(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 May 2003 (15.05.2003)

PCT

(10) International Publication Number
WO 03/040987 A2

(51) International Patent Classification[7]: G06F 19/00

(21) International Application Number: PCT/US02/35305

(22) International Filing Date:
4 November 2002 (04.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/355,542 2 November 2001 (02.11.2001) US

(71) Applicant: SIEMENS CORPORATE RESEARCH, INC. [US/US]; 755 College Road East, Princeton, NJ 08540 (US).

(72) Inventors: RAO, R. Bharat; 2060 St. Andrews Drive, Berwyn, PA 19312 (US). SANDILYA, Sathyakama; 28-12 Phesant Hollow Drive, Plainsboro, NJ 08536 (US).

(74) Agents: PASCHBURG, Donald, B._ et al.; Siemens Corporation - Intellectual Property Dept., 186 Wood Ave. South, Iselin, NJ 08830 (US).

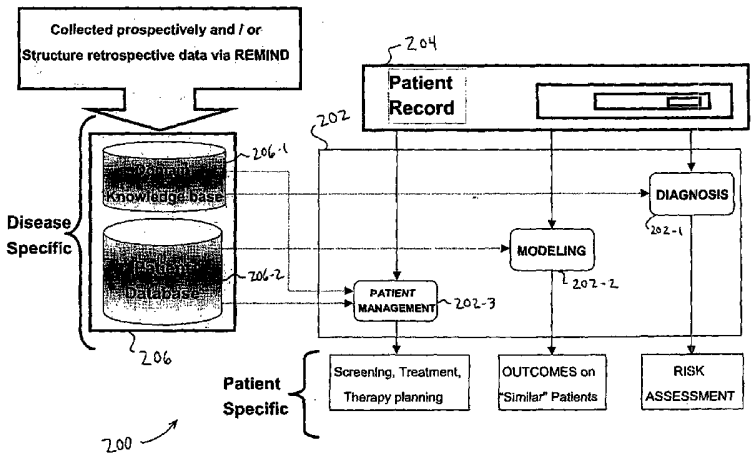
(81) Designated States *(national)*: CA, CN, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PATIENT DATA MINING FOR LUNCH CANCER SCREENING



(57) Abstract: A system and method for lung cancer screening is provided. The system includes a database (206) including structured patient information (206-2) for a patient population and a domain knowledge base (206-1) including information about lung cancer; an individual patient record (204); and a processor (202) for analyzing the patient record with data from the database to determine if a patient has indications of lung cancer. The method includes the steps of inputting patient-specific data into a patient record; performing at least one lung cancer screening procedure on a patient, wherein at least one result from the at least one procedure is inputted into the patient record in a structured format; and analyzing the patient record with a domain knowledge base to determine if the patient has indications of lung cancer.

WO 03/040987 A2

PATIENT DATA MINING FOR LUNG CANCER SCREENING

Cross Reference to Related Applications

This application claims the benefit of U.S. Provisional Application Serial No. 60/335,542, filed on November 2, 2001, which is incorporated by reference herein in its entirety.

Field of the Invention

The present invention relates to medical information processing systems, and, more particularly to a computerized system and method for screening patients for lung cancer, monitoring nodule detection in patients and managing patients exhibiting lung cancer indications.

Background of the Invention

In the United States, lung cancer is the second most common cause of cancer and the leading cause of cancer deaths for both men and women. Survival from lung cancer is dependant on the stage of the cancer. The stage is determined by the size and location of nodules (e.g., tumors), the presence of cancer in the surrounding lymph nodes, and the spread of cancer to distant sites. When lung cancer is treated in its earliest stage, the cure rate approaches 70 % or greater.

Therefore, early detection is crucial for increasing the survival rates for patient with lung cancer.

Traditionally, X-rays have been used to detect nodules in patients showing symptoms of lung cancer. However, the smallest nodule detectable by X-ray is approximately 1 cm, which is an indication of advance growth, and subsequently, survival rates for patients exhibiting these nodules are low. Computerized tomography (CT) scans are capable of detecting lung cancer nodules much smaller than by conventional X-rays. CT scans have a much higher resolution than X-rays and can detect a nodule at only 0.5 mm in diameter.

