

Thus, it is a problem how to objectively accurately and efficiently screen a subject with a high risk in nasopharyngeal cancer for preventing or treating as early as possible.

SUMMARY OF THE INVENTION

5 The present disclosure provides an electronic apparatus, a nasopharyngeal cancer screening and analyzing method, aiming to objectively accurately and efficiently screen a subject with a high risk in nasopharyngeal cancer, for preventing or treating as early as possible.

10 The present disclosure firstly provides an electronic apparatus, the electronic device includes a storage medium and a processor. The storage medium stores a nasopharyngeal cancer screening and analyzing system executed on the processor. The nasopharyngeal cancer screening and analyzing system being executed by the processor includes the following steps:

15 Acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

 Acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed;

20 Extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

 Acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule;

Inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

5 Preferably, the feature tag includes basic feature information, preference habit information, behavior information, and/or social relationship information.

The present disclosure secondly provides a nasopharyngeal cancer screening and analyzing method, the method includes the following steps:

Acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

10 Acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed;

Extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

15 Acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule;

Inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

20 The present disclosure thirdly provides a computer readable medium. The computer readable medium stores a nasopharyngeal cancer screening and analyzing system. The nasopharyngeal cancer screening and analyzing system is executed by a processor for implementing the following steps by the processor:

Acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

Acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature
5 tags from the nasopharyngeal cancer screening request is failed;

Extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

Acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule;

10 Inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is is a subject with a high risk in nasopharyngeal cancer.

The present disclosure acquires feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-
15 screened client is received. In detail, client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request are acquired. When the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed, various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client are extracted. The
20 feature tags are acquired by extracting from the acquired various types feature data based on a preset feature tag extraction rule. The acquired feature tags of the to-be-screened client are inputted into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal

cancer. Thus, the subject with a high risk in nasopharyngeal cancer are objectively accurately and efficiently screened for preventing or treating as early as possible.

BRIEF DESCRIPTION OF THE FIGURES

Implementations of the present disclosure will now be described, by way of
5 example only, with reference to the attached figures, wherein:

FIG. 1 is a flowchart of an embodiment of a nasopharyngeal cancer screening and analyzing method.

FIG. 2 is a flowchart of a training process of a analyzing model in the nasopharyngeal cancer screening and analyzing method.

10 FIG. 3 is a diagram view of an embodiment of a running environment of a nasopharyngeal cancer screening and analyzing system.

FIG. 4 is a diagram view of an embodiment of a nasopharyngeal cancer screening and analyzing system.

DETAILED DESCRIPTION

15 In order to making the technical problems, the technical solutions, and beneficial effects to be more clearer and more understand, the present disclosure is further described in detail with reference to the accompanying drawings and the embodiments. It will be understood that the specific embodiments described herein are merely used for describing the present disclosure, but are not intended to limit the
20 present disclosure.

It needs to be explained that the descriptions related to “first”, “second” and so on in the present disclosure are only description purpose, but are not being understood as indicating or implying a relative significance of itself or implying an amount of the indicated technical feature. Thus, the features limited with “first” or

“second” are indicated or implied including at least one feature. Besides, the technical solutions between different embodiments can be mutual combination, but only based on being achieved by those of an ordinary skill in the art, it should be recognized that the technical solution is not existed when the combination of the technical solutions being mutual contradiction or being unable to be achieved, and it is also not within the protecting range of the present disclosure.

The present disclosure provides a nasopharyngeal cancer screening and analyzing method.

As shown in FIG. 1, a flowchart of a nasopharyngeal cancer screening and analyzing method is provided by the embodiment of the present disclosure.

In one embodiment, the nasopharyngeal cancer screening and analyzing method includes the following steps:

In step S10, acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

The nasopharyngeal cancer screening request includes client attribute data (such as, identification number, or, name and identification number) and/or feature tags of the to-be-screened client. The feature tag includes basic feature information (such as, an involved city belongs to northwest China, a dietary habit of an involved city is spicy, PM2.5 (fine particulate matter) in an involved city often exceeds a specified standard, there are more severe weather in an involved city, a job is in hazardous gas, and so on), preference habit information (such as, an preference of sleep late, an preference of drinking, an preference of barbecue food, an preference of outdoor sports, and so on), behavior information (such as, more frequency in

games, more frequency in overtime, more frequency in takeaway, more frequency in medical serve, and so on), and/or social relationship information (such as, spinsterhood, solitary, infrequency contact with friends). A screening server firstly determines that whether the nasopharyngeal cancer screening request includes
5 feature tags of the client when receiving the nasopharyngeal cancer screening request, and acquires the feature tags of the client.

