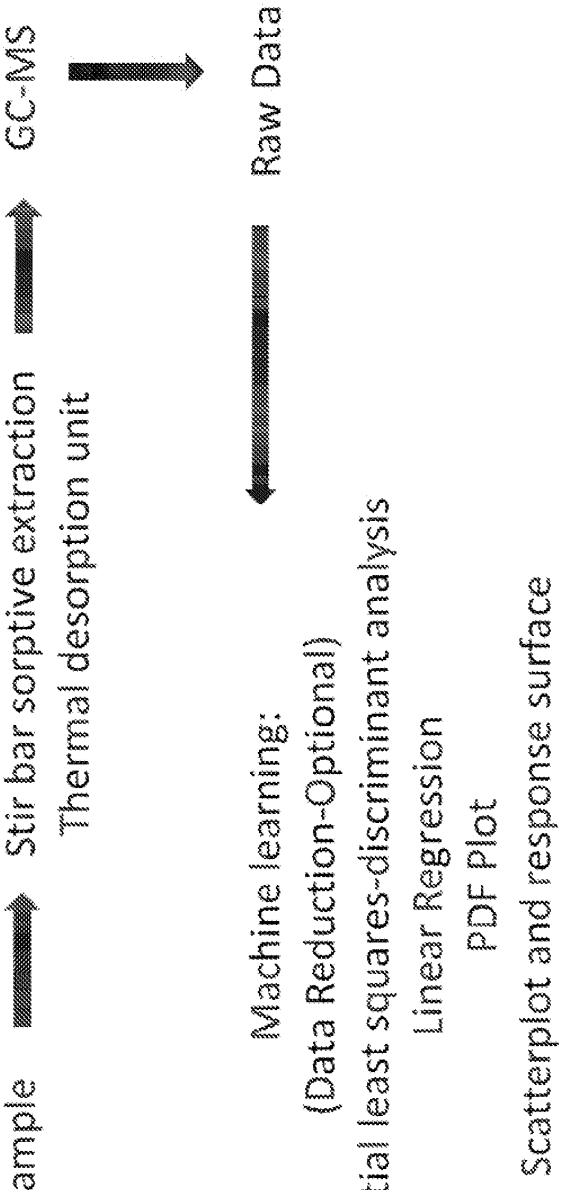
Figure 1


Urine sample
Stir bar sorptive extraction
Thermal desorption unit
GC-MS
Raw Data
Machine learning:
(Data Reduction-Optional)
Partial least squares-discriminant analysis
Linear Regression
PDF Plot
Scatterplot and response surface

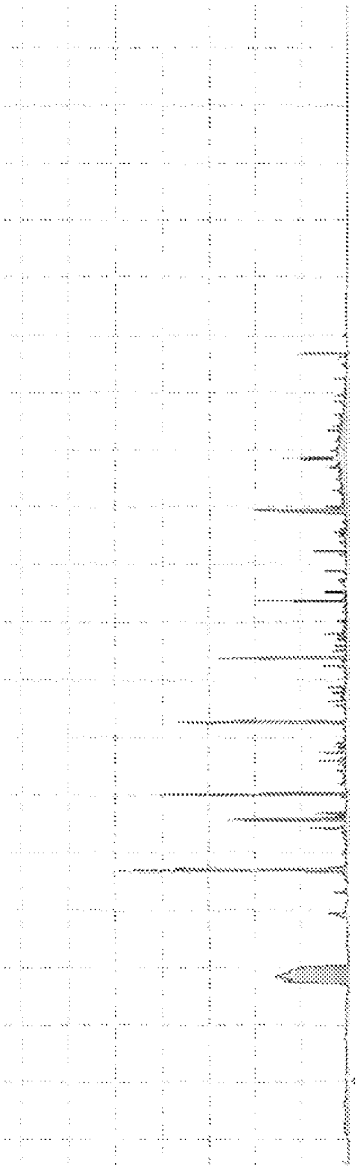


Figure 2

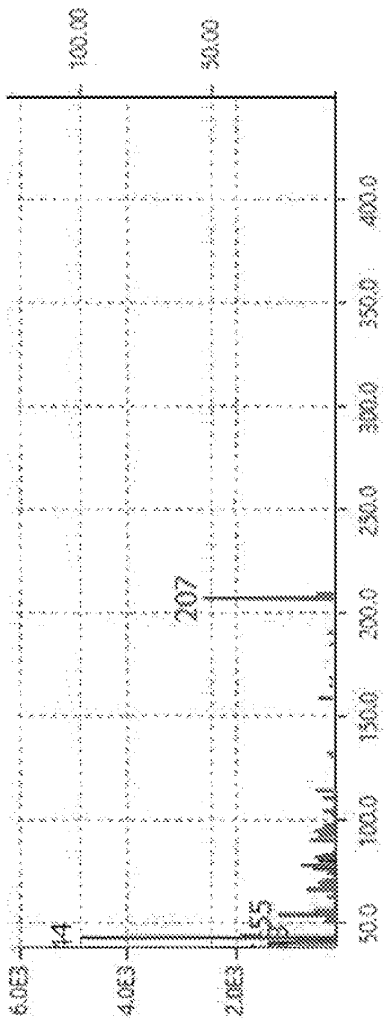

100.00
50.00
6.0E3
4.0E3
2.0E3
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
14
55
207

Figure 3

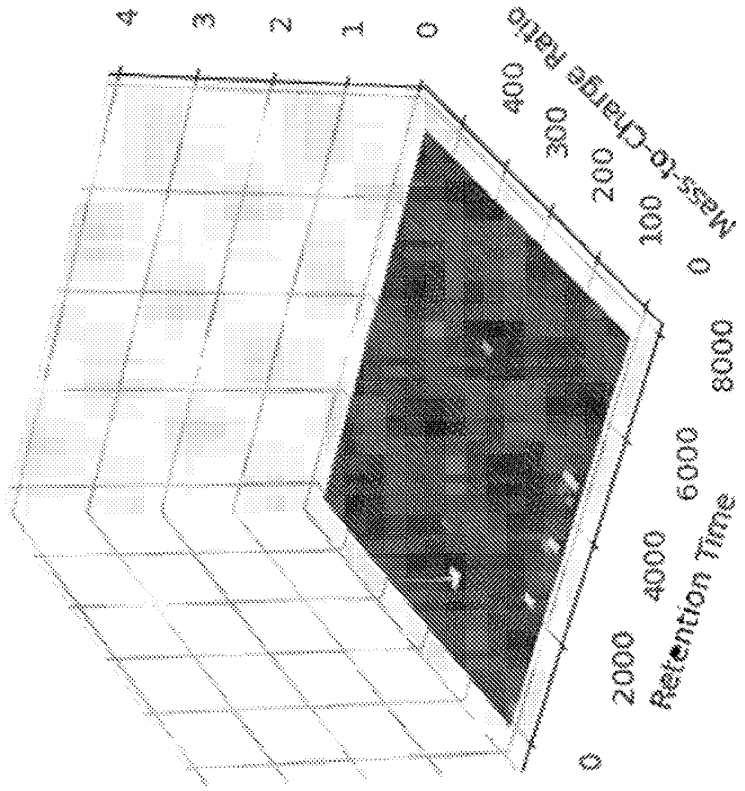

4
3
2
1
0
Mass-to-Charge Ratio
400
300
200
100
0
8000
6000
4000
2000
0
Retention Time

