embodiments, the classifiers are generated according to an analysis of demographic characteristics of the plurality of individuals.

[0036] Optionally, both the current blood test results of the target individual and the historical blood test results of sampled individuals are used for generating expended sets of features which include manipulated and/or weighted values. Optionally, each expended set of features is based on the demographic characteristics of a respective individual, for example as described below.

[0037] Optionally, the one or more classifiers are adapted to one or more demographic characteristics of the target individual. Optionally, the classifiers are selected to match one or more demographic characteristics of the target individual. In such embodiments, different classifiers may be used for women and men and/or for different age groups.

[0038] According to some embodiments of the present invention, there are provided methods and systems of generating one or more classifiers for lung risk evaluation. The methods and systems are based on analysis of a plurality of historical blood test results of each of another of a plurality of sampled individuals and generating accordingly a dataset having a plurality of sets of features each generated according to respective historical blood test results. The dataset is then used to generate and output one or more classifiers, such as K-Nearest neighbors (KNN) classifiers, random forest classifiers, and weighted linear regression classifiers, for example as described above. The classifiers may be provided as modules for execution on client terminals or used as an online service for evaluating lung cancer risk of target individuals based on their current blood test results.

[0039] Classifiers are optionally generated as recited in International Patent Application No. PCT/IL2013/050368 filed on May 2, 2013.

[0040] The following table summarizes the performances of the different classifiers, each generated according to an analysis of a plurality of respective historical blood test results of a plurality of sampled individuals where the blood test results include data from one or two groups, according to some embodiments of the present invention:

	Men 0-30		Men 90-180		Women 0-30		Women 90-180	
	AUC	Sens@95	AUC	Sens@95	AUC	Sens@95	AUC	Sens@95
Optimal (all 4 group parameters)	0.883	56	0.832	39	0.807	44	0.783	33
Only Smx	0.764	29	0.755	21	0.706	23	0.733	21
Only White	0.837	39	0.788	28	0.778	30	0.737	25
Only Red	0.771	29	0.789	26	0.715	26	0.729	22
Only Pits	0.813	39	0.748	27	0.696	27	0.679	12
Smx+White	0.857	46	0.810	35	0.803	37	0.765	28
Smx+Red	0.814	36	0.808	34	0.745	33	0.764	23
Smx+Plts	0.845	42	0.787	28	0.735	34	0.735	22
White+Red	0.857	43	0.816	36	0.793	35	0.762	33
White+Plts	0.862	46	0.800	30	0.788	38	0.747	25
Red+Plts	0.870	47	0.807	34	0.796	39	0.754	29

[0041] Classifier(s) are generated according to an analysis of a plurality of respective historical blood test results of a plurality of sampled individuals where the blood test results include at least one of:

1) White cells test results, for brevity referred to as White and includes at least one of neutrophils count, basophils

count, eosinophils count, lymphocytes count, monocytes count, WBC count, neutrophils percentage, basophils percentage, eosinophils percentage, lymphocytes percentage, monocytes percentage;

2) Platelets cells test results, for brevity referred to as Pits and includes count and/or MPV;

3) Biochemistry test results is selected from a group consisting of Erythrocyte Sedimentation Rate (ESR), Glucose, Urea, Blood Urea Nitrogen (BUN), Creatinine, Sodium, Potassium, Chloride, Calcium, Phosphorus, Uric Acid, Bilirubin Total, Lactate Dehydrogenase (LDH), glutamic oxaloacetic transaminase (GOT), Serum glutamic oxaloacetic transaminase (SGOT), and Glutamate Oxaloacetate, Aspartate transaminase (AST), Aspartate Aminotransferase, glutamate pirovate transaminase (GPT) Serum glutamate pirovate transaminase (SGPT), alanine aminotransferase (ALT), Alkaline Phosphatase (Alk Phos/ALP), gamma glutamyl transpeptidase (GGT), Albumin, CK (Creatine Kinase), Iron, HbA1, B12, Vitamin D, G-6-PD, Lithium, Folic Acid, CRP (C reactive protein), low-density lipoprotein (LDL), high-density lipoprotein (HDL), Triglycerides, Total cholesterol, Amylase, PT (Prothrombin Time), Partial Thromboplastin Time (PTT), Activated Partial Thromboplastin Time (APPT), (International Normalized Ratio (INR), Fibrinogen, Cytidine triphosphate (CPT), Ferritin, glomerular filtration rate (GFR), transferrin, Total iron-binding capacity (TIBC), Unsaturated iron-binding capacity (UIBC).

4) red cells test results, for brevity referred to Red and includes at least one of the following parameters RBC, RDW, MCV, MCHC, Hematocrit, Hemoglobin and MCH.

[0042] For example, classifiers are based only on White or Plts. It should be noted that a classifier based on only one of White or Pits yields better outcome than a classifier based on Smx only (see table above). This is not trivial as smoking history is well documented as being correlated to lung cancer.