Step S20, acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed.

10 If there is no feature tag in the nasopharyngeal cancer screening request, the feature tags cannot be acquired from the nasopharyngeal cancer screening request (the operation of acquiring the feature tags is failed), then, the client attribute data are acquired from the nasopharyngeal cancer screening request for acquiring the feature tags based on the client attribute data of the to-be-screened client.

15 Step S30, extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client.

The screening server communicates with several preset business servers (such as, bank servers, medical servers, insurance servers, instant message (IM) servers,
20 game servers, weather servers, takeaway servers, and/or resumes server). After acquiring the client attribute data, the screening server extracts various feature data from the preset business servers corresponding to the client attribute data of the to-be-screened client (such as, bank loan amount and repayment status information, patient record information (for example, frequency of medical service, disease

category, lasting time of each disease in a predetermined time period), insurance information (for example, involved industry, sex, age, marital status, profession, and so on), usage information of IM account (for example, loading time information of the IM in each day, on-line time in each day), game information (for example, loading time in each day, on-line game time in each day), weather information (days of PM2.5 exceeding the standard in three years), takeaway information (for example, time information of ordering takeaway in each day, takeaway types of the ordered takeaway in each day), resumes information (for example, interests and hobbies, personality, working experience)).

10 Step S40, acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule.

 The screening server stores the preset feature tag extraction rule . Using the preset feature tag extraction rule, feature tags of the extracted various feature data are analyzed after extracting the various feature data of the to-be-screened client, thus
15 the feature tags of the to-be-screened client are obtained.

 Step S50, inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

 The screening server includes a pre-trained analyzing model. The analyzing
20 model is trained by a large volumes of known client data (including feature tags corresponding to the client with or without nasopharyngeal cancer). The acquired feature tags are inputted into the pre-trained analyzing model after the feature tags of the to-be-screened client are acquired by the screening server. The analyzing model

analyzes whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer based on the inputted various feature tags.

The present disclosure acquires feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received. In detail, client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request are acquired. When the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed, various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client are extracted. The feature tags are acquired by extracting from the acquired various types feature data based on a preset feature tag extraction rule. After acquiring the feature tags, the acquired feature tags of the to-be-screened client are inputted into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer. In one embodiment, the feature tags presenting the feature data of the to-be-screened client are inputted into the pre-trained analyzing model for analyzing whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer. The subject with a high risk in nasopharyngeal cancer are objectively accurately and efficiently screened for preventing or treating as early as possible.

As shown in FIG. 2, FIG. 2 shows a flowchart of a training of the analyzing model in the nasopharyngeal cancer screening and analyzing method.

In one embodiment, the analyzing model is a multiple linear regression model. A training processing including:

Step E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data.

The client attribute data can be identification number, or, name and identification number. The screening server communicates with several preset
5 business servers (such as, bank servers, medical servers, insurance servers, IM servers, game servers, weather servers, takeaway servers, and/or resumes server). There are mapping relation of the client attribute data and the feature data in each preset business server. The screening server selects a first preset number of the client (such as 500,000). Then, due to each selected client, feature data corresponding to
10 the client attribute data are respectively extracted from each preset business server according to the client attribute data, thus the selected first preset number of the various feature data of the client are obtained.

Step E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client.

15 The screening server presets the feature tag extraction rule. After extracting the first preset number of the various feature data, the various feature data of each client are respectively analyzed, the feature tags of each client are obtained.

Step E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the
20 nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model.

A database in the screening server records patient records of each client. The screening server searches the database based on the client attribute data to confirm the abnormal clients with the nasopharyngeal cancer and the normal clients without nasopharyngeal cancer in the first preset number of the clients. After confirming the abnormal clients and the normal clients in the first preset number of the clients, each feature tag corresponding to each normal client is set as a training sample, and each feature tag corresponding to each abnormal client is set as another training sample.

Step E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage. A sum of the first percentage and the second percentage is less than or equal to 100%.

