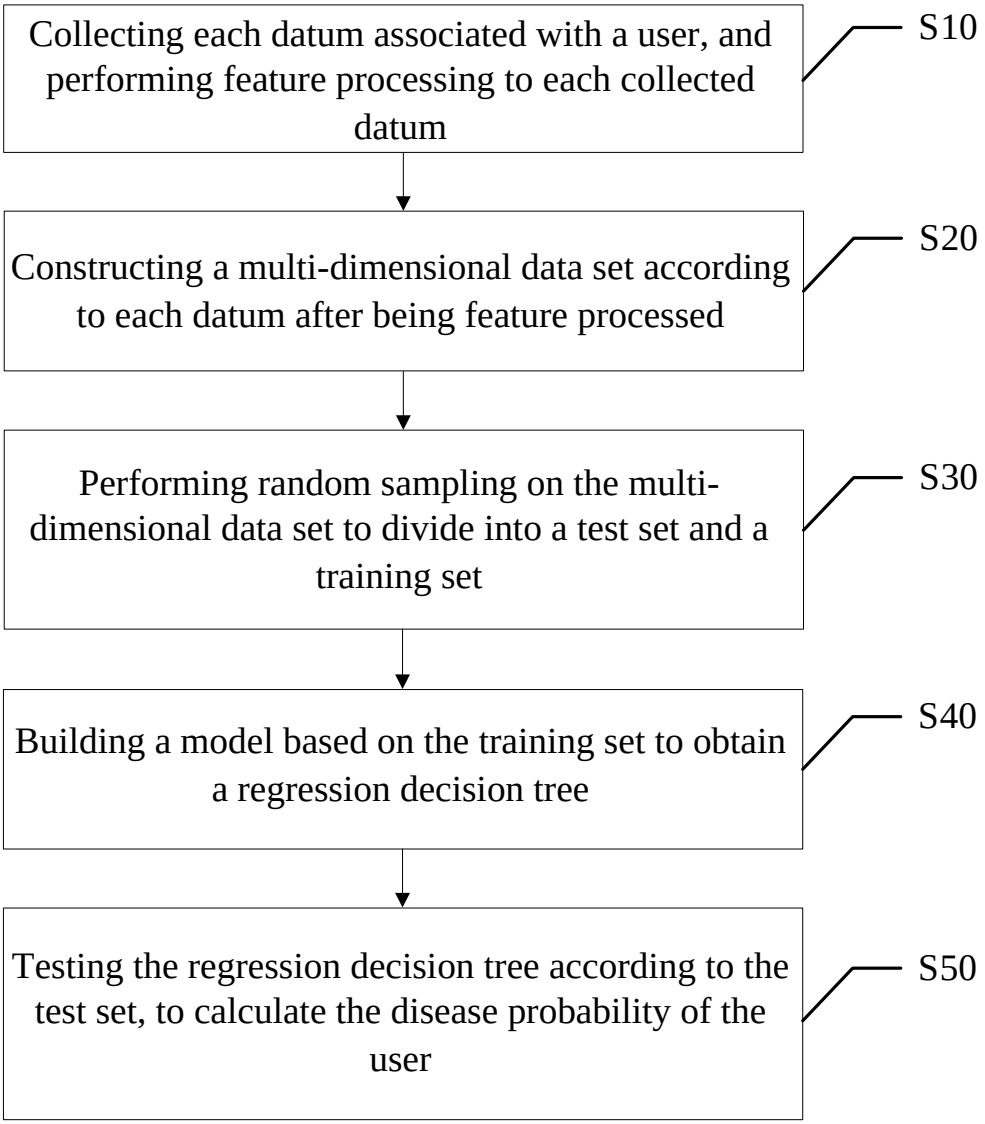
Collecting each datum associated with a user, and performing feature processing to each collected datum
S10
Constructing a multi-dimensional data set according to each datum after being feature processed
S20
Performing random sampling on the multi-dimensional data set to divide into a test set and a training set
S30
Building a model based on the training set to obtain a regression decision tree
S40
Testing the regression decision tree according to the test set, to calculate the disease probability of the user
S50