Although CT scans can detect very small nodules, CT screening is expensive. Determining whether detected nodules are malignant requires multiple CT examinations over several months to make sure that the nodule does not grow. Furthermore, most patients screened have some "junk" in their lungs which may show up as a nodule in a CT scan due to its high resolution. However, every nodule can not be biopsied for several reasons. First, as with multiple CT scans, a biopsy is expensive. Most importantly, a biopsy causes much anxiety in patients due to the fact it is an invasive procedure and it has a certain amount of risk associated with it. Therefore, a protocol for lung cancer screening needs to balance the costs associated with the tests to be performed and the burden

placed upon the patients while maintaining a high quality of care.

In view of the above, there exists a need for improved systems and methods for screening persons for lung cancer, monitoring nodule detection, and managing patients exhibiting lung cancer indications.

Summary of the Invention

According to one aspect of the present invention, a lung cancer screening system is provided including a database including structured patient information for a patient population and a domain knowledge base including information about lung cancer; an individual patient record; and a processor for analyzing the patient record with data from the database to determine if a patient has indications of lung cancer. The database being populated with the structured patient information by data mining structured and unstructured patient records or is populated with information collected prospectively.

In another aspect, the processor further includes a diagnosis module for determining a current state of a patient. The diagnosis module analyzes an imaging study, e.g., a computerized tomography (CT) scan, to detect if a nodule is present and, if a nodule is present, registering the imaging

study (CT scan) with previous imaging studies of the patient to determine growth of the nodule.

In a further aspect of the present invention, the processor further includes a modeling module for analyzing the population-based structured patient information of the database to determine trends in patients with similar characteristics of the patient as determined by the individual patient record. The modeling module predicts a progression of lung cancer in the patient based on a determined trend.

In another aspect of the present invention, the processor further comprises a patient management module for determining a screening protocol for the patient. The patient management module determines an optimal time for a next testing procedure. Additionally, the patient management module balances costs of potential tests to be preformed against a risk of late detection of lung cancer to determine a maximum allowable time between tests for the individual patient.

According to yet another aspect of the present invention, a method for screening for lung cancer is provided. The method includes the steps of inputting patient-specific data into a patient record; performing at least one lung cancer screening procedure on a patient, wherein at least one result from the at least one procedure is inputted into the patient record in a structured format; and analyzing the patient record with a domain patient.

In a further aspect, the method further includes the step of analyzing a database of structured patient information for a patient population to create a model of a similar patient with similar characteristics of the patient based on the patient record; and determining a progression of lung cancer in the patient based on the model. The database of population-based structured patient information is compiled by mining data of population-based patients based on the domain knowledge base, wherein the data is stored in structured and unstructured formats.

In still a further aspect of the present invention, the method includes the step of determining a screening protocol for the patient based on the model. The screening protocol includes an optimal time for a next procedure based on the model.

In another aspect of the present invention, the inputting of patient-specific data into the patient record is performed by mining historical data of the patient, the historical data being in structured and unstructured formats.

In yet another aspect, the performing step includes conducting an imaging study, e.g., a computerized tomography (CT) scan of the patient; and detecting nodules present in the scan. The analyzing step includes registering the CT scan with previous CT scans; and determining growth of the detected nodules over the several scans.

According to a further aspect of the present invention, a program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for screening for lung cancer is provided including the method steps of inputting patient-specific data into a patient record; performing at least one lung cancer screening procedure on a patient, wherein at least one result from the at least one procedure is inputted into the patient record in a structured format; and analyzing the patient record with a domain knowledge base to determine if the patient has indications of lung cancer.

Brief Description of the Drawings

The above and other aspects, features and advantages of the present invention will become more apparent from the following detailed description when taken in conjunction with the accompanying drawings in which:

FIG. 1 is a block diagram of a computer processing system to which the present invention may be applied according to an embodiment of the present invention;

FIG. 2 illustrates an exemplary lung cancer screening system according to an embodiment of the present invention; and

FIG. 3 illustrates a flow diagram for screening, monitoring and managing a patient according to an embodiment of the present invention.

Description of Preferred Embodiments

To facilitate a clear understanding of the present invention, illustrative examples are provided herein which describe certain aspects of the invention. However, it is to be appreciated that these illustrations are not meant to limit the scope of the invention, and are provided herein to illustrate certain concepts associated with the invention.