The first percentage of the training samples in the all the training samples are set as the training set, and the second percentage of the training samples in all the training samples are set as the validation set. For example, the first percentage is 65%, and the second percentage is 35%. Certainly, each of the training set and the validation set includes some training samples corresponding to the normal clients and some training samples corresponding to the abnormal clients.

Step E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training.

Firstly, the analyzing model are trained by the training set, and the accuracy rate of the analyzing result of the analyzing model is verified by the validation set after accomplishing training.

Step E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold.

The screening server presets a preset accuracy threshold (equal to the preset threshold, such as 98%). If the accuracy rate of the analyzing result of the analyzing model based on the validation set is larger than the accuracy rate, it means that the training of the analyzing model meets an expected standard, the screening server confirms the training of the analyzing model is accomplished.

Step E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

If the accuracy rate of the analyzing result of the analyzing model is less than or equal to the accuracy threshold, it means that the training of the analyzing model do not meet the expected standard, the screening server repeats the steps E1, E2, and E3, increases a value of the first preset number (such as increasing 50,000 each time), re-executes the steps E4 and E5 based on the increased training samples, until the requirement of the step E6 are reached.

The preset feature tag extraction rule in the nasopharyngeal cancer screening and analyzing method of the present disclosure includes:

Tag thresholds corresponding to consecutive values of the various feature data are set.

For example, the feature threshold corresponding to the number of days, when the PM2.5 in the involved city exceeds the specified standard in each year for recent three years, is 60 days. When the number of days, when the PM2.5 in the involved city exceeds the specified standard in each year for recent three years, is

larger than the feature threshold, it represents that the PM2.5 in the involved city often exceeds. The feature threshold corresponding to the number of days, when the weather warning level is above a blue warning in the involved city in each year for recent three years, is 55 days. When the number of days, when the weather warning level is above the blue warning in the involved city in each year for recent three years, is larger than the feature threshold, it represents that there are more bad weathers in the involved city. The feature threshold corresponding to the number of days, when the sleeping time is later than 23:00 in the latest recent year, is 100 days. When the number of days, when the sleeping time is later than 23:00 in the latest recent year, is larger than the feature threshold, it represents that the preference is sleep late. The feature threshold corresponding to the number of the times for ordering barbecue takeaway in the latest year is 80 times. When the number of the times for ordering barbecue takeaway in the latest year is larger than the feature threshold, it represents that the preference is barbecue. The feature threshold corresponding to the number of the medical visiting in the latest year is 30 times. When the number of the medical visiting in the latest year is larger than the threshold, it represents that there are more medical visiting times.

Tag ranges corresponding to discrete values of the various feature data are set.

For example, the tag range corresponding to the involved city include a group of cities in northwest China, a group of cities in central China, a group of cities in southern China, and so on. When the involved city is in the group of cities in northwest China, it represents that the tag information corresponding to the involved city is that it belongs to northwest China. The tag range corresponding to the dietary

habit includes a group of spicy cities, a group of greasy cities, a group of light cities, a group of sweet/slat cities, and so on. When the involved city is in the group of spicy cities, it represents that the tag information is that the dietary habit is spicy. The tag range corresponding to the job of the client includes a group of the jobs in harmful gases, a group of the jobs in nonvolatile harmful gases, a group of the jobs in volatile harmful gases, a group of the jobs in harmless gases, and so on. When the job of the client is a job in the group of the jobs in harmful gases, it represents that the tag information corresponding to the job of the client is that the job is a job in harmful gases.

10 The tag information corresponding to the feature data of the various consecutive values of each client is confirmed based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and the tag information corresponding to the feature data of the various discrete values of each client is confirmed based on the mapping relationship
15 between the various feature data category of the discrete values and the tag ranges.

Besides, the present disclosure also provides a nasopharyngeal cancer screening and analyzing system.

Referring to FIG. 3, it is an operating environment of the nasopharyngeal cancer screening and analyzing system 10.

20 In one embodiment, the nasopharyngeal cancer screening and analyzing system 10 is installed and executed in an electronic apparatus 1. The electronic apparatus 1 can be a computer of a desktop computer, a notebook computer, a palmtop computer, or a server. The electronic apparatus 1 can include a storage medium 11, a processor 12, and a display 13, but not being limited thereto. FIG. 3

only shows the components in the electronic apparatus 1, it is understood that the embodiment do not show all the components, there can be more or less components in other embodiments.