Figure 4

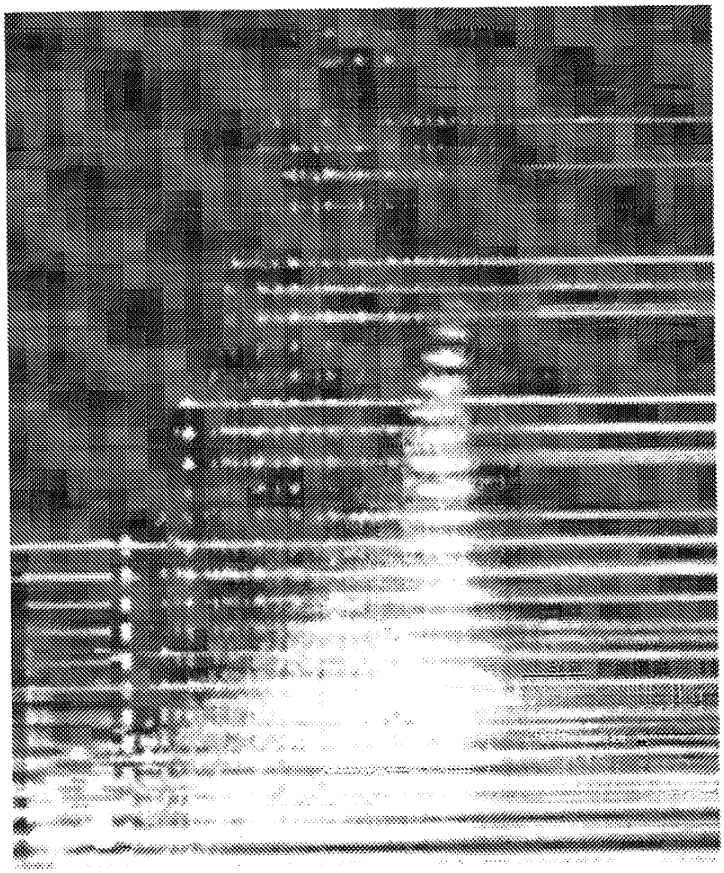


Figure 5

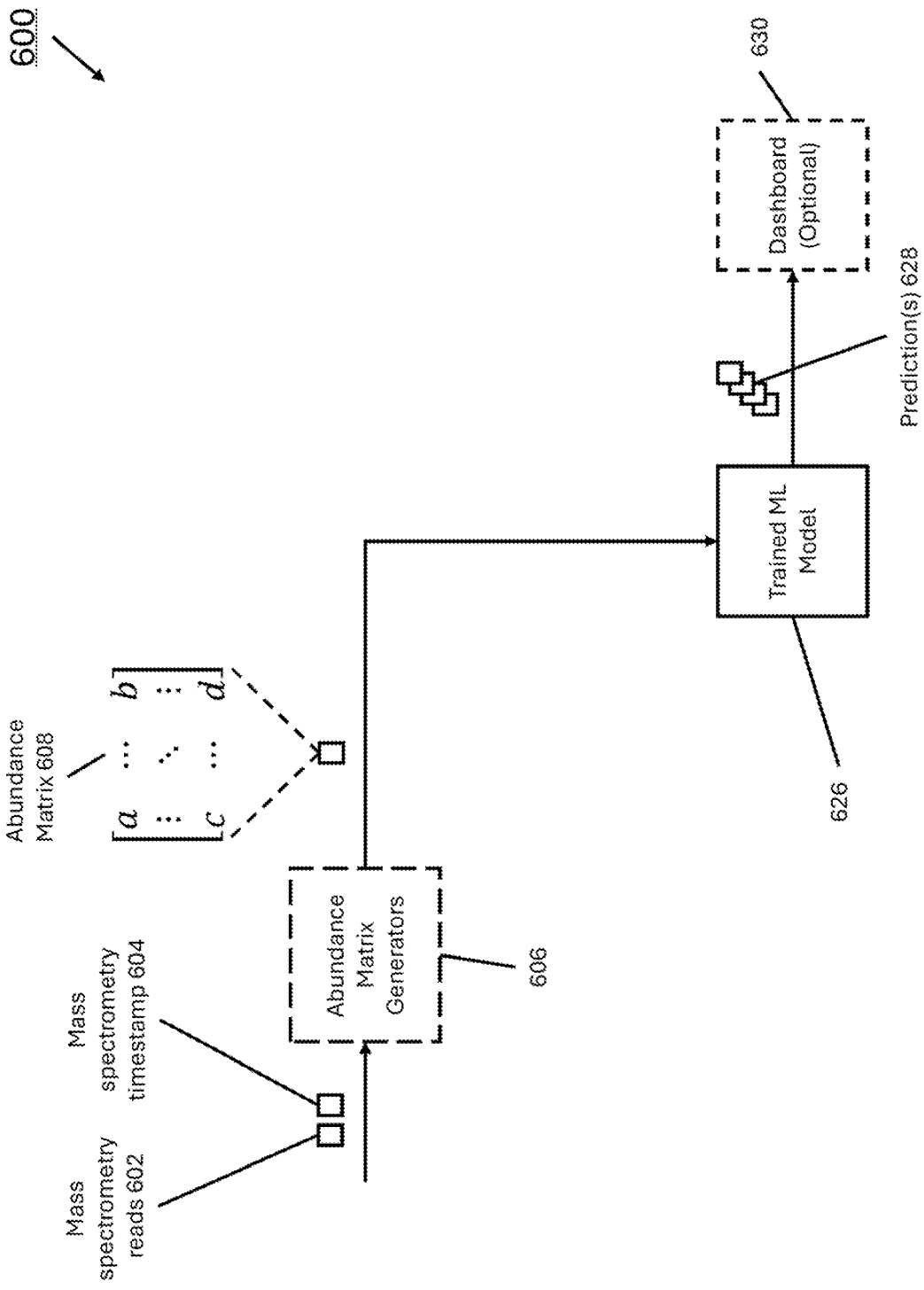

600
Mass spectrometry reads 602
Mass spectrometry timestamp 604
Abundance Matrix Generators
606
Abundance Matrix 608
Trained ML Model
626
Prediction(s) 628
Dashboard (Optional)
630

Figure 6A

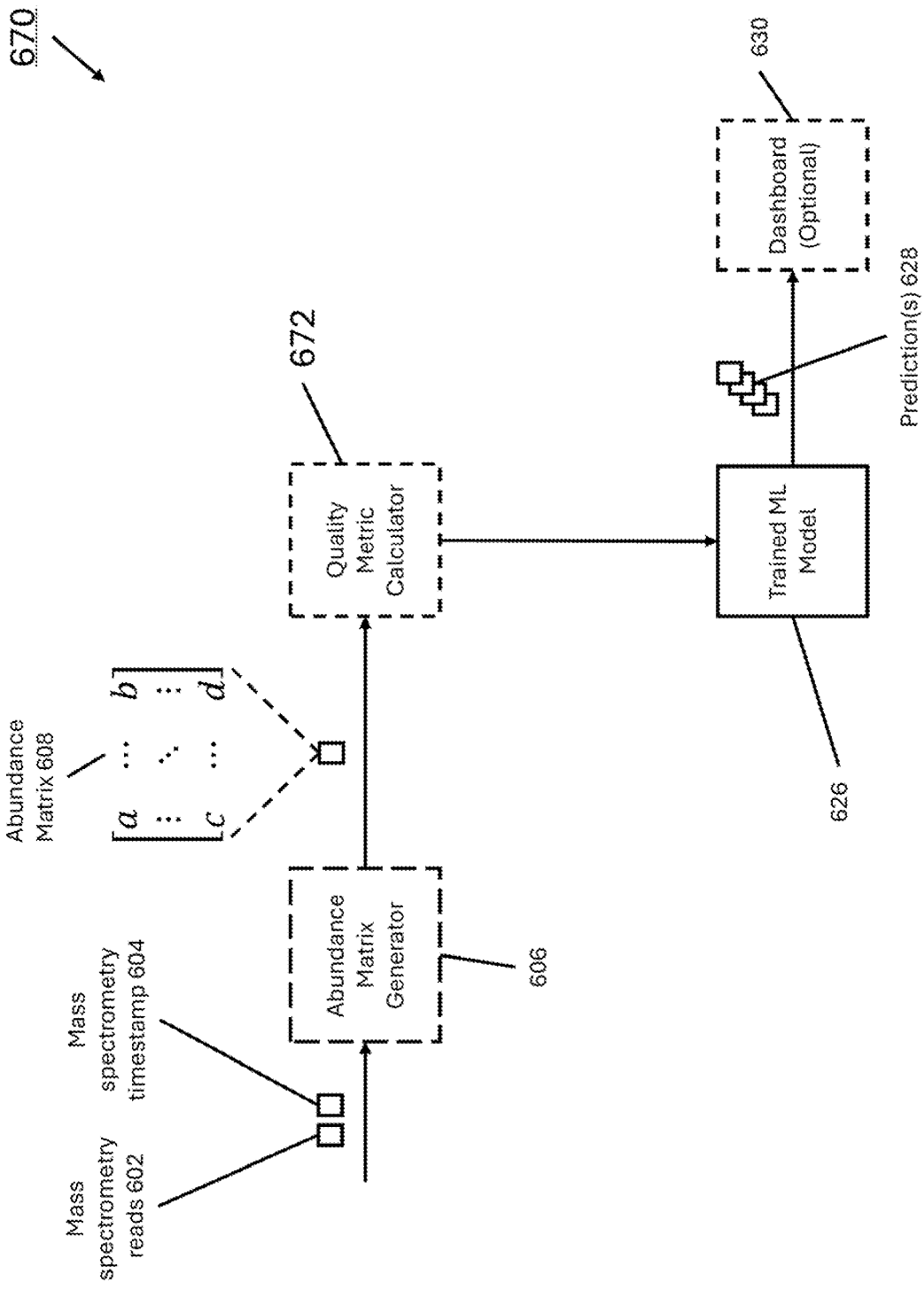

670
Mass spectrometry reads 602
Mass spectrometry timestamp 604
Abundance Matrix Generator
606
Abundance Matrix 608
Quality Metric Calculator
672
Trained ML Model
626
Dashboard (Optional)
630
Prediction(s) 628