It is also to be understood that the present invention may be implemented in various forms of hardware, software, firmware, special purpose processors, or a combination thereof. Preferably, the present invention is implemented in software as a program tangibly embodied on a program storage device. The program may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as one or more central processing units (CPU), a random access memory (RAM), and input/output (I/O) interface(s). The computer platform also includes an operating system and microinstruction code. The various processes and functions described herein may either be part of the microinstruction code or part of the program (or combination thereof) which is executed via the operating system. In addition, various other peripheral devices may be connected to the computer platform such as an additional data storage device and a printing device.

It is to be understood that, because some of the constituent system components and method steps depicted in the accompanying figures are preferably implemented in software, the actual connections between the system components (or the process steps) may differ depending upon the manner in which the present invention is programmed.

FIG. 1 is a block diagram of a computer processing system 100 to which the present invention may be applied according to an embodiment of the present invention. The system 100 includes at least one processor (hereinafter processor) 102 operatively coupled to other components via a system bus 104. A read-only memory (ROM) 106, a random access memory (RAM) 108, an I/O interface 110, a network interface 112, and external storage 114 are operatively coupled to the system bus 104. Various peripheral devices such as, for example, a display device, a disk storage device(e.g., a magnetic or optical disk storage device), a keyboard, and a mouse, may be operatively coupled to the system bus 104 by the I/O interface 110 or the network interface 112.

The computer system 100 may be a standalone system or be linked to a network via the network interface 112. The network interface 112 may be a hard-wired interface. However, in various exemplary embodiments, the network interface 112 can include any device suitable to transmit information to and from another device, such as a universal asynchronous receiver/transmitter (UART), a parallel digital interface, a

software interface or any combination of known or later developed software and hardware. The network interface may be linked to various types of networks, including a local area network (LAN), a wide area network (WAN), an intranet, a virtual private network (VPN), and the Internet.

The external storage 114 may be implemented using a database management system (DBMS) managed by the processor 102 and residing on a memory such as a hard disk. However, it should be appreciated that the external storage 114 may be implemented on one or more additional computer systems. For example, the external storage 114 may include a data warehouse system residing on a separate computer system.

Those skilled in the art will appreciate that other alternative computing environments may be used without departing from the spirit and scope of the present invention.

Referring to FIG. 2, an exemplary lung cancer screening system 200 according to an embodiment of the present invention is illustrated. The lung cancer screening system 200 includes processor 202 which includes a plurality of modules for performing different tasks. The processor 202 is coupled to a structured database 206 compiled for a disease of interest, here, lung cancer. The processor 202 will interact with the structured database 206 to determine certain outputs relating to a specific patient based on the specific patient's record 204. The patient record 204 may include demographic

information, family history, results of initial tests, images from a CT scan, doctors' dictations, etc.

Preferably, the structured database 206 is populated with population-based patient information using data mining techniques described in "Patient Data Mining," by Rao et al., copending U.S. Patent Application Serial No. 10/___,___, (Attorney Docket No. 8706-600) filed herewith, which is incorporated by reference herein in its entirety. That patent application teaches a data mining framework for mining high-quality structured clinical information. The data mining framework preferably includes a data miner, having functions and capabilities as in the REMIND system, commercially available from Siemens Medical Solutions, that mines medical information from computerized patient records (CPRs) based on domain-specific knowledge contained in a knowledge base. The CPRs may be of structured and/or unstructured formats. The domain-specific knowledge may relate to a disease of interest, a hospital, etc. The data miner includes components for extracting information from the CPRs, combining all available evidence in a principled fashion over time, and drawing inferences from this combination process. The mined medical information is stored in the structured CPR database, such as database 206.

Here, the domain knowledge base 206-1 relates to lung cancer and the patient database 206-2 is a structured database populated with information mined from a plurality of

computerized patient records (CPRs) wherein the patients either had lung cancer, exhibited symptoms or indications of lung cancer and/or participate in activities which increase their risk of developing lung cancer, e.g., smoking. Alternatively, the information stored in the patient database 206-2 may be prospectively collected.

The lung cancer screening system 200 interacts with the specific patient record 204 and the structured database 206 to determine the patient present condition, determine the future chances of the patient developing lung cancer and determine suggested future treatment of the patient. Each task performed by the lung cancer screening system 200 is performed by an executable module residing either in the processor of the system 202 and/or in a memory device (e.g., RAM, ROM, external storage, etc.) of the system.