[0043] Optionally, classifiers are based only on 2 of the above group parameters, for example Smx and White, Smx and Plts, Smx and Red, Red and White, Pits and White, and Pits and Red.

[0044] Classifier(s) are generated according to an analysis of age and/or gender of each individual.

[0045] Classifier(s) generate classifications, also referred to as predictions. The classifications are optionally collected to measure performance of each classifier. For example, the measures of performance are selected according to a receiving operating characteristic (ROC) curve. Optionally, specificity at different (5%, 10%, 20%, 50%, and 70%) sensitivity (recall) values are used for identifying the measures. The performances of the different exemplary classifiers are summarized in the tables which respectively have different area under curve (AUC).

[0046] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of a plurality of blood count parameter(s). Exemplary data in such embodiments may be: AUC = 0.94 [0.93, 0.96], OR at sensitivity (SENS 10%) SENS10 = 438 [123, 626], SENS at false positive ratio (FPR 10) = 85.4% [80.7, 89.8], SENS at FPR1 = 41.8% [34.9, 50.5]. Exemplary data herein below is supported by optimal performance as given by records of medical tests from a Memorial Healthcare System (MHS) of a time-window of 0-30 days, of patients at the age group of 50-75. The records are selected according to sensitivity at FPR of 10%.

[0047] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of a plurality of white blood count parameter(s). Exemplary data in such embodiments may be: AUC = 0.94 [0.92, 0.95], OR at SENS10 = 260 [87, 624], SENS at FPR10 = 79.4% [74.7, 84.0], SENS at FPR1 = 38.0% [30.6, 45.4].

[0048] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of a plurality of red blood count parameter(s). Exemplary data in such embodiments may be: AUC = 0.88 [0.86, 0.90], OR at SENS10 = 88 [36, 208], SENS at FPR10 = 64.1% [58.8, 70.0], SENS at FPR1 = 27.4% [20.7, 33.9].

[0049] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of a plurality of platelets count parameter(s). Exemplary data in such embodiments may be: AUC = 0.91 [0.89, 0.92], OR at SENS10 = 149 [41, 614], SENS at FPR10 = 72.9% [66.9, 78.3], SENS at FPR1 = 35.4% [28.1,42.3].

[0050] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of platelets information. Exemplary data in such embodiments may be: AUC = 0.94 [0.92, 0.95], OR at SENS10 = 232 [76, 621], SENS at FPR10 = 80.7% [75.6, 85.2], SENS at FPR1 = 37.7% [31.2, 44.7].

[0051] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of each of white blood counts and any parameter of the red blood counts. Exemplary data in such embodiments may be: AUC = 0.93 [0.92, 0.95], OR at SENS10 = 326 [77, 625], SENS at FPR10 = 80.1% [75.0, 84.8], SENS at FPR1 = 38.2% [32.3, 45.2].

[0052] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of any parameter of the white blood counts and any parameter of the platelets information or any parameter of the white line counts and any biochemistry parameter.

[0053] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of any parameter of the red blood counts and any parameter of the platelets information. In such

embodiments, AUC = 0.92 [0.90, 0.93], OR at SENS10 = 194 [57, 619], SENS at FPR10 = 76.6% [71.1, 81.6], SENS at FPR1 = 35.2% [27.4, 42.2].

[0054] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of the platelets information and any biochemistry parameter.

[0055] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of the red blood counts and any biochemistry parameter.

[0056] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of Neutrophils counts. In such embodiments, AUC = 0.90 [0.89, 0.92], OR at SENS 10 = 44 [23, 89], SENS at FPR10 = 69.2% [62.9, 74.8], SENS at FPR1 = 22.2% [16.9, 28.6].

[0057] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of Hematocrit. In such embodiments, AUC = 0.88 [0.86, 0.90], OR at SENS10 = 88 [36, 208], SENS at FPR10 = 64.1% [58.8, 70.0], SENS at FPR1 = 27.4% [20.7, 33.9].

[0058] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of Platelets count. Exemplary data in such embodiments may be: AUC = 0.91 [0.89, 0.92], OR at SENS10 = 149 [41, 614], SENS at FPR10 = 72.9% [66.9, 78.3], SENS at FPR1 = 35.4% [28.1, 42.3].

[0059] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of white line counts and any parameter of the red line counts, and any parameter of the platelets information. Exemplary data in such embodiments may be: AUC = 0.94 [0.93, 0.96], OR at SENS 10 = 438 [123, 626], SENS at FPR10 = 85.4% [80.7, 89.8], SENS at FPR1 = 41.8% [34.9, 50.5].

[0060] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of any parameter of the white line counts, and any parameter of the red line counts, and any parameter of the platelets information and any biochemistry parameter.