Figure 6B

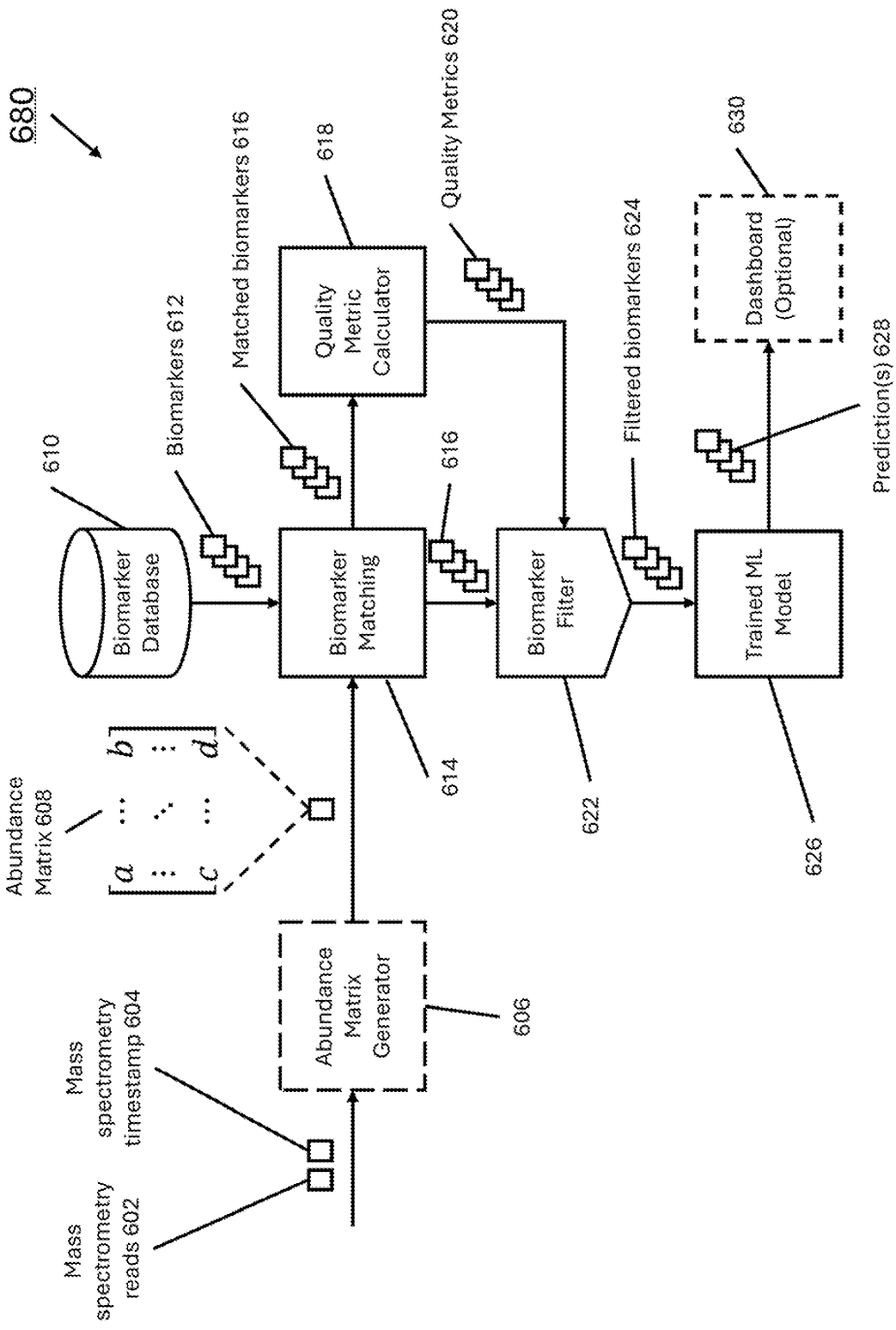

680
Mass spectrometry reads 602
Mass spectrometry timestamp 604
Abundance Matrix Generator
606
Abundance Matrix 608
Biomarker Database
610
Biomarkers 612
Matched biomarkers 616
Quality Metric Calculator
618
Quality Metrics 620
Biomarker Matching
614
616
Biomarker Filter
622
Filtered biomarkers 624
Trained ML Model
626
Prediction(s) 628
Dashboard (Optional)
630

Figure 6C

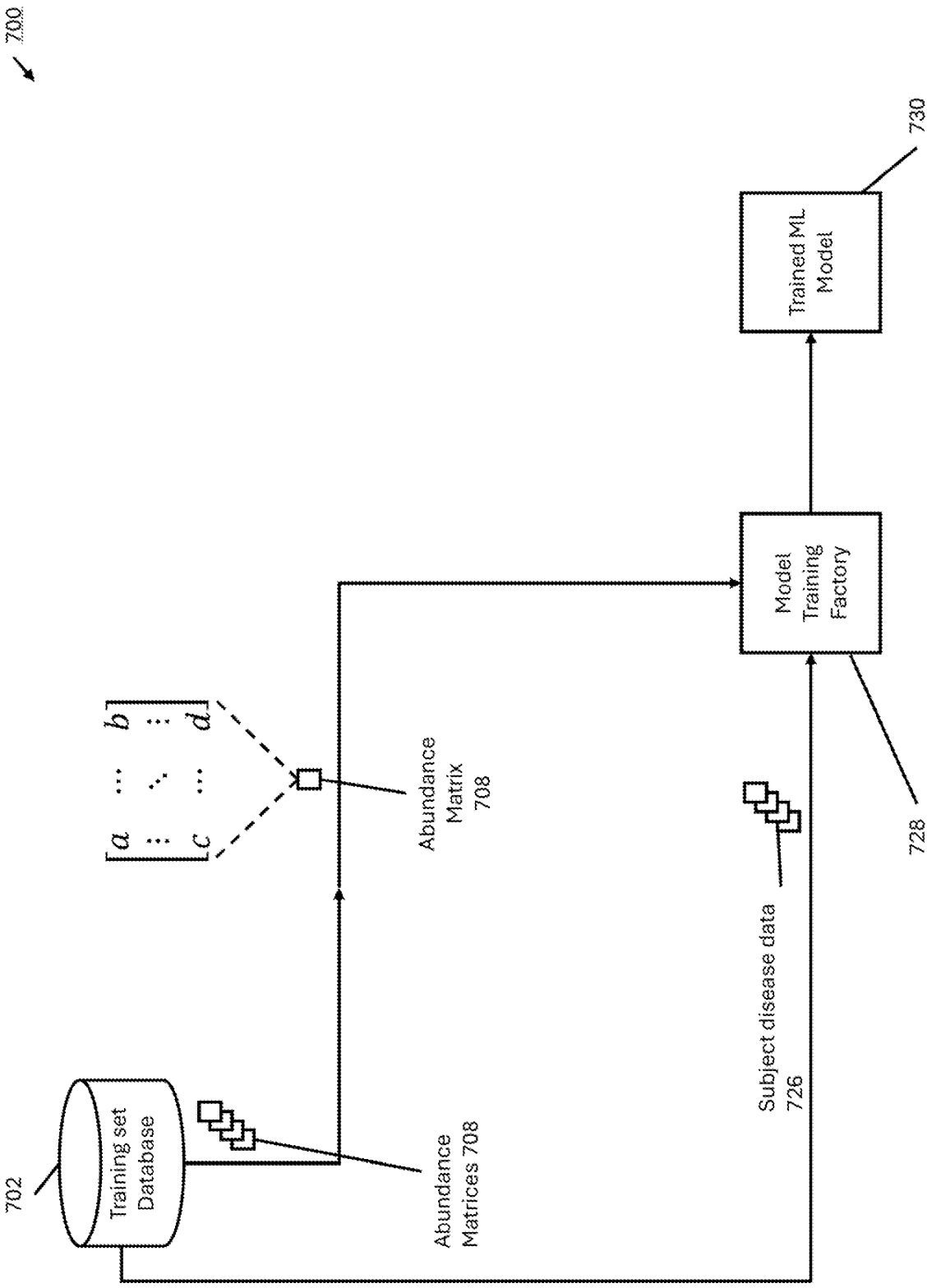

700
Trained ML Model
730
Model Training Factory
728
Subject disease data 726
Abundance Matrix 708
Abundance Matrices 708
Training set Database
702