Fig. 1

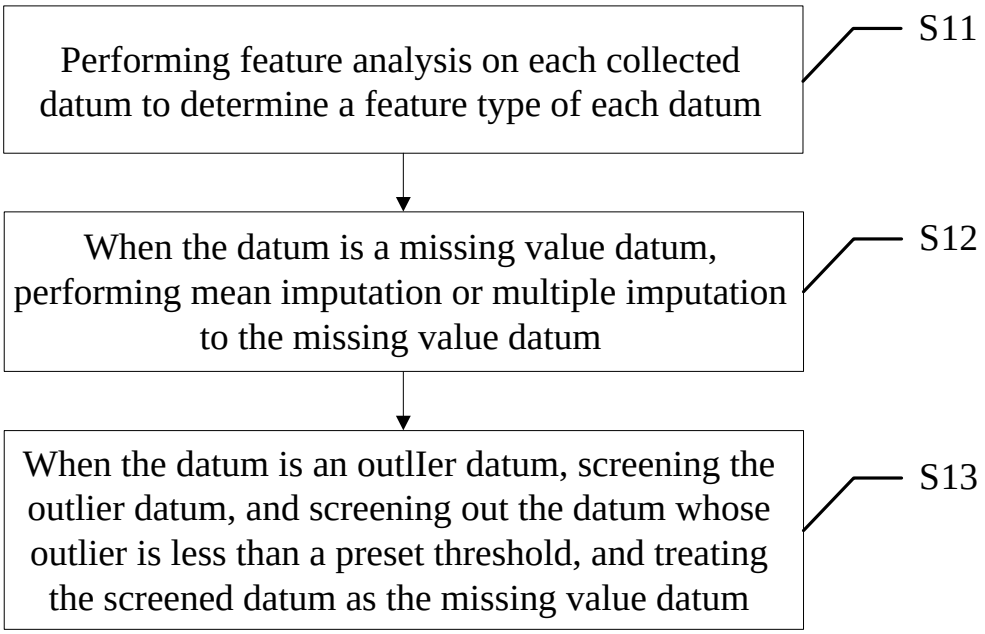
Performing feature analysis on each collected datum to determine a feature type of each datum
S11
When the datum is a missing value datum, performing mean imputation or multiple imputation to the missing value datum
S12
When the datum is an outlIer datum, screening the outlier datum, and screening out the datum whose outlier is less than a preset threshold, and treating the screened datum as the missing value datum
S13

Fig. 2

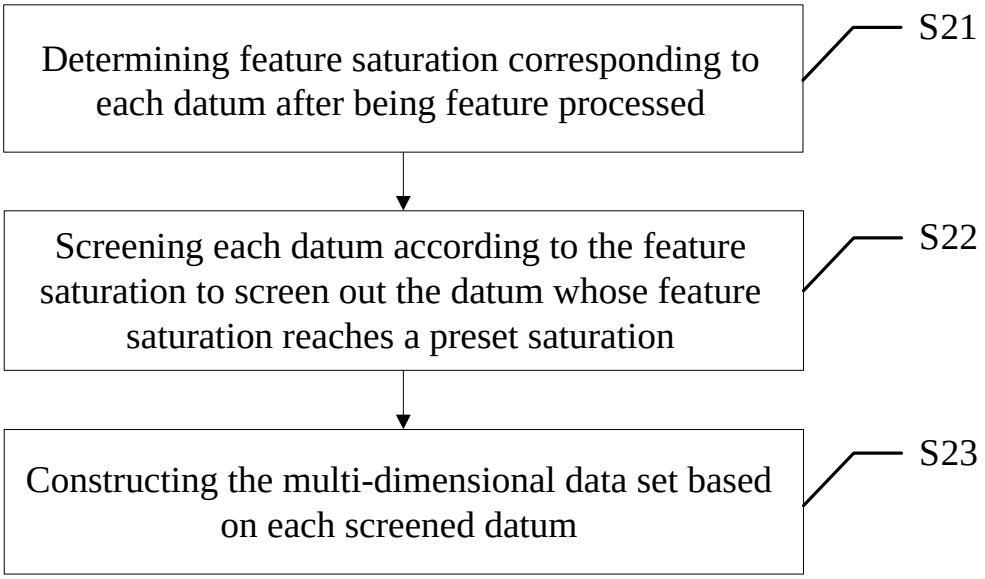
Determining feature saturation corresponding to each datum after being feature processed
S21
Screening each datum according to the feature saturation to screen out the datum whose feature saturation reaches a preset saturation
S22
Constructing the multi-dimensional data set based on each screened datum
S23