In some embodiments, the storage medium 11 can be an internal storage unit
5 of the electronic apparatus 1, such as a hard disk or a memory. In other
embodiments, the storage medium 11 can be an external storage device of the
electronic apparatus 1, such as a plug-in hard disk in the electronic apparatus 1, a
smart media card (SMC), a secure digital (SD), a flash card, and so on. Further, the
storage medium 11 can includes the internal storage unit and the external storage
10 device of the electronic apparatus 1. The storage medium 11 stores various types of
softwares of the electronic device 1 and various data, such as program codes of the
nasopharyngeal cancer screening and analyzing system 10, and so on. The storage
medium 11 also can provisionally store the data being outputted or to-be outputted.

In some embodiments, the processor 12 can be a central processing unit
15 (CPU), a micro processor, or any other data processing chip. The processor 12 is
used for running the program code stored in the storage medium 11 or processing the
data, such as a nasopharyngeal cancer screening and analyzing system 10, and so on.

In some embodiments, the display 13 can be a LED display, a LCD display, a
LCD touch display, an touchable organic light emitting diode (OLED) touch display,
20 and so on. The display 13 is used for displaying processing information in the
electronic apparatus 1 and visible user interface. The components 11-13 in the
electronic apparatus 1 communicates with each other through a system bus.

Referring to FIG. 4, it is a diagram view of an embodiment of the
nasopharyngeal cancer screening and analyzing system 10. In the embodiment, the

nasopharyngeal cancer screening and analyzing system 10 can be divided into one or more modules stored in the storage medium 11, which can be executed by the one or more processor (the processor 12 in the embodiment) for implementing the present disclosure. For example, in FIG. 4, the nasopharyngeal cancer screening and
5 analyzing system 10 can be divided into a first acquiring module 101, a second acquiring module 1020, an extraction module 103, a first analyzing module 104, and a second analyzing module 105. The module in the present disclosure is a series of instruction segments being capable of implementing a specified function, which is better for describing an executing processing of the nasopharyngeal cancer screening
10 and analyzing system 10 in the electronic apparatus 1 by comparing with programs.

The first acquiring module 101 acquires feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received.

The nasopharyngeal cancer screening request includes client attribute data
15 (such as, identification number, or, name and identification number) and/or feature tags of the to-be-screened client. The feature tags include basic information (such as, an involved city belongs to northwest China, a dietary habit of an involved city is spicy, PM2.5 (fine particulate matter) in an involved city often exceeds a specified standard, there are more severe weather in an involved city, a job is in hazardous gas,
20 and so on), preference habit information (such as, an preference of sleep late, an preference of drinking, an preference of barbecue food, an preference of outdoor sports, and so on), behavior information (such as, more frequency in games, more frequency in overtime, more frequency in takeaway, more frequency in medical serve, and so on), and/or social relationship information (such as, spinsterhood,

solitary, infrequency contact with friends). A screening server firstly determines that whether the nasopharyngeal cancer screening request includes feature tags of the client when receiving the nasopharyngeal cancer screening request, and acquires the feature tags of the client.

5 The second acquiring module 102 acquires client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed.

10 If there is no feature tag in the nasopharyngeal cancer screening request, the feature tags cannot be acquired from the nasopharyngeal cancer screening request (the operation of acquiring the feature tags is failed), then, the client attribute data are acquired from the nasopharyngeal cancer screening request for acquiring the feature tags based on the client attribute data of the to-be-screened client.

15 The extraction module 103 extracts various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client.

20 The screening server communicates with several preset business servers (such as, bank servers, medical servers, insurance servers, instant message (IM) servers, game servers, weather servers, takeaway servers, and/or resumes server). After acquiring the client attribute data, the screening server extracts various feature data from the preset business servers corresponding to the client attribute data of the to-be-screened client (such as, bank loan amount and repayment status information, patient record information (for example, frequency of medical service, disease category, lasting time of each disease in a predetermined time period), insurance

information (for example, involved industry, sex, age, marital status, profession, and so on), usage information of IM account (for example, loading time information of the IM in each day, on-line time in each day), game information (for example, loading time in each day, on-line game time in each day), weather information (days of PM2.5 exceeding the standard in three years), takeaway information (for example, time information of ordering takeaway in each day, takeaway types of the ordered takeaway in each day), resumes information (for example, interests and hobbies, personality, working experience)).