Figure 7A

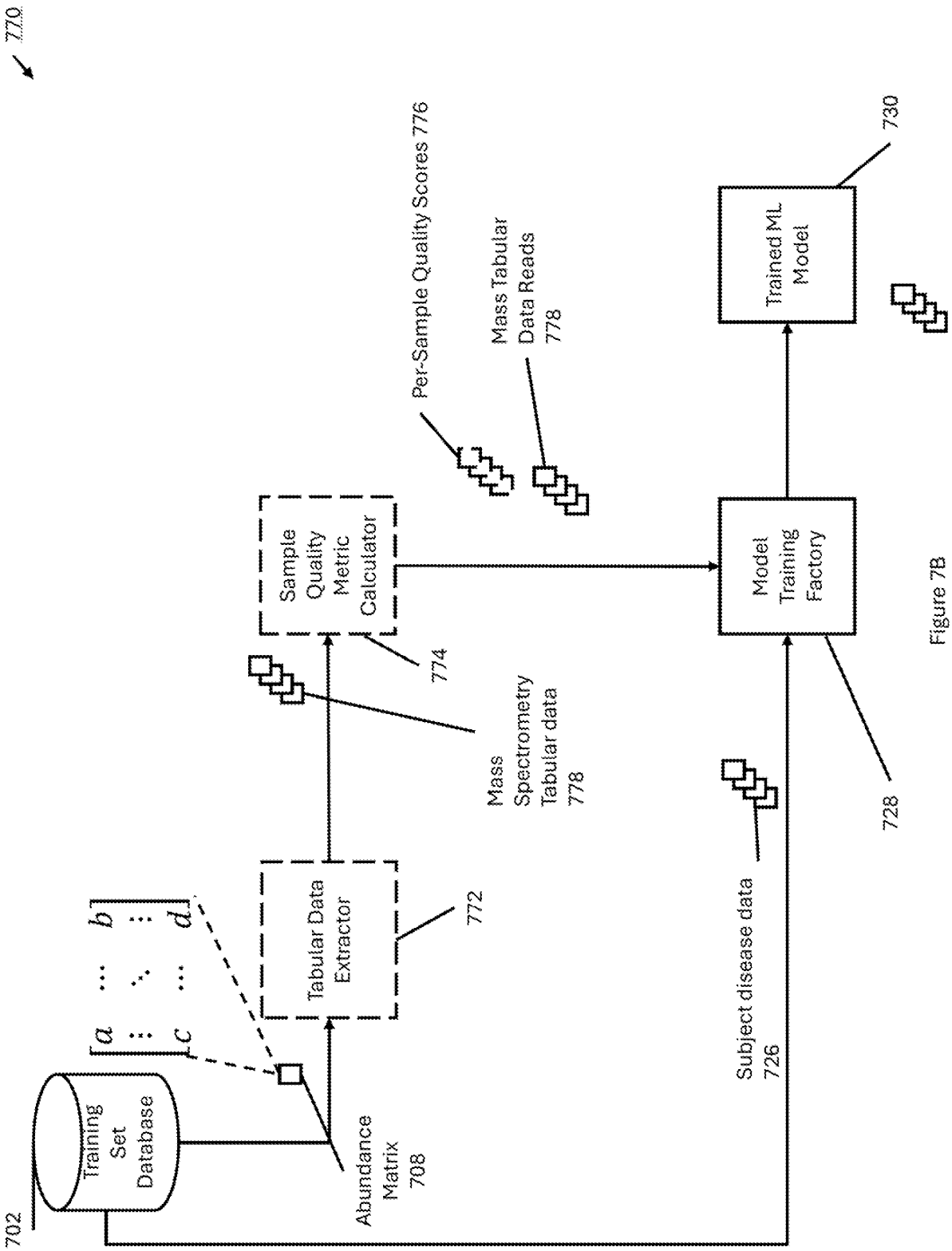

770
Training Set Database
702
Abundance Matrix 708
Tabular Data Extractor
772
Mass Spectrometry Tabular data 778
Sample Quality Metric Calculator
774
Per-Sample Quality Scores 776
Mass Tabular Data Reads 778
Model Training Factory
728
Subject disease data 726
Trained ML Model
730