A diagnosis module 202-1 interacts with the patient record 204 and the domain knowledge base 206-1 to determine the current state of the patient, e.g., a diagnosis, and any risk assessment. The diagnosis module 202-1 will combine all available information about the patient and perform a probabilistic inference on patient-specific issues based on the domain knowledge base 206-1, using techniques described in "Patient Data Mining for Diagnosis and Projections of Patient States," by Rao et al., copending U.S. Patent Application Serial No. 10/___,___, (Attorney Docket No. 8706- 624) filed herewith, which is incorporated by reference herein in its

entirety. This may entail reviewing an imaging study, e.g., a CT scan, included in the patient record 204 and determining if the CT scan includes a nodule and, if so, determining the size of the nodule.

In one embodiment, the diagnosis module 202-1 may perform volumetric serial studies on several CT scans taken over a period of time. Each CT scan image will be inputted into the screening system and any nodules discovered will be extracted by imaging processes known in the art. Each nodule will be analyzed for shape, size, position and risk. Each successive CT scan will be registered with the previous scan to estimate a growth distribution of the nodules over time. The domain knowledge base 206-1 will then be used to determine, based on the size of the nodule, rate of growth, if the nodule is a cause for concern. For example, if a detected nodule is greater than 1cm, the diagnosis module may determine there is a 20% probability that the screened patient is in Stage I.

Additionally, the domain knowledge base 206-1 will impose criteria on other aspects of the patient record 204 to determine if there is a cause for concern. For example, the system may increase the probability that the patient is in Stage I if it is determined the patient is a heavy smoker. Alternatively, the system may decrease the probability that the patient is in Stage I if the patient record shows that the detected nodule has been the same size for more than two years.

Similarly, nodule information may be extracted from a radiologist's dictations by a natural language processing module. The diagnosis module 202-1 will use the domain knowledge base to reconcile any conflicts between the image processing and the natural language processing. For example, if the natural language processing has extracted from a doctor's dictation that the patient exhibits no indications of lung cancer but the image processing indicates the patient has several nodules which have doubled in size over the last three months, the system will, based on growth rates in the knowledge base, assign a high probability to Stage I lung cancer.

Additionally, the diagnosis module 202-1 may employ the domain knowledge base 206-1 to establish relationships between test values and demographics to other variables.

A modeling module 202-2 interacts with the patient record 204 and the patient database 206-2 to determine the potential progression of the patient or to determine future chances of the disease occurring. The modeling module 202-2 reviews the patient database 206-2 for records of patients with similar characteristics of the current patient as determined by the patient record 204. For example, if the current patient is a male, age forty who smokes five packs of cigarettes a day, the modeling module will extract only those patient records that are with an acceptable range of the current patient's characteristics. The modeling module 202-2 will conduct a

retrospective CPR analysis on the structured database 206 to look for trends of outcomes of these "similar" patients to predict the progression of the lung cancer in the current patient, for example, automatically identify interesting patterns between genetic markers, outcomes, demographics and therapy, for instance, white males, age forty who have been smoking five packs of cigarettes a day for twenty years tend to develop nodules at age 52. All recommendations or prognosis may be shown as statistics on similar patients.

A patient management module 202-3 interacts with the patient record 204, domain knowledge base 206-1 and patient database 206-2 to determine a screening protocol for the patient, e.g., an optimal time for a next procedure and/or an optimal procedure to be next performed. For example, if a detected nodule in a patient has grown 100% in three months, the system will recommend a follow-up CT scan in three months; alternatively, if the nodule exhibits no growth, the system will recommend a follow-up CT scan in twelve months.

Additionally, the patient management module 202-3 will generate a treatment and therapy planning guideline for the patient. When compiling a proposed protocol for the specific patient, the patient management module 202-2 will balance the costs of the potential tests to be performed against a risk of late detection of lung cancer with the model generated above to find a maximum allowable time between tests to ensure quality of care. For example, if a detected nodule has no

growth over the last eighteen months, the system may recommend a follow-up visit to be two years later since no growth indicates a lower risk for developing cancer.

Referring to FIG. 3, a work flow diagram illustrates how the lung cancer screening system can assist an appropriate medical professional in diagnosing, monitoring and managing a patient with lung cancer and/or lung cancer indications.