The first analyzing module 104 acquires feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule.

The screening server stores the preset feature tag extraction rule. Using the preset feature tag extraction rule, feature tags of the extracted various feature data are analyzed after extracting the various feature data of the to-be-screened client, thus the feature tags of the to-be-screened client are obtained.

The second analyzing module 105 inputs the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

The screening server includes a pre-trained analyzing model. The analyzing model is trained by a large volumes of known client data (including feature tags corresponding to the client with or without nasopharyngeal cancer). The acquired feature tags are inputted into the pre-trained analyzing model after the feature tags of the to-be-screened client are acquired by the screening server. The analyzing model

analyzes whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer based on the inputted various feature tags.

The present disclosure acquires feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received. In detail, client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request are acquired. When the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed, various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client are extracted. The feature tags are acquired by extracting from the acquired various types feature data based on a preset feature tag extraction rule. After acquiring the feature tags, the acquired feature tags of the to-be-screened client are inputted into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer. In one embodiment, the feature tags presenting the feature data of the to-be-screened client are inputted into the pre-trained analyzing model for analyzing whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer. The subject with a high risk in nasopharyngeal cancer are objectively accurately and efficiently screened for preventing or treating as early as possible.

In one embodiment, the analyzing model is a multiple linear regression model. A training processing including (as shown in FIG. 2):

Step E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data.

The client attribute data can be identification number, or, name and identification number. The screening server communicates with several preset business servers (such as, bank servers, medical servers, insurance servers, IM servers, game servers, weather servers, takeaway servers, and/or resumes server).

5 There are mapping relation of the client attribute data and the feature data in each preset business server. The screening server selects a first preset number of the client (such as 500,000). Then, due to each selected client, feature data corresponding to the client attribute data are respectively extracted from each preset business server according to the client attribute data, thus the selected first preset number of the

10 various feature data of the client are obtained.

Step E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client.

The screening server presets the feature tag extraction rule. After extracting the first preset number of the various feature data, the various feature data of each

15 client are respectively analyzed, the feature tag of each client are obtained.

Step E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to

20 each normal client and each abnormal client as a training sample of the analyzing model.

A database in the screening server records patient records of each client. The screening server searches the database based on the client attribute data to confirm the abnormal clients with the nasopharyngeal cancer and the normal clients without

nasopharyngeal cancer in the first preset number of the clients. After confirming the abnormal clients and the normal clients in the first preset number of the clients, each feature tag corresponding to each normal client is set as a training sample, and each feature tag corresponding to each abnormal client is set as another training sample.

- 5 Step E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage. A sum of the first percentage and the second percentage is less than or equal to 100%.

 The first percentage of the training samples in the all the training samples are set as the training set, and the second percentage of the training samples in all the
10 training samples are set as the validation set. For example, the first percentage is 65%, and the second percentage is 35%. Certainly, each of the training set and the validation set includes some training samples corresponding to the normal clients and some training samples corresponding to the abnormal clients.

- Step E5, training the analyzing model with the feature tags of each normal
15 client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training.

 Firstly, the analyzing model are trained by the training set, and the accuracy rate of the analyzing result of the analyzing model is verified by the validation set
20 after accomplishing training.

 Step E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold.

 The screening server presets a preset accuracy threshold (equal to the preset threshold, such as 98%). If the accuracy rate of the analyzing result of the analyzing

model based on the validation set is larger than the accuracy rate, it means that the training of the analyzing model meets an expected standard, the screening server confirms the training of the analyzing model is accomplished.

Step E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

If the accuracy rate of the analyzing result of the analyzing model is less than or equal to the accuracy threshold, it means that the training of the analyzing model do not meet the expected standard, the screening server repeats the steps E1, E2, and E3, increases a value of the first preset number (such as increasing 50,000 each time), re-executes the steps E4 and E5 based on the increased training samples, until the requirement of the step E6 are reached.

The preset feature tag extraction rule in the nasopharyngeal cancer screening and analyzing method of the present disclosure includes:

Tag thresholds corresponding to consecutive values of the various feature data are set.