[0061] Optionally, each of the plurality of sampled individuals in the individual records includes past value(s) as well as the current value(s) of any parameter from CBC or biochemistry and family history of lung cancer, any parameter of the white blood counts and family history of lung cancer, any information of the red blood count and family history of lung cancer, any parameter from CBC or biochemistry and body mass index (BMI), any parameter of the white blood counts and BMI, any information of the red blood count and BMI. any parameter from CBC or biochemistry and comorbidity (e.g. COPD), any parameter of the white blood counts and comorbidity, any information of the red blood count and comorbidity, any parameter from CBC or biochemistry and socio-economic indicators (e.g. Education level) any parameter of the white blood counts and socio-economic indicators, and/or any information of the red blood count and socio-economic indicators.

[0062] It is expected that during the life of a patent maturing from this application many relevant systems and methods will be developed and the scope of the term a processor, a display, and user interface is intended to include all such new technologies *a priori.*

[0063] As used herein the term "about" refers to $\pm$ 10 %.

[0064] The terms "comprises", "comprising", "includes", "including", "having" and their conjugates mean "including but not limited to". This term encompasses the terms "consisting of" and "consisting essentially of".

[0065] The phrase "consisting essentially of" means that the composition or method may include additional ingredients and/or steps, but only if the additional ingredients and/or steps do not materially alter the basic and novel characteristics of the claimed composition or method.

[0066] As used herein, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a compound" or "at least one compound" may include a plurality of compounds, including mixtures thereof.

[0067] The word "exemplary" is used herein to mean "serving as an example, instance or illustration". Any embodiment described as "exemplary" is not necessarily to be construed as preferred or advantageous over other embodiments and/or to exclude the incorporation of features from other embodiments.

[0068] The word "optionally" is used herein to mean "is provided in some embodiments and not provided in other embodiments". Any particular embodiment of the invention may include a plurality of "optional" features unless such features conflict.

[0069] Throughout this application, various embodiments of this invention may be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[0070] Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases "ranging/ranges between" a first indicate number and a second indicate number

and "ranging/ranges from" a first indicate number "to" a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals therebetween.

[0071] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination or as suitable in any other described embodiment of the invention. Certain features described in the context of various embodiments are not to be considered essential features of those embodiments, unless the embodiment is inoperative without those elements.

Claims

1. A lung cancer evaluating system, comprising:

   a processor;
   a memory unit which stores at least one classifier generated according to an analysis of a plurality of historical blood test results of each of another of a plurality of sampled individuals;
   an input unit which receives a plurality of current blood test results taken from a blood of a target individual; and
   a lung cancer evaluating module which evaluates, using said processor, a lung cancer risk of said target individual by classifying, using said at least one classifier, a set of features extracted from said plurality of current blood test results;
   characterized by:

   each of said plurality of historical and current blood test results comprises at least the following 18 blood test results: red blood cells (RBC); white blood cell count - WBC (CBC); mean platelet volume (MPV); hemoglobin (HGB); hematocrit (HCT); mean cell volume (MCV); mean cell hemoglobin (MCH); mean corpuscular hemoglobin concentration (MCHC); red cell distribution width (RDW); platelet count (CBC); eosinophils count; neutrophils percentage; monocytes percentage; eosinophils percentage; basophils percentage; neutrophils count; monocytes count; and Platelets hematocrit (PCT); and
   said at least one classifier comprises a member of a group consisting of: a weighted linear regression classifier, a K-Nearest neighbors (KNN) classifier, and a random forest classifier.

2. The lung cancer evaluating system of claim 1, wherein said each of said plurality of historical and current blood test results comprises Biochemistry results which are selected from a group consisting of Erythrocyte Sedimentation Rate (ESR), Glucose, Urea, Blood Urea Nitrogen (BUN), Creatinine, Sodium, Potassium, Chloride, Calcium, Phosphorus, Uric Acid, Bilirubin Total, Lactate Dehydrogenase (LDH), glutamic oxaloacetic transaminase (GOT), Serum glutamic oxaloacetic transaminase (SGOT), and Glutamate Oxaloacetate, Aspartate transaminase (AST), Aspartate Aminotransferase, glutamate pirovate transaminase (GPT) Serum glutamate pirovate transaminase (SGPT), alanine aminotransferase (ALT), Alkaline Phosphatase (Alk Phos/ALP), gamma glutamyl transpeptidase (GGT), Albumin, CK (Creatine Kinase), Iron, HbA1, B12, Vitamin D, G-6-PD, Lithium, Folic Acid, CRP (C reactive protein), low-density lipoprotein (LDL), high-density lipoprotein (HDL), Triglycerides, Total cholesterol, Amylase, PT (Prothrombin Time), Partial Thromboplastin Time (PTT), Activated Partial Thromboplastin Time (APPT), (International Normalized Ratio (INR), Fibrinogen, Cytidine triphosphate (CPT), Ferritin, glomerular filtration rate (GFR), transferrin, Total iron-binding capacity (TIBC), Unsaturated iron-binding capacity (UIBC).