An asymptomatic patient 302 submits himself to a lung cancer screening procedure 306. Before the procedure takes place, a patient record 304 is created for the patient. It is to be understood that patient record 304 is the same as patient record 204 used in the lung cancer screening system 200 of FIG. 2. The patient record 304 is populated with data from a questionnaire (for example, information such as demographics, family history, smoking history, etc.), results of initial tests, genetic markers, radiologists' clinical findings, etc. Additionally, the patient record may be populated by mining data from the historical records of the patient, as described in the copending application identified above (Attorney Docket No. 8706-600). The lung cancer screening procedure 306 may include any of the following: a physical examination, chest X-ray, CT scan, positron emission tomography (PET) scan, a magnetic resonance imaging (MRI) scan, sputum cytology, bronchoscopy, blood work, pulmonary function tests, etc.

The results of the lung cancer screening procedures are inputted to the system and stored in a structure format in the patient record 304. Before being stored in the patient record, any CT scan 308 performed is processed for nodule detection and management via volumetrics serial studies 310 performed by the diagnosis module 202-1. For an initial CT scan, the scan is analyzed to determine if any nodules of interest are present via the diagnosis module 202-1 of the lung cancer screening system 200. If any nodules are present, their position, size, and/or features are determined and are stored in the patient record. The results from the volumetrics study of the nodules are combined in a report 312 with the results of the other procedures 314 to be presented to the appropriate medical professional 316, for example, a radiologist or oncologist. The report may include a probabilistic determination of the current state of the patient, e.g., there is at least a 20% chance of malignant lung cancer.

Based on the report 312, the appropriate medical professional will make a determination of the current state of the patient. This determination may be made solely on the report generated by the lung cancer screening system. Alternatively, the determination may be made by comparing the results in the report to a model of a similar patient generated by the modeling module 202-2 of the system 200. If cancer is suspected, the patient will be sent for a further diagnostic workflow, for example, a biopsy.

If cancer is not suspected at this time, the medical professional's determinations and any further comments are added to the patient record 304. If any suspicious nodules are detected, the patient record is reviewed for patient management 318 to determine an appropriate screening protocol, i.e., the timing of the next visit, what tests are to be performed next, etc., via the patient management module 202-3 of the lung cancer screening system 200. The screening protocol may be determined for the specific patient by querying the structured database 206 with the patient record 304 to find similar patients and to determine their outcomes in accordance with their proscribed treatments and/or therapies. The screening protocol for the specific patient may entail determining the optimal time for the next CT scan, blood test, etc. and/or identifying all potential lung cancer incidents before an adverse event occurs, e.g., stage II cancer or metastases.

The lung cancer screening system and method of the present invention provides for a rapid review of a large volume data set. The system and method allows all available information to be used in diagnosing, monitoring and managing patients with lung cancer and/or patients exhibiting indications of lung cancer. By accessing all available information of a specific patient and a plurality of patients, the system and method will assess the need for a patient to get diagnostic workup or remain in screening, estimate when

the patient should return for follow-up study and project lung cancer risk into the future.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be affected therein by one skilled in the art without departing from the scope or spirit of the invention.

What Is Claimed Is:

1. A lung cancer screening system comprising:

a database including structured patient information for a patient population and a domain knowledge base including information about lung cancer;

an individual patient record; and

a processor for analyzing the patient record with data from the database to determine if a patient has indications of lung cancer.

2. The system as in claim 1, wherein the database is populated with the structured patient information by data mining structured and unstructured patient records.

3. The system as in claim 1, wherein the database is populated with information collected prospectively.

4. The system as in claim 1, wherein the processor further comprising a diagnosis module for determining a current state of a patient.

5. The system as in claim 4, wherein the diagnosis module analyzes an imaging study to detect if a nodule is present and, if a nodule is present, registering the imaging study with previous imaging studies of the patient to determine growth of the nodule.

6. The system as in claim 5, wherein the imaging study is a computed tomography (CT) scan.

7. The system as in claim 1, wherein the processor further comprising a modeling module for analyzing the population-based structured patient information of the database to determine trends in patients with similar characteristics of the patient as determined by the individual patient record.

8. The system as in claim 1, wherein the individual patient record is created by mining structured and unstructured patient information.

9. The system as in claim 7, wherein the modeling module predicts a progression of lung cancer in the patient based on a determined trend.

10. The system as in claim 1, wherein the processor further comprises a patient management module for determining a screening protocol for the patient.

11. The system as in claim 10, wherein the patient management module determines a time for a next testing procedure for the individual patient.

12. The system as in claim 11, wherein the next testing procedure is a computerized tomography (CT) scan.

13. The system as in claim 8, wherein the patient management module determines a testing procedure to be next performed for the individual patient and determines a time for the testing procedure for the individual patient.