For example, the feature threshold corresponding to the number of days, when the PM2.5 in the involved city exceeds the specified standard in each year for recent three years, is 60 days. When the number of days, when the PM2.5 in the involved city exceeds the specified standard in each year for recent three years, is larger than the feature threshold, it represents that the PM2.5 in the involved city often exceeds. The feature threshold corresponding to the number of days, when the weather warning level is above a blue warning in the involved city in each year for recent three years, is 55 days. When the number of days, when the weather warning

level is above the blue warning in the involved city in each year for recent three years, is larger than the feature threshold, it represents that there are more bad weather in the involved city. The feature threshold corresponding to the number of days, when the sleeping time is later than 23:00 in the latest recent year, is 100 days.

5 When the number of days, when the sleeping time is later than 23:00 in the latest recent year, is larger than the feature threshold, it represents that the preference is sleep late. The feature threshold corresponding to the number of the times for ordering barbecue takeaway in the latest year is 80 times. When the number of the times for ordering barbecue takeaway in the latest year is larger than the feature

10 threshold, it represents that the preference is barbecue. The feature threshold corresponding to the number of the medical visiting in the latest year is 30 times. When the number of the medical visiting in the latest year is larger than the threshold, it represents that there are more medical visiting times.

Tag ranges corresponding to discrete values of the various feature data are

15 set.

For example, the tag range corresponding to the involved city include a group of cities in northwest China, a group of cities in central China, a group of cities in southern China, and so on. When the involved city is in the group of cities in northwest China, it represents that the tag information corresponding to the involved

20 city is that it belongs to northwest China. The tag range corresponding to the dietary habit includes a group of spicy cities, a group of greasy cities, a group of light cities, a group of sweet/slat cities, and so on. When the involved city is in the group of spicy cities, it represents that the tag information is that the dietary habit is spicy. The tag range corresponding to the job of the client includes a group of the jobs in

harmful gases, a group of the jobs in nonvolatile harmful gases, a group of the jobs in volatile harmful gases, a group of the jobs in harmless gases, and so on. When the job of the client is a job in the group of the jobs in harmful gases, it represents that the tag information corresponding to the job of the client is that the job is a job in
5 harmful gases.

The tag information corresponding to the feature data of the various consecutive values of each client is confirmed based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and the tag information corresponding to the feature data of the various
10 discrete values of each client is confirmed based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

Further, the present disclosure further provides a computer readable medium. The computer readable medium stores a nasopharyngeal cancer screening and analyzing system. The nasopharyngeal cancer screening and analyzing system can
15 be implemented by at least one processor for executing the nasopharyngeal cancer screening and analyzing method in any of the above embodiments by the at least one processor.

The foregoing implementations are merely preferably embodiments of the present disclosure, and are not intended to limit the protection scope of the present
20 disclosure. Any equivalent structure variation using the present disclosure and drawings, being directly or indirectly used in other related technical fields shall all fall into the protection scope of the present disclosure.

CLAIMS

What is claimed is:

1. An electronic apparatus, wherein the electronic apparatus comprises a storage medium and a processor; the storage medium stores a nasopharyngeal cancer screening and analyzing system executed in the processor; the nasopharyngeal cancer screening and analyzing is executed by the processor for performing following steps:

acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed;

extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule; and

inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

2. The electronic apparatus of claim 1, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data;

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

3. The electronic apparatus of claim 1, wherein the analyzing model is a multiple linear regression model; and the steps of training the analyzing model comprises:

E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data;

E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client;

E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model;

E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage; a sum of the first percentage and the second percentage is less than or equal to 100%;

E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training;

E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold; and

E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

4. The electronic apparatus of claim 3, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data; and

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

5. The electronic apparatus of claim 1, wherein the feature tags comprises basic feature information, preference habit information, behavior information, and/or social relationship information.

6. The electronic apparatus of claim 5, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data;

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

7. The electronic apparatus of claim 5, wherein the analyzing model is a multiple linear regression model; the steps of training the analyzing model comprises:

E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data;

E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client;

E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model;

E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage; a sum of the first percentage and the second percentage is less than or equal to 100%

E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training;

E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold; and

E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

8. The electronic apparatus of claim 5, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data;

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

9. A nasopharyngeal cancer screening and analyzing method comprising the following steps:

acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed;

extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule; and

inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

10. The nasopharyngeal cancer screening and analyzing method of claim 9, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data;

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

11. The nasopharyngeal cancer screening and analyzing method of claim 9, wherein the analyzing model is a multiple linear regression model; the steps of training the analyzing model comprises:

E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data;

E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client;

E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model;

E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage; a sum of the first percentage and the second percentage is less than or equal to 100%;

E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training;

E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold; and

E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

12. The nasopharyngeal cancer screening and analyzing method of claim 11, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data; and

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

13. The nasopharyngeal cancer screening and analyzing method of claim 9, wherein the feature tags comprises basic feature information, preference habit information, behavior information, and/or social relationship information.