3. The lung cancer evaluating system of claim 1, wherein at least one of said plurality of sets of features further comprising at least one demographic parameter of each of said plurality of sampled individuals such that said at least one classifier is further trained using said at least one demographic parameter; wherein said at least one demographic parameter is a member of a group consisting of gender, age, residential zone, race and socio-economic characteristic;
   wherein said each of said plurality of historical and current blood test results comprises results of white blood cells blood test results including at least one of neutrophils count, basophils count, eosinophils count, lymphocytes count, monocytes count, WBC count, neutrophils percentage, basophils percentage, eosinophils percentage, lymphocytes percentage, monocytes percentage.

4. The lung cancer evaluating system of claim 1, wherein said each of said plurality of historical and current blood test results comprises platelets blood test results including at least one of platelets count and MPV.

5. The lung cancer evaluating system of claim 1, wherein said each of said plurality of historical and current blood test results comprises a combination of smoking history and red cell test results including at least one of RBC, RDW, MCH, MCV, MCHC, Hematocrit, and Hemoglobin.

6. The lung cancer evaluating system of claim 1, wherein each of said plurality of historical and current blood test results comprises results of both neutrophils percentage/count and lymphocytes percentage/count.

7. The lung cancer evaluating system of claim 1, wherein each of said plurality of historical and current blood test results comprises results of Platelets hematocrit (PCT).

8. The lung cancer evaluating system of claim 1, wherein each of said plurality of historical and current blood test results comprises results of both HGB and HCT.

9. A method of generating a classifier for a lung cancer risk evaluation, comprising:

providing a processor and a memory unit which stores at least one classifier generated according to an analysis of a plurality of historical blood test results of each of another of a plurality of sampled individuals;
receiving a plurality of current blood test results taken from a blood of a target individual; and
evaluating, using said processor, a lung cancer risk of said target individual by classifying, using said at least one classifier, a set of features extracted from said plurality of current blood test results;
characterized by:

each of said plurality of historical and current blood test results comprises at least the following 18 blood test results: red blood cells (RBC); white blood cell count - WBC (CBC); mean platelet volume (MPV); hemoglobin (HGB); hematocrit (HCT); mean cell volume (MCV); mean cell hemoglobin (MCH); mean corpuscular hemoglobin concentration (MCHC); red cell distribution width (RDW); platelet count (CBC); eosinophils count; neutrophils percentage; monocytes percentage; eosinophils percentage; basophils percentage; neutrophils count; monocytes count; and Platelets hematocrit (PCT), and
said at least one classifier comprises a member of a group consisting of: a weighted linear regression classifier, a K-Nearest neighbors (KNN) classifier, and a random forest classifier.

10. The method of claim 9, wherein at least one of said plurality of sets of features further comprising at least one demographic parameter of each of said plurality of sampled individuals such that said at least one classifier is further trained using said at least one demographic parameter; wherein said at least one demographic parameter is a member of a group consisting of gender, age, residential zone, race and socio-economic characteristic; wherein said each of said plurality of historical and current blood test results comprises results of white blood cells blood test results including at least one of neutrophils count, basophils count, eosinophils count, lymphocytes count, monocytes count, WBC count, neutrophils percentage, basophils percentage, eosinophils percentage, lymphocytes percentage, monocytes percentage.

Patentansprüche

1. Lungenkrebsbewertungssystem, umfassend:

einen Prozessor;
eine Speichereinheit, die mindestens einen Klassifizierer speichert, der gemäß einer Analyse einer Vielzahl von historischen Bluttestergebnissen eines jeden anderen von einer Vielzahl von untersuchten Individuen erzeugt wurde;
eine Eingabeeinheit, die eine Vielzahl von aktuellen Bluttestergebnissen erhält, die aus dem Blut eines Zielindividuums entnommen wurden; und
ein Lungenkrebsbewertungsmodul, das unter Verwendung des Prozessors ein Lungenkrebsrisiko des Zielindividuums bewertet, indem es unter Verwendung des mindestens einen Klassifizierers einen Satz von Merkmalen klassifiziert, die aus der Vielzahl von aktuellen Bluttestergebnissen extrahiert wurden;

gekennzeichnet durch:

jedes der Vielzahl von historischen und aktuellen Bluttestergebnissen umfasst mindestens die folgenden 18

Bluttestergebnisse: rote Blutkörperchen (RBC); Anzahl der weißen Blutkörperchen - WBC (CBC); mittleres Thrombozytenvolumen (MPV); Hämoglobin (HGB); Hämatokrit (HCT); mittleres Zellvolumen (MCV); mittleres Zellhämoglobin (MCH); mittlere korpuskulare Hämoglobinkonzentration (MCHC); Breitenverteilung der roten Blutkörperchen (RDW); Thrombozytenzahl (CBC); Eosinophilenzahl; Neutrophile in Prozent; Monozyten in Prozent; Eosinophile in Prozent; Basophile in Prozent; Neutrophilenzahl; Monozytenzahl; und Thrombozyten-Hämatokrit (PCT); und

der mindestens eine Klassifizierer umfasst ein Mitglied einer Gruppe, bestehend aus: einem Klassifizierer mit gewichteter linearer Regression, einem K-nächste-Nachbarn (KNN)-Klassifizierer und einem Random-Forest-Klassifizierer.

2. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse biochemische Ergebnisse umfasst, die ausgewählt sind aus einer Gruppe bestehend aus Erythrozytensedimentationsrate (ESR), Glukose, Harnstoff, Blut-Harnstoff-Stickstoff (BUN), Kreatinin, Natrium, Kalium, Chlorid, Kalzium, Phosphor, Harnsäure, Bilirubin Gesamt, Laktat-Dehydrogenase (LDH), Glutamin-Oxalessigsäure-Transaminase (GOT), Serum-Glutamin-Oxalessigsäure-Transaminase (SGOT) und Glutamat-Oxalacetat, Aspartat-Transaminase (AST), Aspartat-Aminotransferase, Glutamat-Pirovat-Transaminase (GPT) Serum-Glutamat-Pirovat-Transaminase (SGPT), Alanin-Aminotransferase (ALT), Alkalische Phosphatase (Alk Phos/ALP), Gamma-Glutamyl-Transpeptidase (GGT), Albumin, CK (Kreatinkinase), Eisen, HbA1, B12, Vitamin D, G-6-PD, Lithium, Folsäure, CRP (C-reaktives Protein), Low-Density-Lipoprotein (LDL), High-Density-Lipoprotein (HDL), Triglyceride, Gesamtcholesterin, Amylase, PT (Prothrombinzeit), partielle Thromboplastinzeit (PTT), aktivierte partielle Thromboplastinzeit (APPT), (Internationales normalisiertes Verhältnis (INR), Fibrinogen, Cytidintriphosphat (CPT), Ferritin, glomeruläre Filtrationsrate (GFR), Transferrin, Gesamteisenbindungskapazität (TIBC), ungesättigte Eisenbindungskapazität (UIBC).

3. Lungenkrebsbewertungssystem nach Anspruch 1, wobei mindestens einer der mehreren Sätze von Merkmalen ferner mindestens einen demographischen Parameter von jedem der mehreren untersuchten Individuen umfasst, so dass der mindestens eine Klassifizierer unter Verwendung des mindestens einen demographischen Parameters weiter trainiert wird; wobei der mindestens eine demographische Parameter ein Mitglied einer Gruppe ist, die aus Geschlecht, Alter, Wohngebiet, Rasse und sozioökonomischem Merkmal besteht; wobei jedes der Vielzahl von historischen und aktuellen Bluttestergebnissen Ergebnisse von Bluttestergebnissen für weiße Blutkörperchen umfasst, die mindestens eines der folgenden umfassen: Neutrophilenzahl, Basophilenzahl, Eosinophilenzahl, Lymphozytenzahl, Monozytenzahl, WBC Zahl, Neutrophile in Prozent, Basophile in Prozent, Eosinophile in Prozent, Lymphozyten in Prozent, Monozyten in Prozent.

4. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse Thrombozyten-Bluttestergebnisse umfasst, die mindestens eines von Thrombozytenzahl und MPV enthalten.

5. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse eine Kombination aus der Rauchergeschichte und rote Blutkörperchen-Testergebnissen umfasst einschließlich mindestens eines von RBC, RDW, MCH, MCV, MCHC, Hämatokrit und Hämoglobin.

6. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse die Ergebnisse sowohl von Neutrophilenzahl/ -Prozentsatz als auch von Lymphozytenzahl/ -Prozentsatz umfasst.

7. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse die Ergebnisse des Thrombozyten-Hämatokrit (PCT) umfasst.

8. Lungenkrebsbewertungssystem nach Anspruch 1, wobei jedes der Vielzahl von historischen und aktuellen Bluttestergebnissen Ergebnisse sowohl von HGB als auch von HCT umfasst.

9. Verfahren zur Erzeugung eines Klassifizierers für eine Lungenkrebsrisikobewertung, umfassend:

Bereitstellung eines Prozessors und einer Speichereinheit, die mindestens einen Klassifizierer speichert, der gemäß einer Analyse einer Vielzahl von historischen Bluttestergebnissen eines jeden anderen aus einer Vielzahl von untersuchten Individuen erzeugt wurde;

Erhalt einer Vielzahl aktueller Bluttestergebnisse, die aus dem Blut eines Zielindividuums entnommen wurden;

und
Bewertung eines Lungenkrebsrisikos des Zielindividuums unter Verwendung des Prozessors, indem unter Verwendung des mindestens einen Klassifizierers ein Satz von Merkmalen klassifiziert wird, die aus der Vielzahl der aktuellen Bluttestergebnisse extrahiert wurden;