Figure 7B

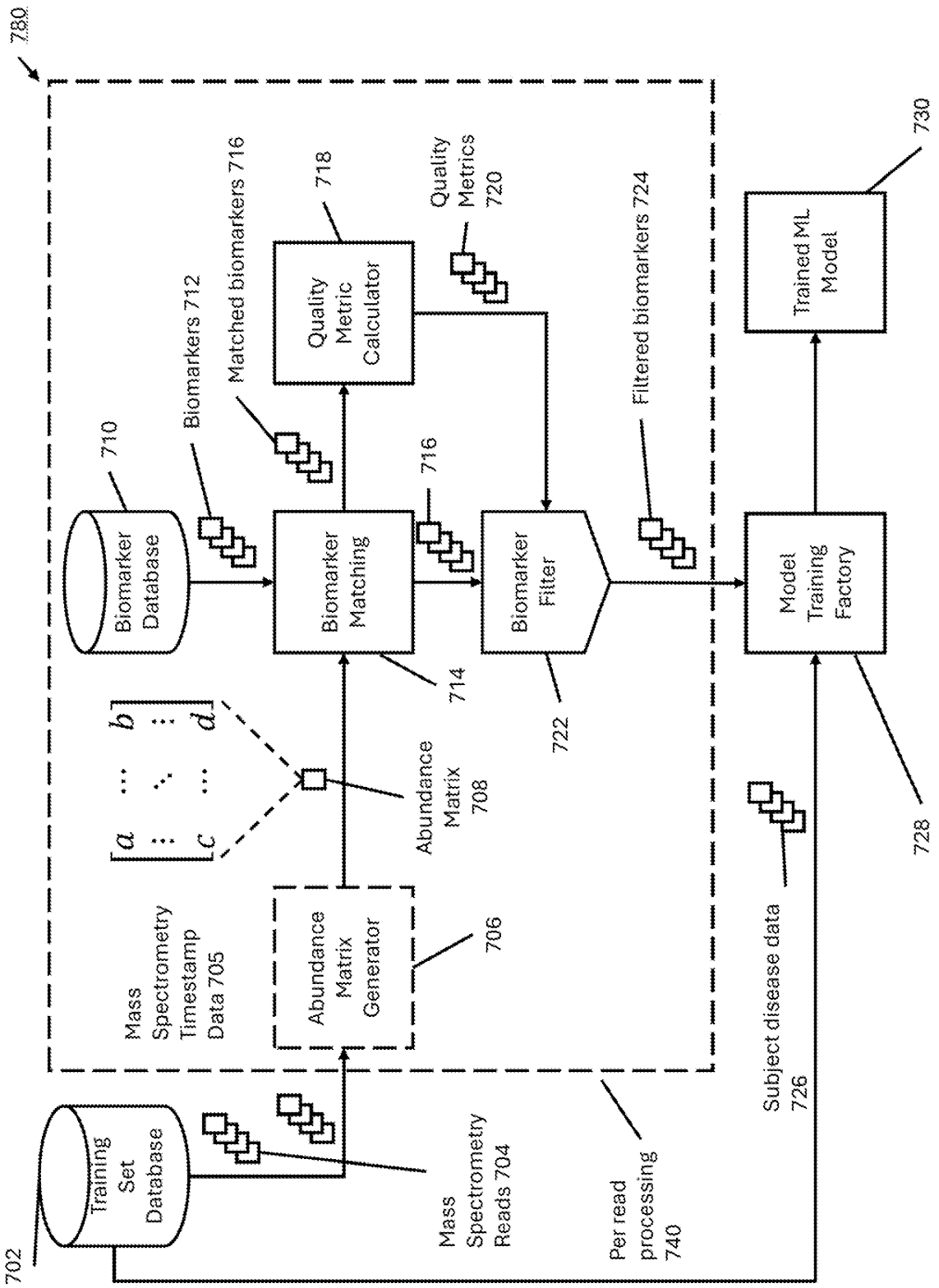

780
Training Set Database
702
Mass Spectrometry Reads 704
Per read processing 740
Mass Spectrometry Timestamp Data 705
Abundance Matrix Generator
706
Abundance Matrix 708
Biomarker Database
710
Biomarkers 712
Biomarker Matching
714
Matched biomarkers 716
Quality Metric Calculator
718
Quality Metrics 720
Biomarker Filter
722
Filtered biomarkers 724
Model Training Factory
728
Subject disease data 726
Trained ML Model
730

Figure 7C

800

History of uploads

File name	Bladder cancer	Kidney cancer	Prostate cancer
Dec 30, 2024			
Male_68_Bladder_Cancer.D	Positive	Negative	Negative
Male_76_Healthy.D	Negative	Negative	Negative
Male_68_Bladder_Cancer.D	Positive	Negative	Negative
Male_76_Healthy.D	Negative	Negative	Negative
Dec 26, 2024			
Male_68_Bladder_Cancer.D	Positive	Negative	Negative

Figure 8A

850

History of uploads

File name	Bladder Cancer	Kidney Cancer	Prostate Cancer
Today			
Patient_5.D	1st Positive	Negative	2nd Positive
Patient_4.D	2nd Positive	1st Positive	3rd Positive
Patient_3.D	1st Positive	2nd Positive	Negative
Patient_2.D	Negative	Negative	Negative

Figure 8B

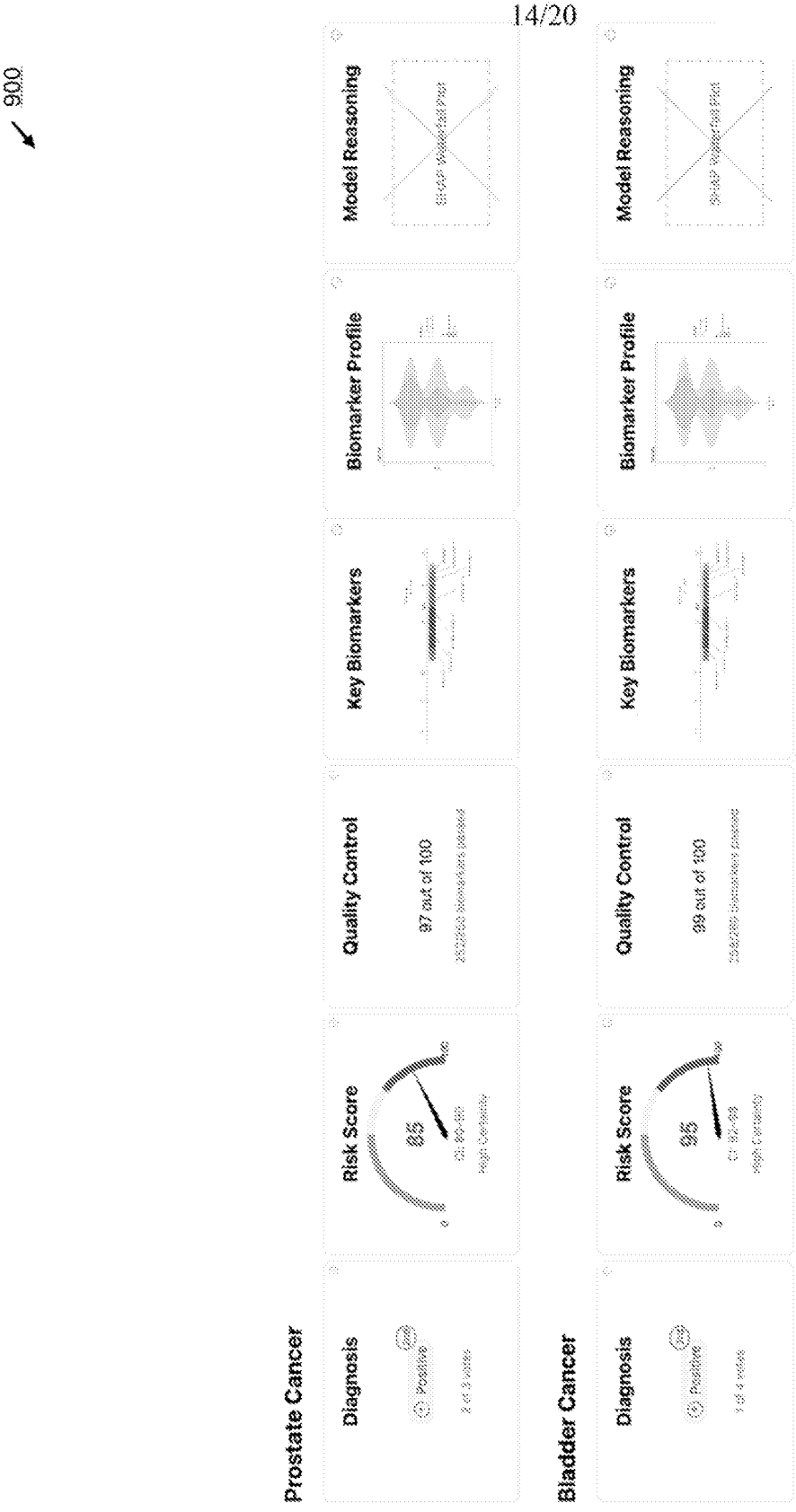

900
Prostate Cancer
Diagnosis
Positive
Risk Score
85
Quality Control
97 out of 100
Key Biomarkers
Biomarker Profile
Model Reasoning
Bladder Cancer
Diagnosis
Positive
Risk Score
95
Quality Control
99 out of 100
Key Biomarkers
Biomarker Profile
Model Reasoning