14. The system as in claim 11, wherein the patient management module balances costs of potential tests to be performed against a risk of late detection of lung cancer to determine a

maximum allowable time between tests for the individual patient.

15. A method for screening for lung cancer, the method comprising the steps of:

inputting patient-specific data into a patient record;

performing at least one lung cancer screening procedure on a patient, wherein at least one result from the at least one procedure is inputted into the patient record in a structured format; and

analyzing the patient record with a domain knowledge base to determine if the patient has indications of lung cancer.

16. The method as in claim 15, further comprising the step of diagnosing a current state of the patient.

17. The method as in claim 15, further comprising the steps of

analyzing a database of structured patient information for a patient population to create a model of a similar patient with similar characteristics of the patient based on the patient record; and

determining a progression of lung cancer in the patient based on the model.

18. The method as in claim 17, wherein the database of population-based structured patient information is compiled by mining data of population-based patients based on the domain knowledge base, wherein the data is stored in structured and unstructured formats.

19. The method as in claim 17, further comprising the step of determining a screening protocol for the patient based on the model.

20. The method as in claim 19, wherein the screening protocol includes a time for a next procedure for the individual patient based on the model.

21. The method as in claim 20, wherein the next testing procedure is a computerized tomography (CT) scan.

22. The method as in claim 19, further comprising the steps of determining a testing procedure to be next performed for the individual patient and determining a time for the testing procedure for the individual patient.

23. The method as in claim 20, further comprising the step of balancing costs of potential tests to be performed against a risk of late detection of lung cancer to determine a maximum allowable time between tests for the individual patient.

24. The method as in claim 15, wherein the inputting of patient-specific data into the patient record is performed by mining historical data of the patient, the historical data being in structured and unstructured formats.

25. The method as in claim 15, wherein the performing step includes

conducting an imaging study of the patient; and

detecting nodules present in the imaging study.

26. The method as in claim 25, wherein the analyzing step includes

registering the imaging study with previous imaging studies; and

determining growth of the detected nodules over the imaging studies.

27. The method as in claim 25, wherein the imaging study is a computerized tomography (CT) scan.

28. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for screening for lung cancer, the method steps comprising:

inputting patient-specific data into a patient record;

performing at least one lung cancer screening procedure on a patient, wherein at least one result from the at least one procedure is inputted into the patient record in a structured format; and

analyzing the patient record with a domain knowledge base to determine if the patient has indications of lung cancer.