gekennzeichnet durch:
jedes der Mehrzahl von historischen und aktuellen Bluttestergebnissen umfasst mindestens die folgenden 18 Bluttestergebnisse: rote Blutkörperchen (RBC); Anzahl der weißen Blutkörperchen - WBC (CBC); mittleres Thrombozytenvolumen (MPV); Hämoglobin (HGB); Hämatokrit (HCT); mittleres Zellvolumen (MCV); mittleres Zellhämoglobin (MCH); mittlere korpuskulare Hämoglobinkonzentration (MCHC); Breitenverteilung der roten Blutkörperchen (RDW); Thrombozytenzahl (CBC); Eosinophilenzahl; Neutrophile in Prozent; Monozyten in Prozent, Eosinophile in Prozent, Basophile in Prozent, Neutrophilenzahl, Monozytenzahl; und Thrombozyten-Hämatokrit (PCT), und
der mindestens eine Klassifizierer umfasst ein Mitglied einer Gruppe, die besteht aus: einem Klassifizierer mit gewichteter linearer Regression, einem K-nächste-Nachbarn (KNN)-Klassifizierer und einem Random-Forest-Klassifizierer.

10. Verfahren nach Anspruch 9, wobei mindestens einer der mehreren Sätze von Merkmalen ferner mindestens einen demographischen Parameter von jedem der mehreren untersuchten Individuen umfasst, so dass der mindestens eine Klassifizierer unter Verwendung des mindestens einen demographischen Parameters weiter trainiert wird;
wobei der mindestens eine demographische Parameter ein Mitglied einer Gruppe ist, die aus Geschlecht, Alter, Wohngebiet, Rasse und sozioökonomischen Merkmalen besteht; wobei jedes der mehreren historischen und aktuellen Bluttestergebnisse Ergebnisse von Bluttestergebnissen für weiße Blutkörperchen umfasst, einschließlich mindestens eines der folgenden: Neutrophilenzahl, Basophilenzahl, Eosinophilenzahl, Lymphozytenzahl, Monozytenzahl, WBC Zahl, Neutrophile in Prozent, Basophile in Prozent, Eosinophile in Prozent, Lymphozyten in Prozent, Monozyten in Prozent.

Revendications

1. Système d'évaluation du cancer du poumon comprenant :

un processeur ;
une unité de mémoire qui stocke au moins un classificateur généré selon une analyse d'une pluralité de résultats de tests sanguins historiques pour chaque autre individu d'une pluralité d'individus échantillonnés ;
une unité d'entrée qui reçoit une pluralité de résultats de tests sanguins courants obtenus à partir du sang d'un individu cible ; et
un module d'évaluation du cancer du poumon qui évalue, en utilisant ledit processeur,
un risque de cancer du poumon dudit individu cible en classant, en utilisant ledit au moins un classificateur, un ensemble de caractéristiques extraites de ladite pluralité de résultats de tests sanguins courants ;
caractérisé par :

chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend au moins les 18 résultats de tests sanguins suivants : globules rouges (RBC) ; numération des globules blancs - WBC (CBC) ; volume plaquettaire moyen (MPV) ; hémoglobine (HGB) ; hématocrite (HCT) ; volume globulaire moyen (MCV) ; hémoglobine cellulaire moyenne (MCH) ; concentration corpusculaire moyenne en hémoglobine (MCHC) ; indice de distribution des globules rouges (RDW) ; numération plaquettaire (CBC) ; numération des éosinophiles ; pourcentage des neutrophiles ; pourcentage des monocytes ; pourcentage des éosinophiles ; pourcentage des basophiles ; numération des neutrophiles ; numération des monocytes ; et hématocrite plaquettaire (PCT) ; et
ledit au moins un classificateur comprend un élément d'un groupe constitué : d'un classificateur par régression linéaire pondérée, d'un classificateur par les K plus proches voisins (KNN), et d'un classificateur par forêts aléatoires.

2. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats d'analyse biochimique qui sont sélectionnés parmi un groupe constitué de la vitesse de sédimentation des érythrocytes (ESR), du glucose, de l'urée, de l'azote uréique du sang (BUN), de la créatinine, du sodium, du potassium, du chlorure, du calcium, du phosphore, de l'acide urique, de la bilirubine totale, de la déshydrogénase lactique (LDH), de la transaminase glutamo-oxaloa-

cétique (GOT), du sérum de transaminase glutamo-oxaloacétique (SGOT), et du glutamate oxaloacétate, de l'aspartate transaminase (AST), de l'aspartate aminotransférase, du glutamate pirovate transaminase (GPT), du sérum de glutamate pirovate transaminase (SGPT), de l'alanine aminotransférase (ALT), de la phosphatase alcaline (Alk Phos/ALP), de la gamma glutamyle transpeptidase (GGT), de l'albumine, de la CK (créatine kinase), du fer, de l'HbA1, de la B12, de la vitamine D, G-6-PD, du lithium, de l'acide folique, de la CRP (protéine C réactive), de la lipoprotéine de faible densité (LDL), de la lipoprotéine de haute densité (HDL), des triglycérides, du cholestérol total, de l'amylase, du PT (temps de prothrombine), du temps de thromboplastine partielle (PTT), du temps de thromboplastine partielle activée (APPT), du rapport international normalisé (INR), du fibrinogène, du cytidine triphosphate (CPT), de la ferritine, du taux de filtration glomérulaire (GFR), de la transferrine, de la capacité totale de liaison du fer (TIBC), de la capacité de liaison du fer insaturé (UIBC).

3. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel au moins l'un de ladite pluralité d'ensembles de caractéristiques comprend en outre au moins un paramètre démographique de chacun de ladite pluralité d'individus échantillonnés de sorte que ledit au moins un classificateur est en outre entraîné en utilisant ledit au moins un paramètre démographique ; dans lequel ledit au moins un paramètre démographique est un élément d'un groupe constitué du sexe, de l'âge, de la région de résidence, de l'origine ethnique et de caractéristiques socio-économiques ; dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats de tests sanguins sur les globules blancs incluant au moins l'un parmi la numération des neutrophiles, la numération des basophiles, la numération des éosinophiles, la numération des lymphocytes, la numération des monocytes, la numération des WBC, le pourcentage des neutrophiles, le pourcentage des basophiles, le pourcentage des éosinophiles, le pourcentage des lymphocytes, le pourcentage des monocytes.

4. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats de tests sanguins sur les plaquettes, incluant au moins l'un de la numération des plaquettes et du MPV.

5. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend une combinaison d'antécédents de tabagisme et de résultats d'analyse sur les globules rouges incluant au moins l'un parmi RBC, RDW, MCH, MCV, MCHC, hématocrite et hémoglobine.

6. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats à la fois sur la numération/le pourcentage des neutrophiles et sur la numération/le pourcentage des lymphocytes.

7. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats sur l'hématocrite plaquettaire (PCT).

8. Système d'évaluation du cancer du poumon selon la revendication 1, dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats à la fois sur l'HGB et l'HCT.

9. Procédé pour générer un classificateur pour l'évaluation d'un risque de cancer du poumon, comprenant :

la fourniture d'un processeur et d'une unité de mémoire qui stocke au moins un classificateur généré selon une analyse d'une pluralité de résultats de tests sanguins historiques pour chaque autre individu d'une pluralité d'individus échantillonnés ;
la réception d'une pluralité de résultats de tests sanguins courants obtenus à partir du sang d'un individu cible ; et l'évaluation, en utilisant ledit processeur, d'un risque de cancer du poumon pour ledit individu cible en classant, en utilisant ledit au moins un classificateur, un ensemble de caractéristiques extraites de ladite pluralité de résultats de tests sanguins courants ;
caractérisé par :

chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend au moins les 18 résultats de tests sanguins suivants : globules rouges (RBC) ; numération des globules blancs - WBC (CBC) ; volume plaquettaire moyen (MPV) ; hémoglobine (HGB) ; hématocrite (HCT) ; volume globulaire moyen (MCV) ; hémoglobine cellulaire moyenne (MCH) ; concentration corpusculaire moyenne en hémoglobine (MCHC) ; indice de distribution des globules rouges (RDW) ; numération plaquettaire (CBC) ; numération des éosinophiles ; pourcentage des neutrophiles ; pourcentage des monocytes ; pourcentage des

éosinophiles ; pourcentage des basophiles ; numération des neutrophiles ; numération des monocytes ; et hématocrite plaquettaire (PCT), et
ledit au moins un classificateur comprend un élément d'un groupe constitué : d'un classificateur par régression linéaire pondérée, d'un classificateur par les K plus proches voisins (KNN), et d'un classificateur par forêts aléatoires.

10. Procédé selon la revendication 9, dans lequel au moins l'un de ladite pluralité d'ensembles de caractéristiques comprend en outre au moins un paramètre démographique de chacun de ladite pluralité d'individus échantillonnés de sorte que ledit au moins un classificateur est en outre entraîné en utilisant ledit au moins un paramètre démographique ; dans lequel ledit au moins un paramètre démographique est un élément d'un groupe constitué du sexe, de l'âge, de la région de résidence, de l'origine ethnique et de caractéristiques socio-économiques ; dans lequel chacun de ladite pluralité de résultats de tests sanguins historiques et courants comprend des résultats de tests sanguins sur les globules blancs incluant au moins l'un parmi la numération des neutrophiles, la numération des basophiles, la numération des éosinophiles, la numération des lymphocytes, la numération des monocytes, la numération des WBC, le pourcentage des neutrophiles, le pourcentage des basophiles, le pourcentage des éosinophiles, le pourcentage des lymphocytes, le pourcentage des monocytes.