Figure 9

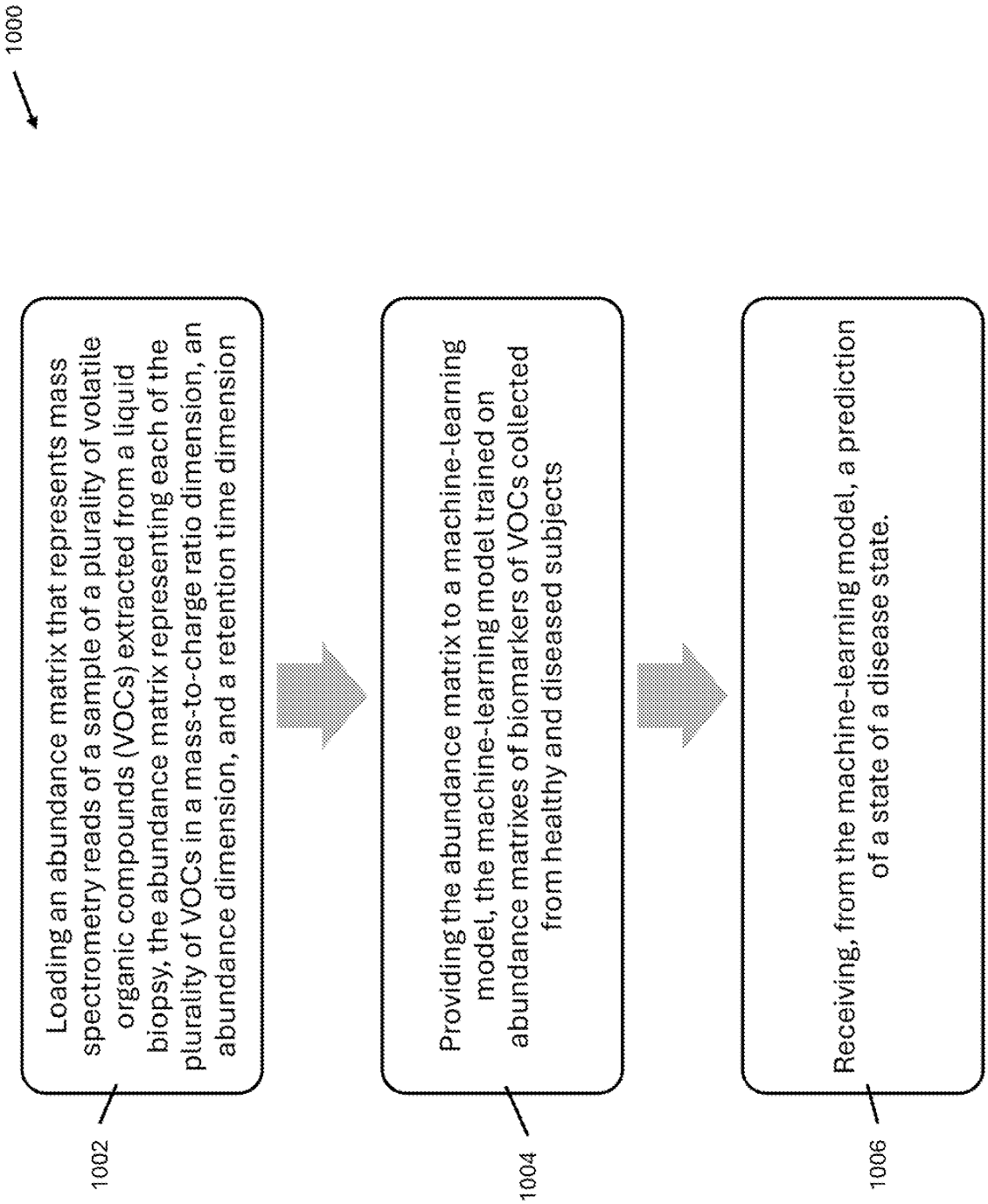

1000
Loading an abundance matrix that represents mass spectrometry reads of a sample of a plurality of volatile organic compounds (VOCs) extracted from a liquid biopsy, the abundance matrix representing each of the plurality of VOCs in a mass-to-charge ratio dimension, an abundance dimension, and a retention time dimension
1002
Providing the abundance matrix to a machine-learning model, the machine-learning model trained on abundance matrixes of biomarkers of VOCs collected from healthy and diseased subjects
1004
Receiving, from the machine-learning model, a prediction of a state of a disease state.
1006

Figure 10A

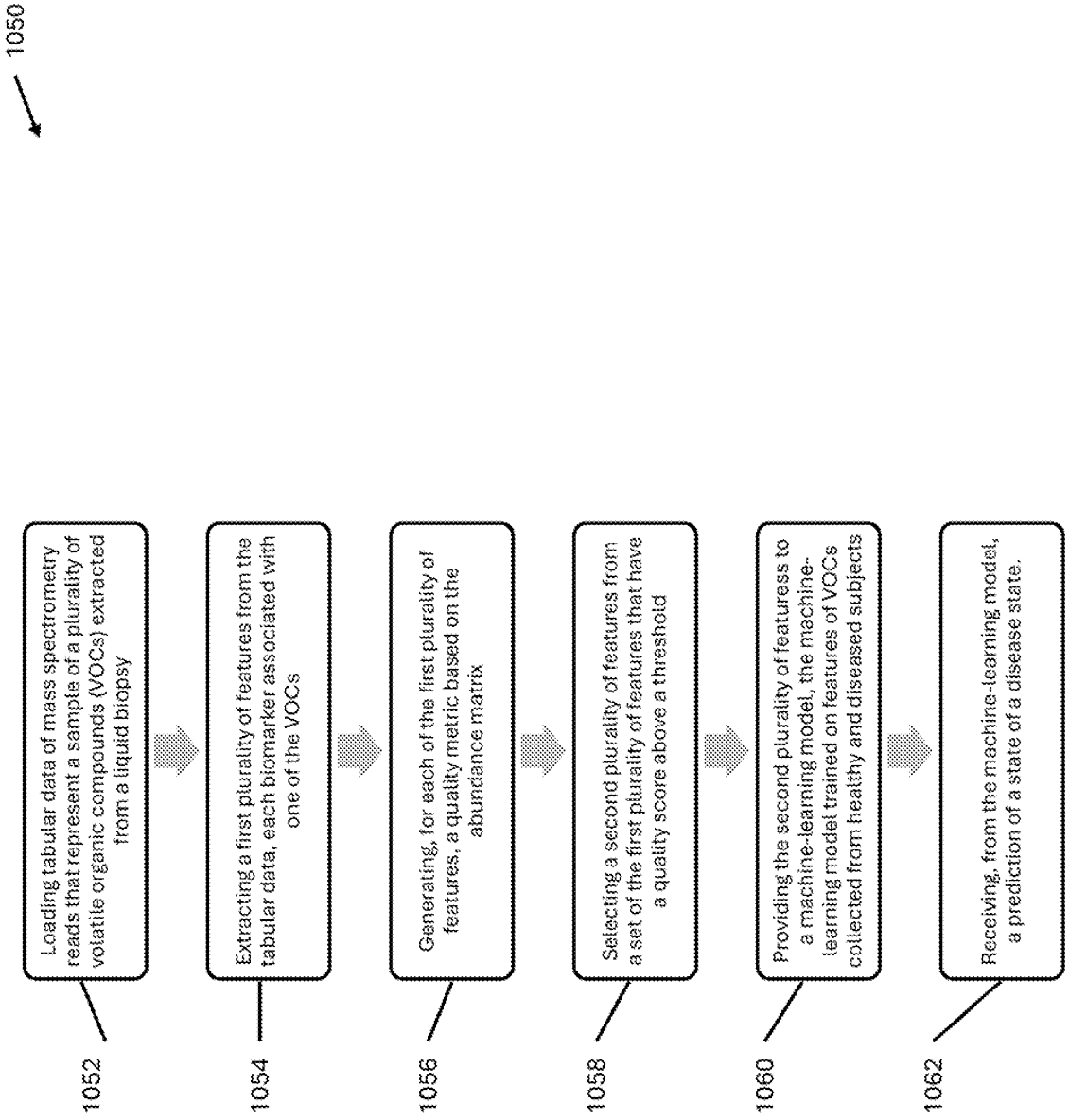

1050
Loading tabular data of mass spectrometry reads that represent a sample of a plurality of volatile organic compounds (VOCs) extracted from a liquid biopsy
1052
Extracting a first plurality of features from the tabular data, each biomarker associated with one of the VOCs
1054
Generating, for each of the first plurality of features, a quality metric based on the abundance matrix
1056
Selecting a second plurality of features from a set of the first plurality of features that have a quality score above a threshold
1058
Providing the second plurality of featuress to a machine-learning model, the machine-learning model trained on features of VOCs collected from healthy and diseased subjects
1060
Receiving, from the machine-learning model, a prediction of a state of a disease state.
1062