Fig. 3

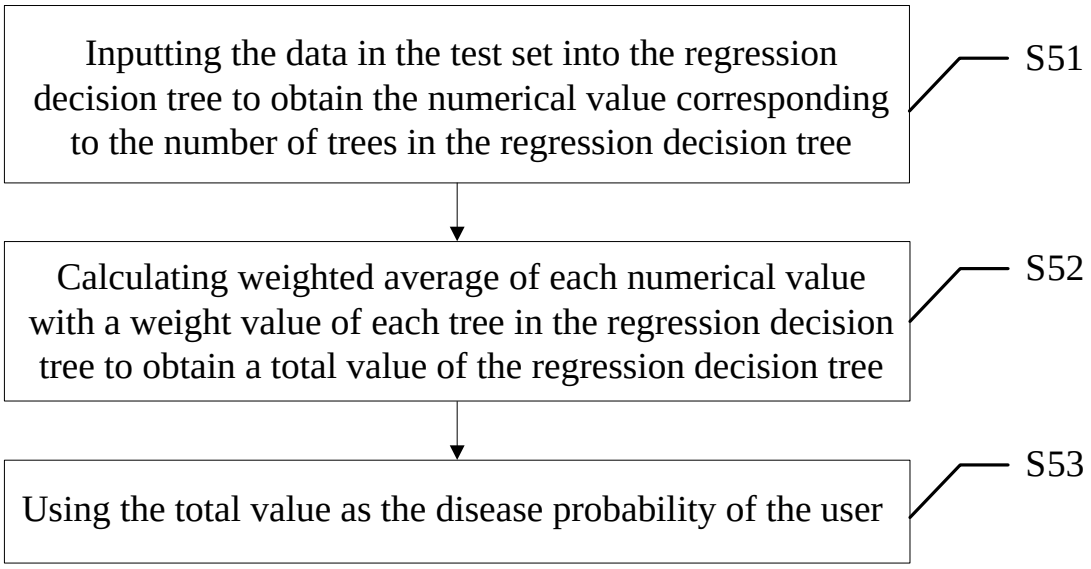
Inputting the data in the test set into the regression decision tree to obtain the numerical value corresponding to the number of trees in the regression decision tree
S51
Calculating weighted average of each numerical value with a weight value of each tree in the regression decision tree to obtain a total value of the regression decision tree
S52
Using the total value as the disease probability of the user
S53

Fig. 4

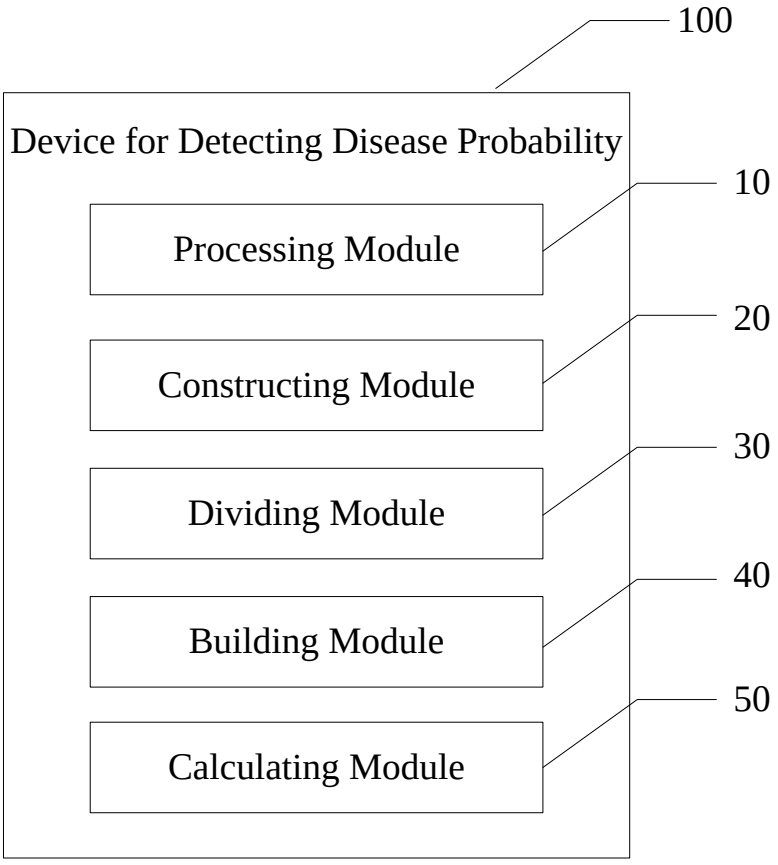
100
Device for Detecting Disease Probability
Processing Module
10
Constructing Module
20
Dividing Module
30
Building Module
40
Calculating Module
50