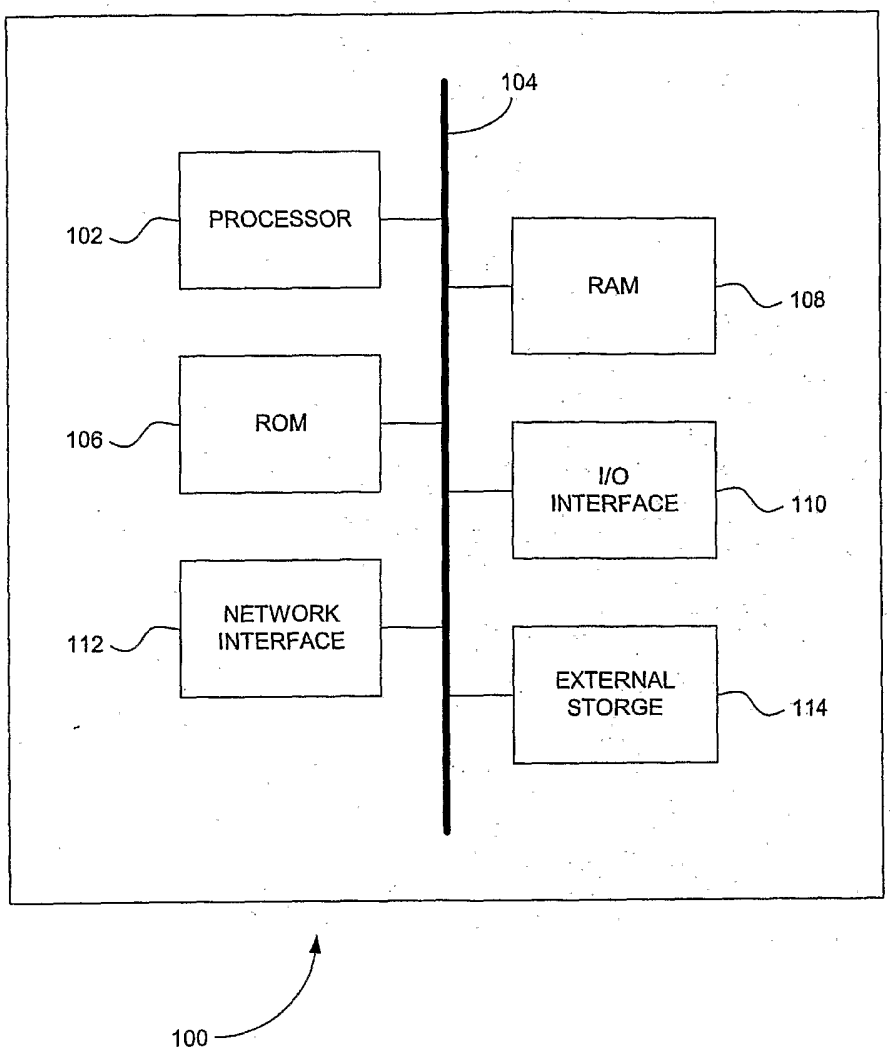

104
PROCESSOR
102
RAM
108
ROM
106
I/O
INTERFACE
110
NETWORK
INTERFACE
112
EXTERNAL
STORGE
114
100

FIG. 1

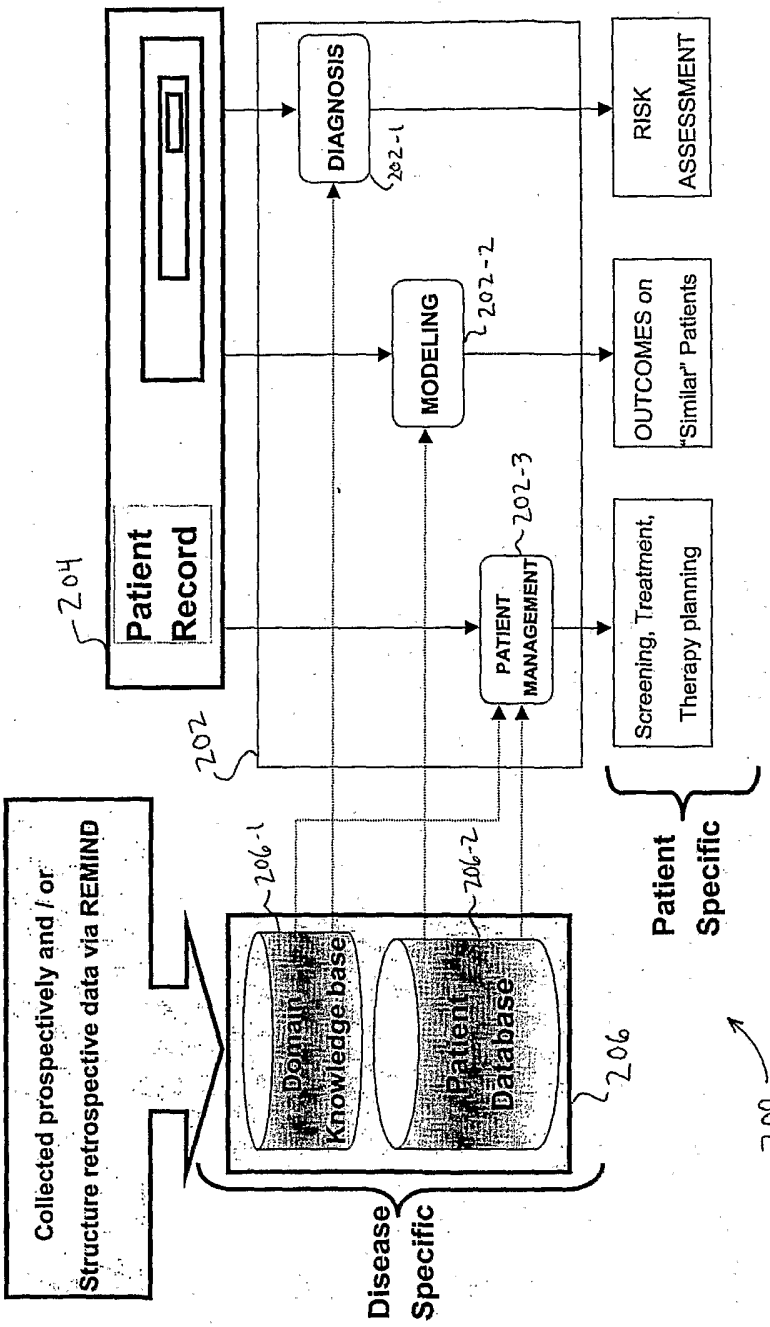

Collected prospectively and / or
Structure retrospective data via REMIND
Domain Knowledge base
Patient Database
206-1
206-2
206
Disease Specific
204
Patient Record
202
DIAGNOSIS
202-1
MODELING
202-2
PATIENT MANAGEMENT
202-3
RISK ASSESSMENT
OUTCOMES on "Similar" Patients
Screening, Treatment, Therapy planning