14. The nasopharyngeal cancer screening and analyzing method of claim 13, wherein the preset feature tag extraction rule comprises:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data;

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

15. The nasopharyngeal cancer screening and analyzing method of claim 13, wherein the analyzing model is a multiple linear regression model; the steps of training the analyzing model comprising:

E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data;

E2, analyzing the various extracted feature data of each client according to the preset feature tag extraction rule for confirming the feature tags of each client;

E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model;

E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage; a sum of the first percentage and the second percentage is less than or equal to 100%;

E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training;

E6, ending the training operation of the analyzing model is finished when the accuracy rate is larger than a preset threshold; and

E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

16. The nasopharyngeal cancer screening and analyzing method of claim 15, wherein the preset feature tag extraction rule comprises:

- setting tag thresholds corresponding to consecutive values of the various feature data;

- setting tag ranges corresponding to discrete values of the various feature data;

- confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.

17. A computer readable medium, the computer readable medium stores a nasopharyngeal cancer screening and analyzing system; the nasopharyngeal cancer screening and analyzing system is executed by at least one processor for performing following steps:

- acquiring feature tags from a nasopharyngeal cancer screening request, when the nasopharyngeal cancer screening request from a to-be-screened client is received;

- acquiring client attribute data of the to-be-screened client from the nasopharyngeal cancer screening request, when the operation of acquiring the feature tags from the nasopharyngeal cancer screening request is failed;

- extracting various types feature data from several preset business servers corresponding to the acquired client attribute data of the to-be-screened client;

acquiring feature tags of the to-be-screened client by analyzing the acquired various types feature data based on a preset feature tag extraction rule; and

inputting the feature tags of the to-be-screened client into a pre-trained analyzing model for determining whether the to-be-screened client is a subject with a high risk in nasopharyngeal cancer.

18. The computer readable medium of claim 17, wherein the feature tags comprises basic feature information, preference habit information, behavior information, and/or social relationship information.

19. The computer readable medium of claim 17, wherein the analyzing model is a multiple linear regression model; the steps of training the analyzing model comprising:

E1, extracting a first preset number of the various feature data from the several preset businesses servers according to the preset client attribute data;

E2, analyzing the various extracted feature data of each client corrodng to the preset feature tag extraction rule for confirming the feature tags of each client;

E3, confirming abnormal clients with nasopharyngeal cancer in the first preset number of the client according to the preset mapping relation of the nasopharyngeal cancer and the client attribute data, confirming the feature tags corresponding to the abnormal client, and setting the feature tags corresponding to each normal client and each abnormal client as a training sample of the analyzing model;

E4, dividing the training samples into a training set with a first percentage and a validation set with a second percentage; wherein a sum of the first percentage and the second percentage is less than or equal to 100%;

E5, training the analyzing model with the feature tags of each normal client and each abnormal client in the training set, and verifying an accuracy rate of the analyzing model based on the feature tags of each normal client and each abnormal client in the validation set after accomplishing training;

E6, ending the training operation of the analyzing model when the accuracy rate is larger than a preset threshold; and

E7, executing the steps E1, E2, and E3 for increasing a number of the training samples, and re-executing the steps E4 and E5 based on the increased training samples when the accuracy rate is less than or equal to the accuracy rate.

20. The computer readable medium of claim 17, wherein the preset feature tag extraction rule comprising:

setting tag thresholds corresponding to consecutive values of the various feature data;

setting tag ranges corresponding to discrete values of the various feature data; and

confirming tag information corresponding to the feature data of the various consecutive values of each client based on the mapping relationship between the various feature data category of the consecutive values and the feature thresholds, and confirming the tag information corresponding to the feature data of the various discrete values of each client based on the mapping relationship between the various feature data category of the discrete values and the tag ranges.