Fig. 5

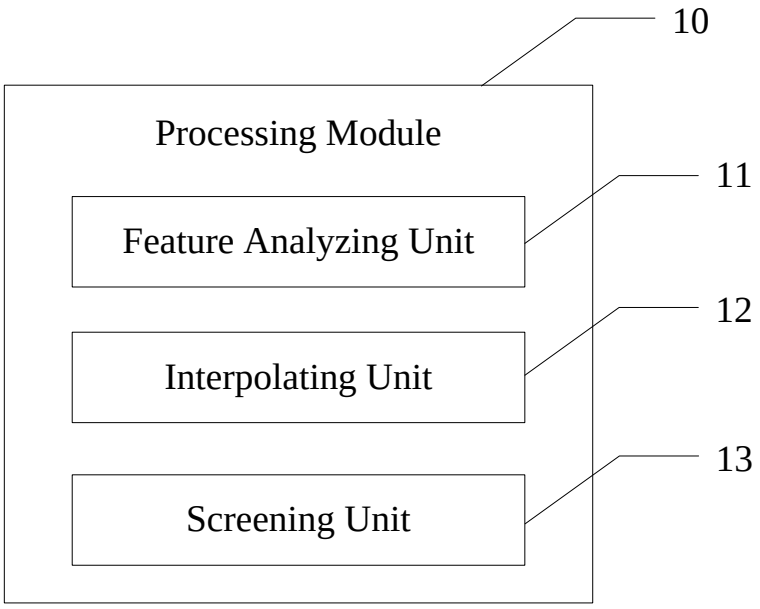
10
Processing Module
11
Feature Analyzing Unit
12
Interpolating Unit
13
Screening Unit

Fig. 6

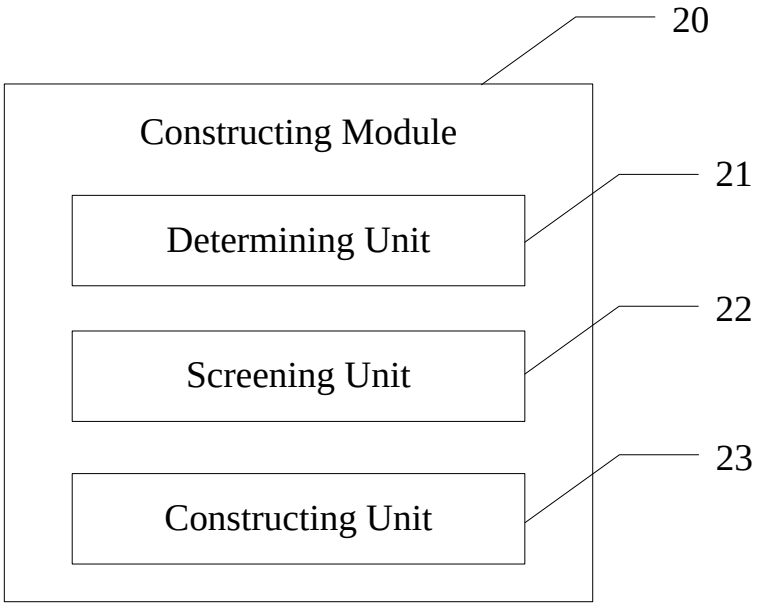
20
Constructing Module
21
Determining Unit
22
Screening Unit
23
Constructing Unit

