Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2990367 A1 2016/12/29

(21) 2 990 367

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2016/06/24
(87) Date publication PCT/PCT Publication Date: 2016/12/29
(85) Entrée phase nationale/National Entry: 2017/12/20
(86) N° demande PCT/PCT Application No.: IB 2016/053797
(87) N° publication PCT/PCT Publication No.: 2016/207862
(30) Priorité/Priority: 2015/06/25 (US62/184,796)

(51) Cl.Int./Int.Cl. *A61B 5/02* (2006.01), *A61B 90/00* (2016.01), *G06F 15/18* (2006.01), *G06F 19/00* (2018.01)

(71) Demandeur/Applicant:
ANALYTICS FOR LIFE INC., CA

(72) Inventeurs/Inventors:
BURTON, TIMOTHY, CA;
RAMCHANDANI, SHYAMLAL, CA;
GUPTA, SUNNY, CA

(74) Agent: MARKS & CLERK

(54) Titre : PROCEDES ET SYSTEMES UTILISANT UNE ANALYSE MATHEMATIQUE ET L'APPRENTISSAGE AUTOMATIQUE AFIN DE DIAGNOSTIQUER UNE MALADIE

(54) Title: METHODS AND SYSTEMS USING MATHEMATICAL ANALYSIS AND MACHINE LEARNING TO DIAGNOSE DISEASE

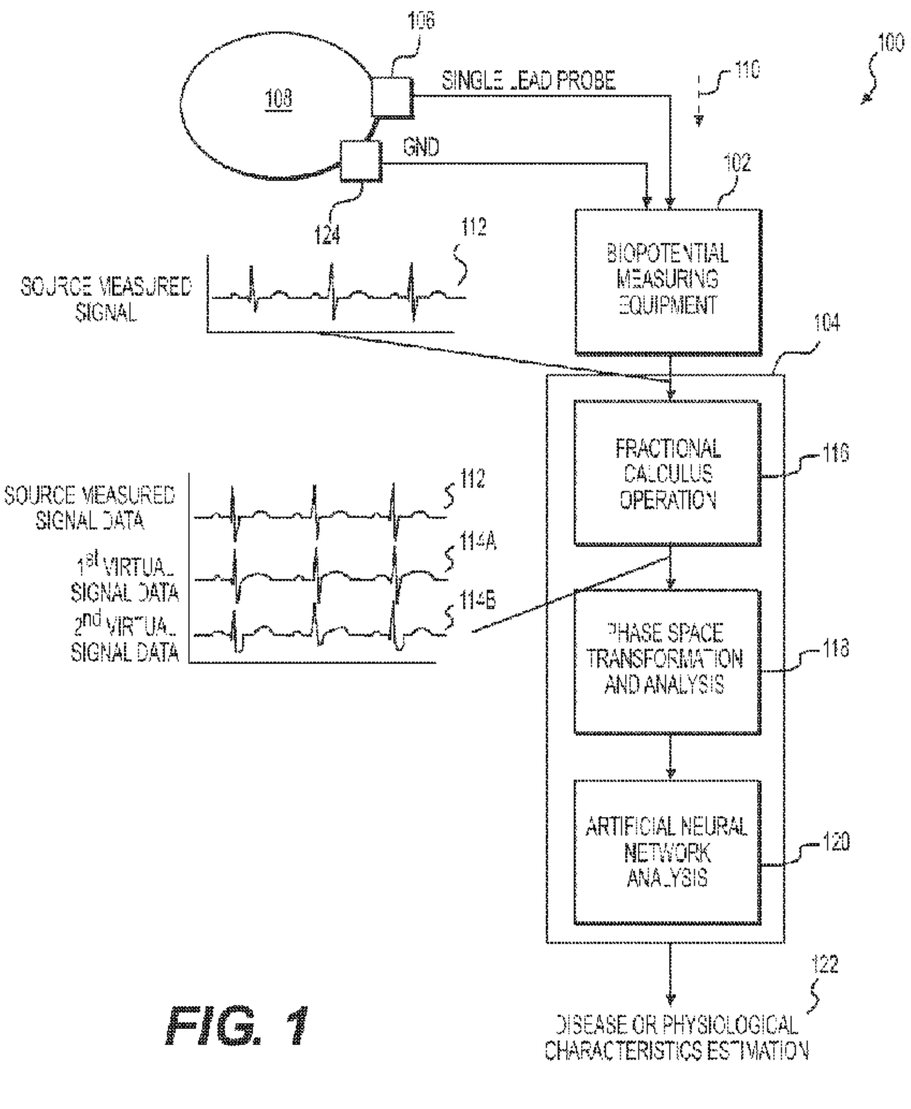

FIG. 1

(57) Abrégé/Abstract:
Exemplified method and system facilitates monitoring and/or evaluation of disease or physiological state using mathematical analysis and machine learning analysis of a biopotential signal collected from a single electrode. The exemplified method and system creates, from data of a singularly measured biopotential signal, via a mathematical operation (i.e., via numeric fractional derivative calculation of the signal in the frequency domain), one or more mathematically-derived biopotential signals (e.g., virtual biopotential signals) that is used in combination with the measured biopotential signals to generate a multi-dimensional phase-space representation of the body (e.g., the heart). By mathematically modulating (e.g., by expanding or contracting) portions of a given biopotential signal, in the frequency domain, the numeric- based operation gives emphasis or de-emphasis to certain measured frequencies of the biopotential signals, which, when coupled with machine learning, facilitates improved diagnostics of certain pathologies.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2016 (29.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/207862 A1

(51) International Patent Classification:
A61B 5/02 (2006.01) *G06F 15/18* (2006.01)
A61B 90/00 (2016.01) *G06F 19/00* (2011.01)

(21) International Application Number:
PCT/IB2016/053797

(22) International Filing Date:
24 June 2016 (24.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/184,796 25 June 2015 (25.06.2015) US

(72) Inventors; and
(71) Applicants : **BURTON, Timothy** [CA/CA]; 573 Evered Avenue, Ottawa, Ontario K1Z 5L1 (CA). **RAMCHANDANI, Shyamlal** [CA/CA]; 82 Beverley Street, Kingston, Ontario K7L 3Y6 (CA). **GUPTA, Sunny** [CA/CA]; 18 Miller Place, Amherstview, Ontario K7N 1X6 (CA).

(74) Agent: **TANPITUKPONGSE, Paul**; 999 Peachtree Street Ne, Suite 1300, Atlanta, Georgia 30309 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) Title: METHODS AND SYSTEMS USING MATHEMATICAL ANALYSIS AND MACHINE LEARNING TO DIAGNOSE DISEASE