Figure 10B

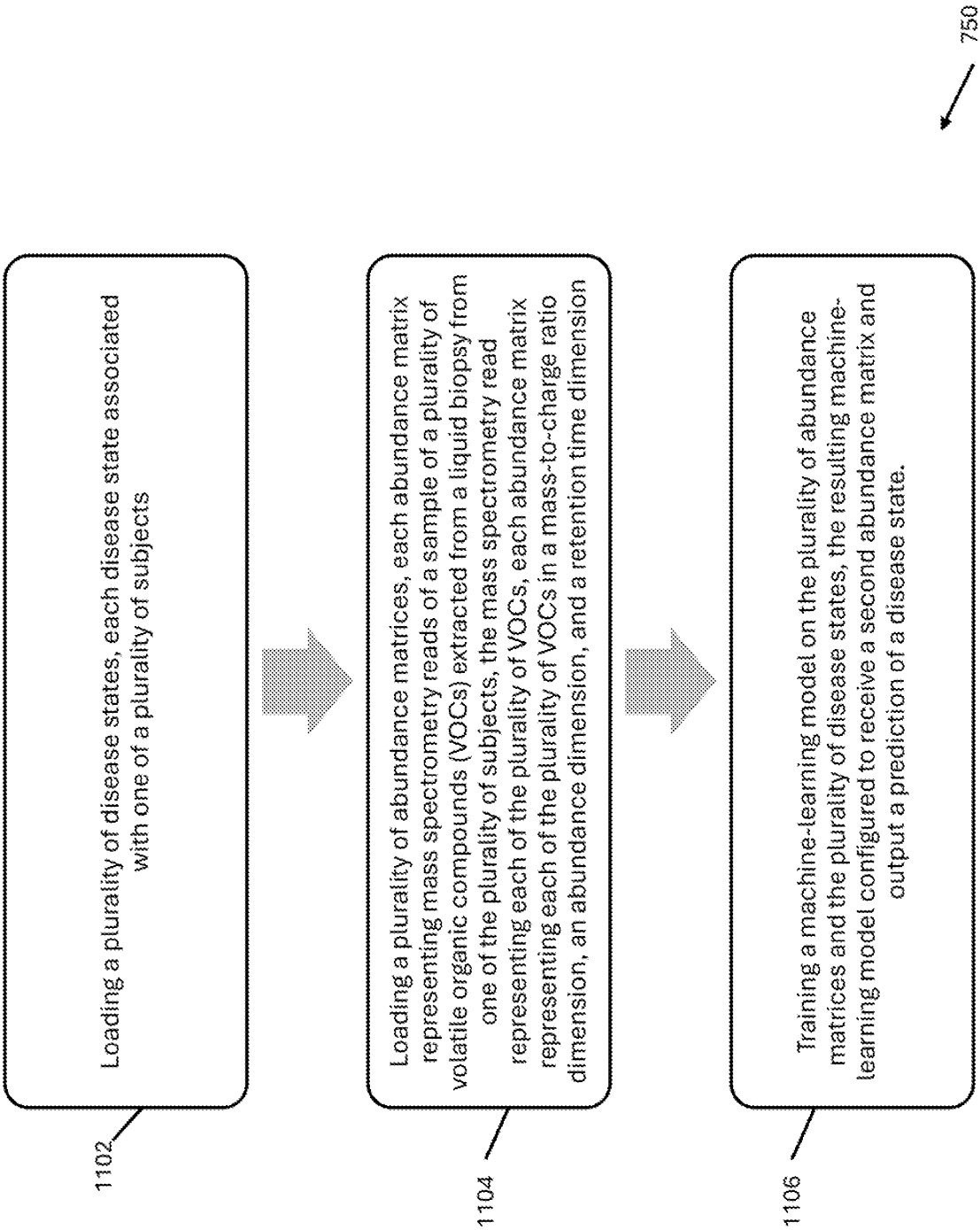

750
Loading a plurality of disease states, each disease state associated with one of a plurality of subjects
1102
Loading a plurality of abundance matrices, each abundance matrix representing mass spectrometry reads of a sample of a plurality of volatile organic compounds (VOCs) extracted from a liquid biopsy from one of the plurality of subjects, the mass spectrometry read representing each of the plurality of VOCs, each abundance matrix representing each of the plurality of VOCs in a mass-to-charge ratio dimension, an abundance dimension, and a retention time dimension
1104
Training a machine-learning model on the plurality of abundance matrices and the plurality of disease states, the resulting machine-learning model configured to receive a second abundance matrix and output a prediction of a disease state.
1106

Figure 11A

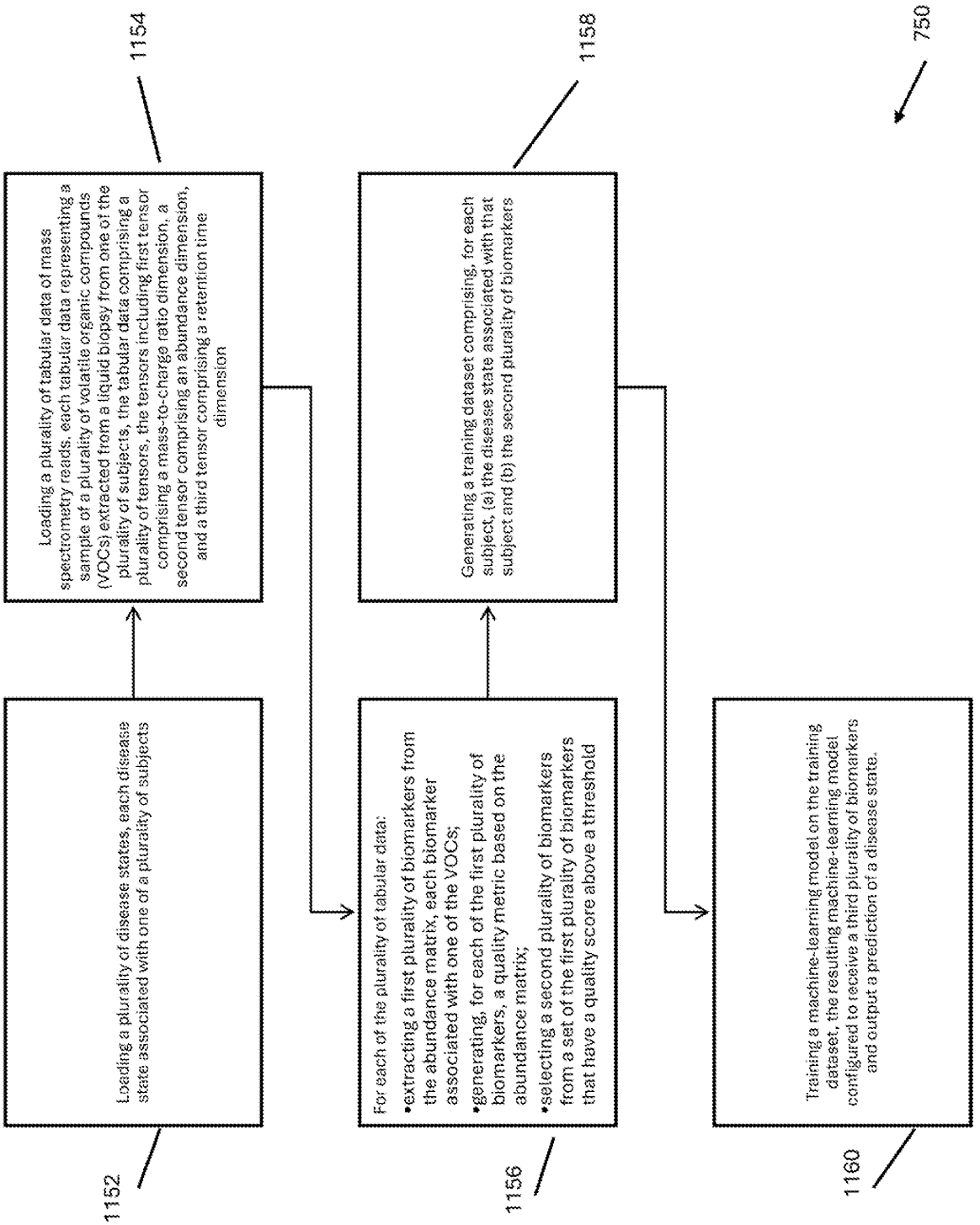

750
1154
1158
Loading a plurality of tabular data of mass spectrometry reads, each tabular data representing a sample of a plurality of volatile organic compounds (VOCs) extracted from a liquid biopsy from one of the plurality of subjects, the tabular data comprising a plurality of tensors, the tensors including first tensor comprising a mass-to-charge ratio dimension, a second tensor comprising an abundance dimension, and a third tensor comprising a retention time dimension
Generating a training dataset comprising, for each subject, (a) the disease state associated with that subject and (b) the second plurality of biomarkers
Loading a plurality of disease states, each disease state associated with one of a plurality of subjects
For each of the plurality of tabular data:
•extracting a first plurality of biomarkers from the abundance matrix, each biomarker associated with one of the VOCs;
•generating, for each of the first plurality of biomarkers, a quality metric based on the abundance matrix;
•selecting a second plurality of biomarkers from a set of the first plurality of biomarkers that have a quality score above a threshold
Training a machine-learning model on the training dataset, the resulting machine-learning model configured to receive a third plurality of biomarkers and output a prediction of a disease state.
1152
1156
1160