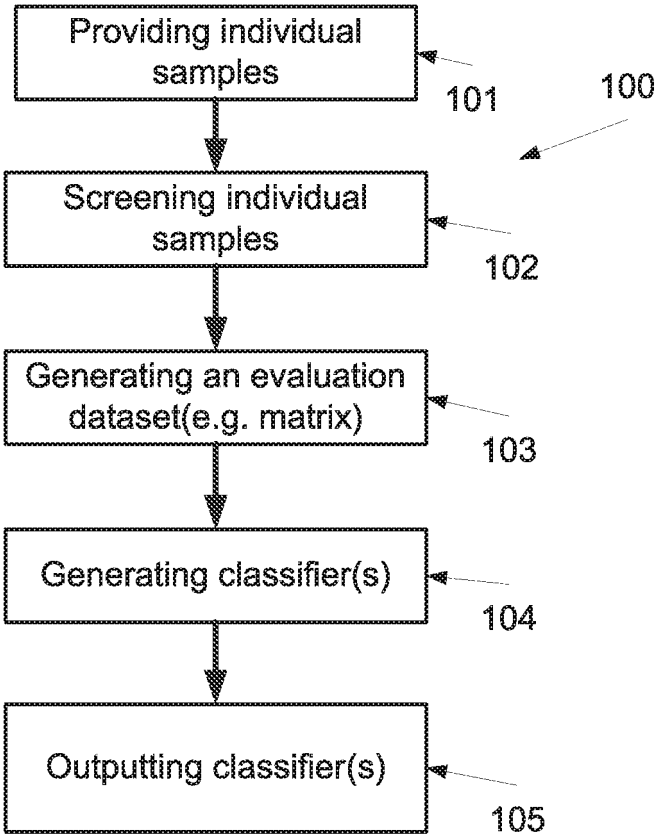
Providing individual samples
101
100
Screening individual samples
102
Generating an evaluation dataset(e.g. matrix)
103
Generating classifier(s)
104
Outputting classifier(s)
105

FIG. 1

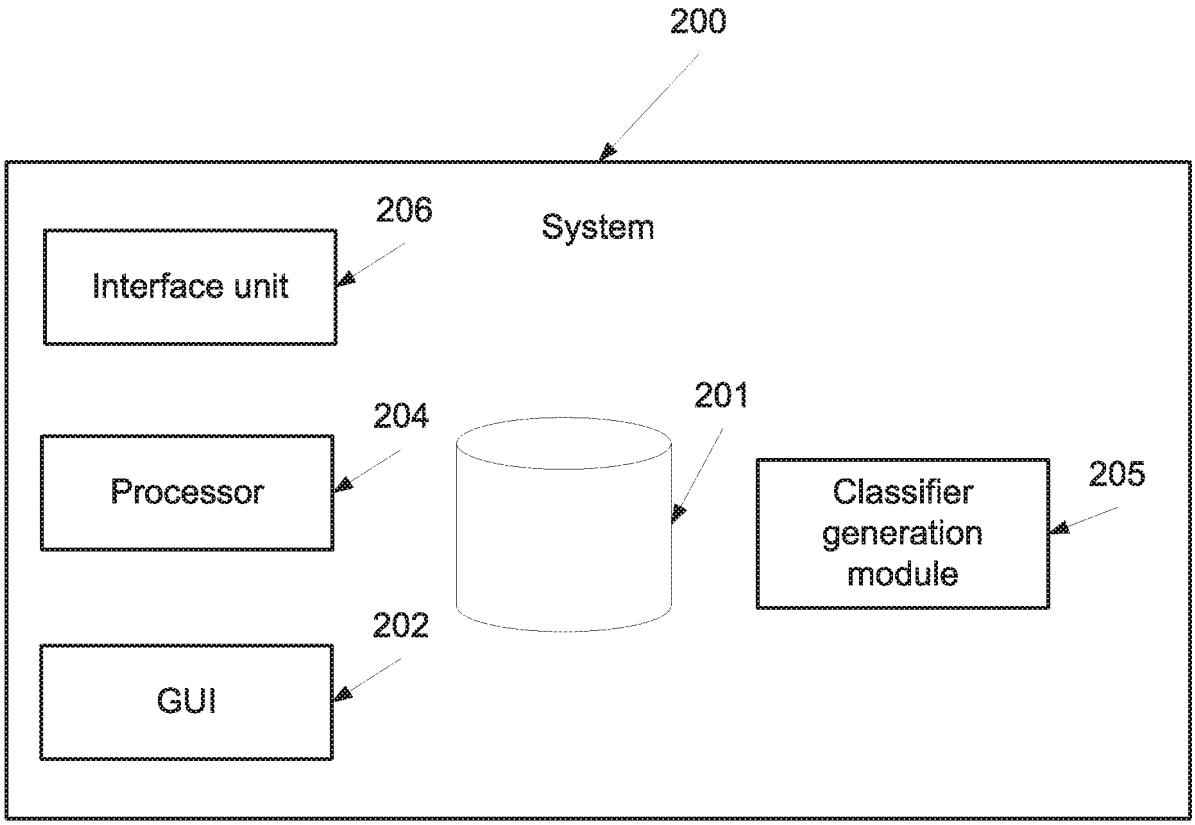
200
System
206
Interface unit
204
Processor
201
202
GUI
205
Classifier generation module

FIG. 2

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 1969363 B1 **[0005]**
- EP 2518507 A1 **[0005]**
- WO 2010030697 A1 **[0005]**
- IL 2013050368 W **[0039]**