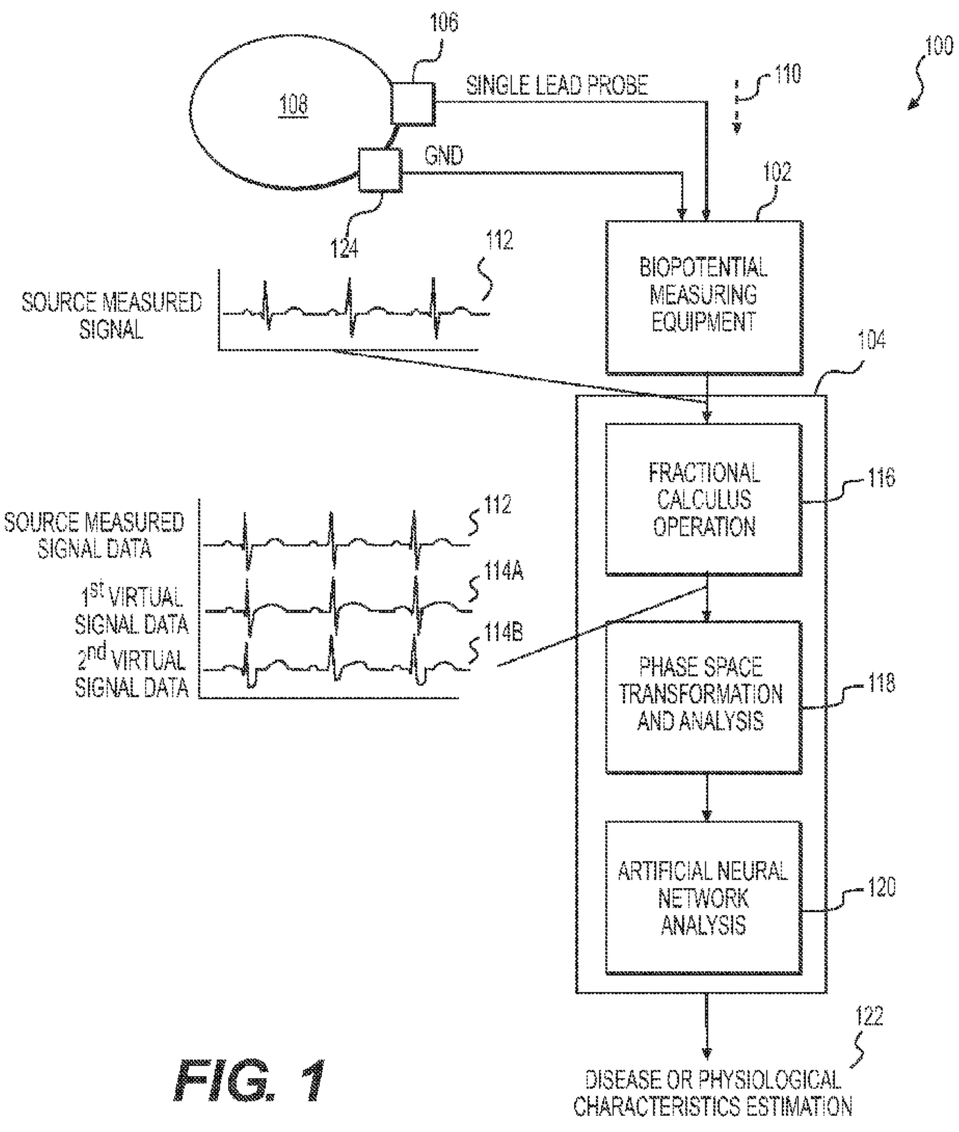

FIG. 1

(57) Abstract: Exemplified method and system facilitates monitoring and/or evaluation of disease or physiological state using mathematical analysis and machine learning analysis of a biopotential signal collected from a single electrode. The exemplified method and system creates, from data of a singularly measured biopotential signal, via a mathematical operation (i.e., via numeric fractional derivative calculation of the signal in the frequency domain), one or more mathematically-derived biopotential signals (e.g., virtual biopotential signals) that is used in combination with the measured biopotential signals to generate a multi-dimensional phase-space representation of the body (e.g., the heart). By mathematically modulating (e.g., by expanding or contracting) portions of a given biopotential signal, in the frequency domain, the numeric-based operation gives emphasis or de-emphasis to certain measured frequencies of the biopotential signals, which, when coupled with machine learning, facilitates improved diagnostics of certain pathologies.

WO 2016/207862 A1

METHODS AND SYSTEMS USING MATHEMATICAL ANALYSIS AND MACHINE LEARNING TO DIAGNOSE DISEASE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to, and the benefit of, U.S. Provisional Application No. 62/184,796, filed June 25, 2015, titled "Latent Teratogen-Induced Heart Deficits Are Unmasked Postnatally with Mathematical Analysis and Machine Learning on ECG Signals," the content of which is incorporated by reference herein in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to methods and systems to diagnose cardiac pathologies via mathematical and machine learning analysis on biopotential signals.

BACKGROUND

[0003] Congenital heart defects (CHDs) are the most common birth anomaly, with ventricular septal defects (VSDs) being the most prevalent category of congenital heart defects (CHDs). Clinically, about 80% of ventricular septal defects (VSDs) resolve spontaneously within the first year, but little is known about the long-term consequences of these resolved VSDs on postnatal heart function.

[0004] What are needed are devices, systems and methods that overcome challenges in the present art, some of which are described above.