Figure 11B

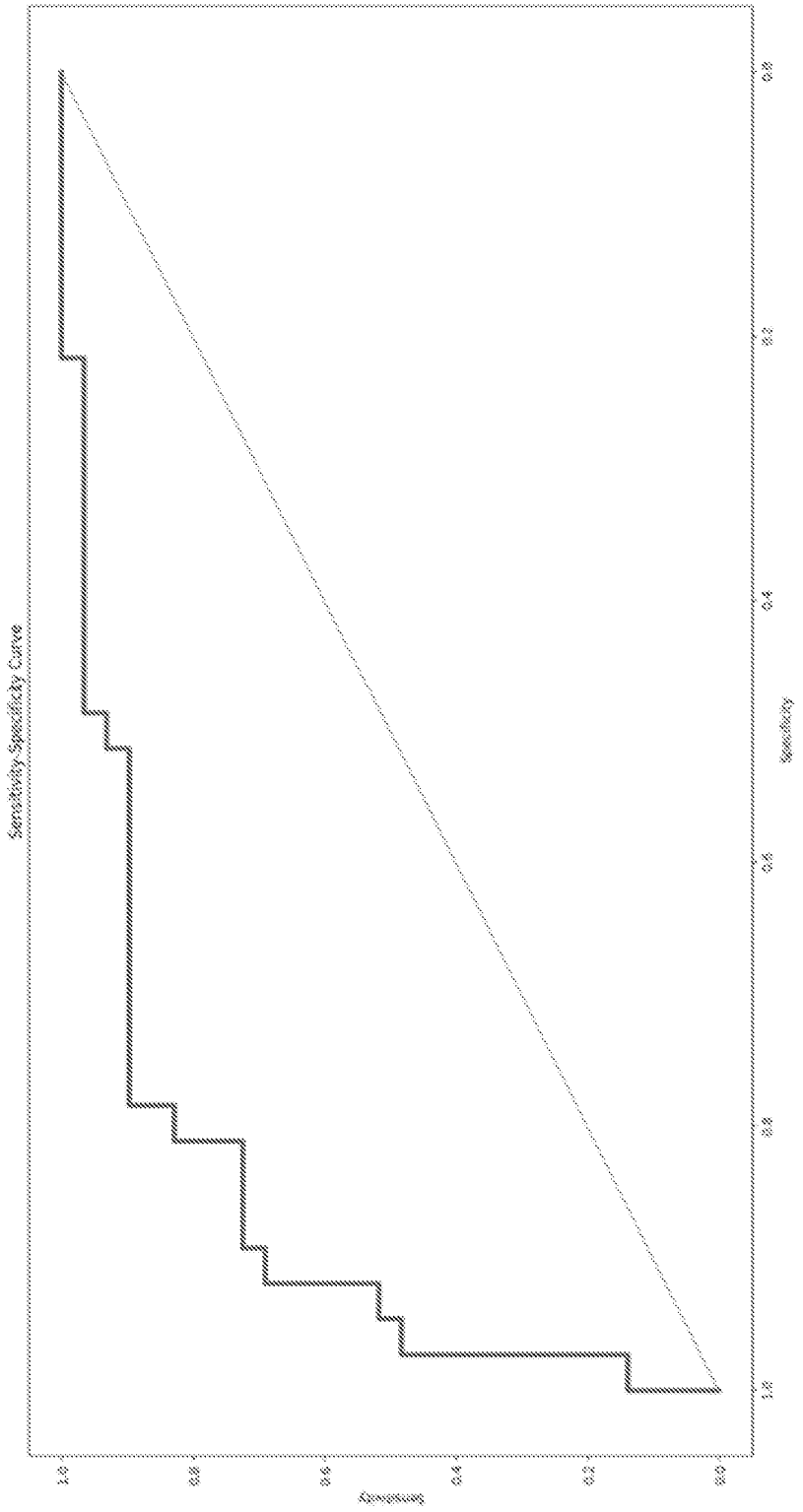

Sensitivity Specificity Curve
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

Figure 12

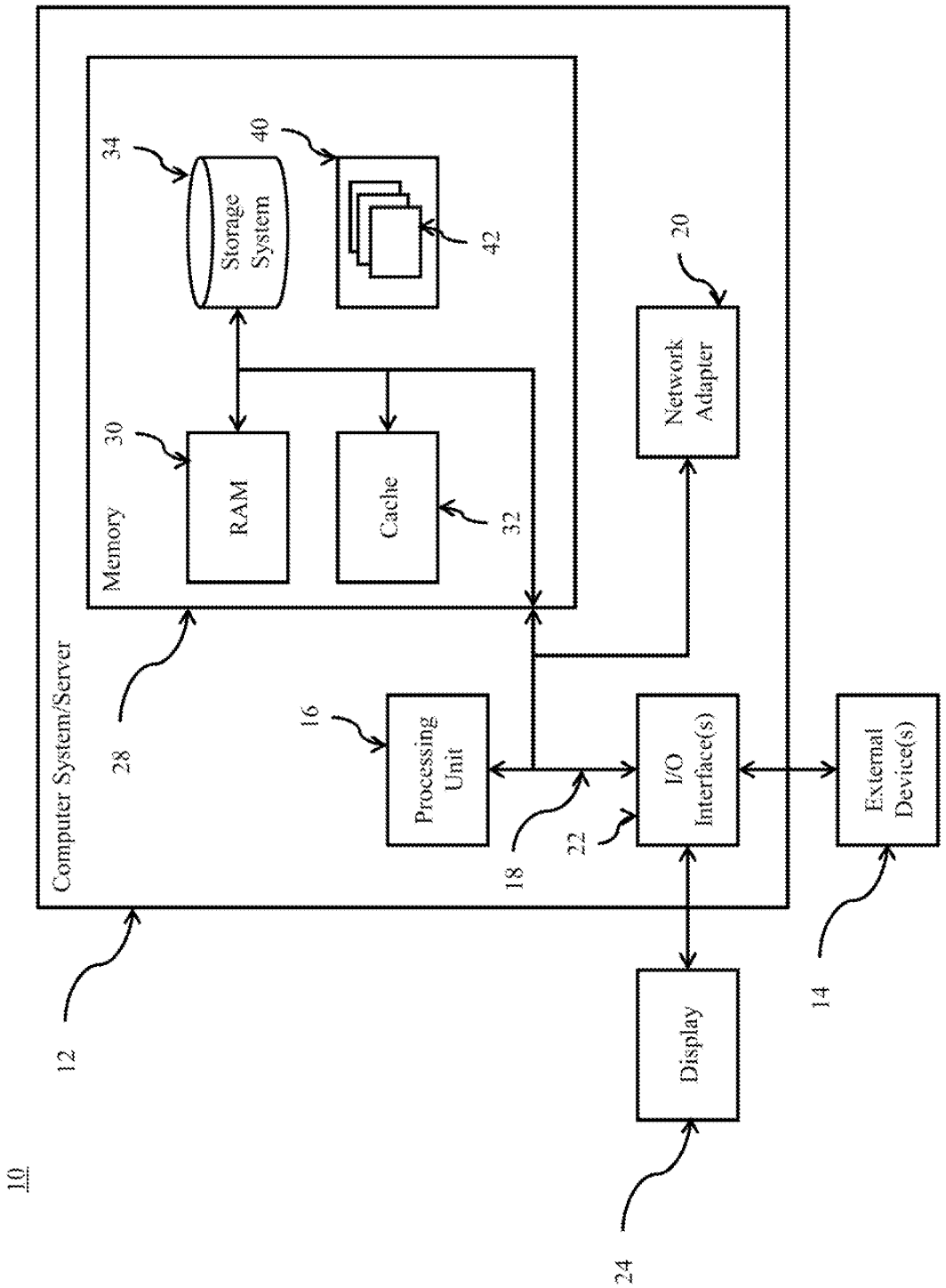

10
Computer System/Server
12
Memory
28
34
Storage System
40
42
30
RAM
Cache
32
20
Network Adapter
16
Processing Unit
18
22
I/O Interface(s)
External Device(s)
14
Display
24

Figure 13