Fig. 7

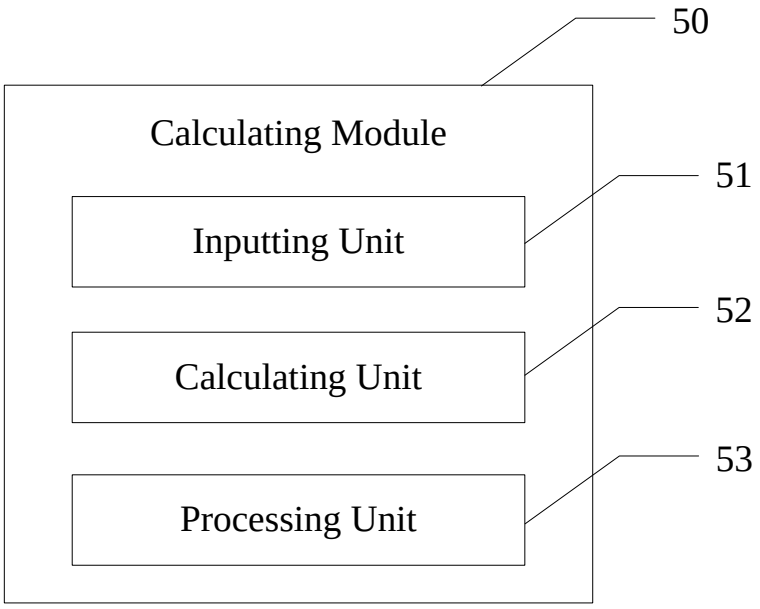
50
Calculating Module
51
Inputting Unit
52
Calculating Unit
53
Processing Unit

Fig. 8

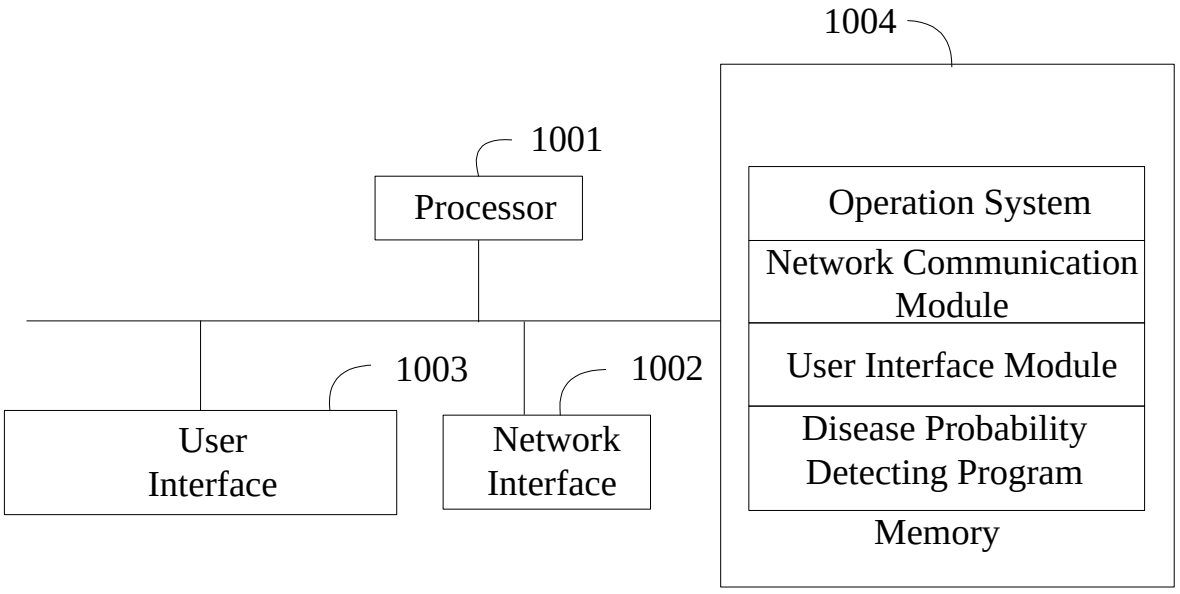
1004
1001
Processor
Operation System
Network Communication Module
User Interface Module
Disease Probability Detecting Program
Memory
1003
1002
User Interface
Network Interface

Fig. 9