Patient Specific
200

FIG. 2

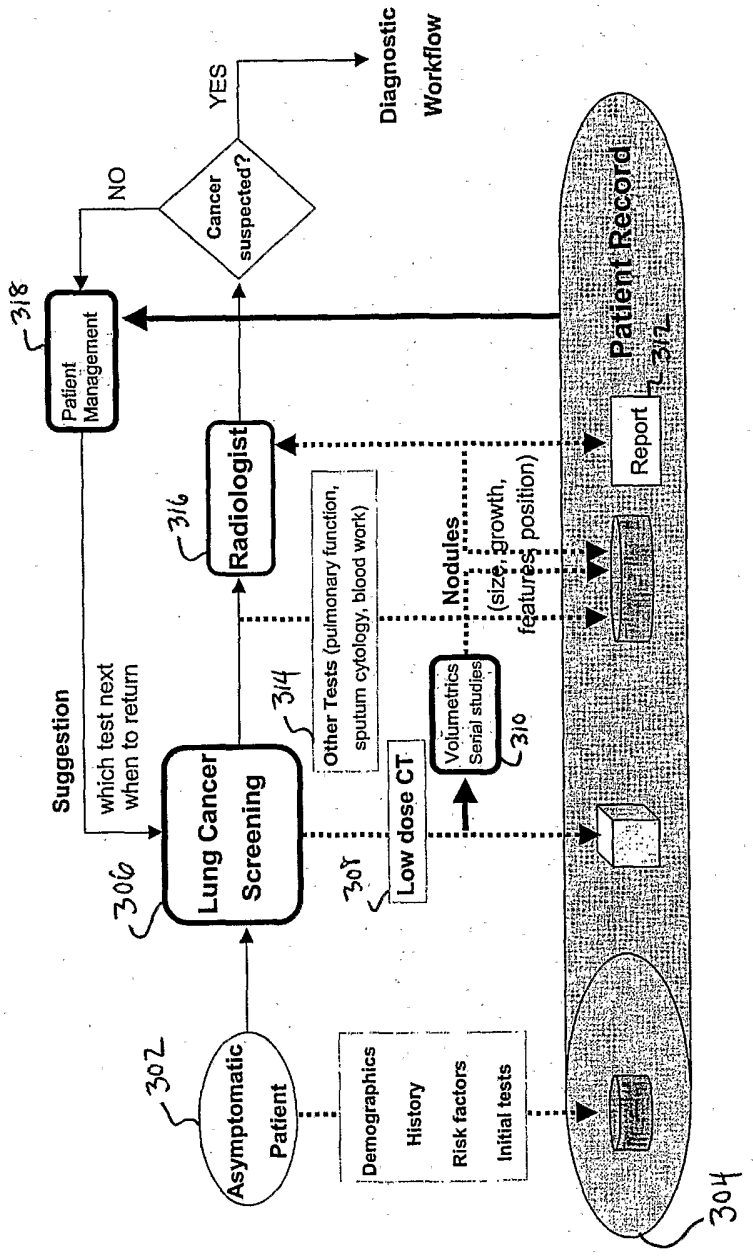

Asymptomatic Patient
302
Demographics
History
Risk factors
Initial tests
Lung Cancer Screening
306
Low dose CT
308
Other Tests (pulmonary function, sputum cytology, blood work)
314
Volumetrics
Serial studies
310
Nodules
(size, growth, features, position)
Radiologist
316
Cancer suspected?
NO
YES
Diagnostic Workflow
Patient Management
318
Suggestion
which test next
when to return
Patient Record
312
Report
304

FIG. 3