SUMMARY

[0005] Exemplified method and system facilitates monitoring and/or evaluation of disease or physiological state using mathematical analysis and machine learning analysis of a biopotential signal collected from a single electrode. The exemplified method and system creates, from data of a singularly measured biopotential signal, via a mathematical operation (i.e., via fractional derivative calculation), one or more mathematically-derived biopotential signals (e.g., virtual biopotential signals) that is used in combination with the measured

biopotential signals to generate a multi-dimensional phase-space representation of the body (e.g., the heart). In some embodiments, the fractional derivative of a fraction of a real number or a fraction of an integer number is applied numerically to the measured biopotential signals, or a portion thereof, in the frequency domain, to increase the dimensionality of the measured signal data. In some embodiments, the fractional derivative of an integer is applied numerically to the measured biopotential signals, or a portion thereof, in the frequency domain, to increase the dimensionality of the measured signal data. By mathematically modulating (e.g., by expanding or contracting) portions of a given biopotential signal, in the frequency domain, the numeric-based operation gives emphasis or de-emphasis to certain measured frequencies of the biopotential signals, which, when coupled with machine learning, facilitates improved diagnostics of certain pathologies and facilitates diagnostics in real-time (or near real time).

[0006] To this end, the exemplified method and system facilitates the measurements of biopotential signals using a single electrode lead to obtain diagnostics results. In addition, the exemplified method and system may be used to enhance measurements collected by multiple physical leads by, in effect, multiplying the physical effects with virtual lead that provide a different vantage point or perspective from the original physical measurement that improves the subsequent analysis.

[0007] A clinical animal model study was performed, the study illustrates that the exemplified method and system facilitates clinically-relevant diagnosis of physiologic conditions. In particular, the clinical animal model study illustrates that resolved congenital heart defects harboring hidden cardiovascular dysfunction can be detected using the exemplified method and system.

[0008] In an aspect, a method is disclosed of pre-processing data to extract variables for use in machine learning to diagnose a pathology. The method includes receiving a biopotential signal data associated with a subject, said biopotential signal data being associated with a biopotential signal collected from one or more electrical leads; generating, via a processor, a first

and a second fractional derivative signal data by numerically performing one or more fractional derivative operations (e.g., a numeric fractional derivative operation) of the biopotential signal data in a frequency domain and converting a result of the one or more fractional derivative operations to a time domain signal data, wherein each of the first and second generated fractional derivative signal data comprises a same length and a same sampling frequency as the biopotential signal data; and generating, via the processor, a three-dimensional space wherein each corresponding value of the biopotential signal data, the first fractional derivative signal data, and the second fractional derivative signal data forms a three-dimensional point in said space, wherein geometric features and dynamical properties of the three-dimensional space are used as variables representative of the subject in machine learning to detect one or more diagnosable pathology of the subject.

[0009] In some embodiments, the first fractional derivative signal data is generated by performing a first numeric fractional derivative of a first order value on the biopotential signal data in the frequency domain and by performing an inversed transformation (e.g., inversed FFT) on the fractional derived signal data to convert the fractional derived signal data to a time domain signal data. In some embodiments, the inversed transformation comprises an inversed Fast Fourier Transform (inversed FFT) operation.

[0010] In some embodiments, the second fractional derivative signal data is generated by performing a second numeric fractional derivative of a second order value on the biopotential signal data in the frequency domain and by performing an inversed transformation (e.g., inversed FFT) on the fractional derived signal data to convert the fractional derived signal data to a time domain signal data.

[0011] In some embodiments, each of the first fractional derivative signal data and the second fractional derivative signal data comprises a time domain signal data.

[0012] In some embodiments, the first fractional derivative signal data is generated by a fractional derivative of an order of pi/2.

[0013] In some embodiments, the second fractional derivative signal data is generated by a fractional derivative of an order of 0.5.

[0014] In some embodiments, the geometric features and dynamical properties of the three-dimensional space are generated by performing a MMP (modified matching pursuit) algorithm of the three-dimensional point in said space.

[0015] In some embodiments, the biopotential signal data is associated with a biopotential signal collected from a single electrical lead.

[0016] In some embodiments, the single electrical lead collected measurements of the biopotential signal at a location selected from the group consisting of a chest line of the subject, a waistline of the subject, a wrist of the subject, a pelvic line of the subject, a neck of the subject, an ankle of the subject, a forehead of the subject, and an arm line of the subject.

[0017] In some embodiments, the method includes generating, via a processor, an alpha shape of the three-dimensional point in said space, wherein the geometric features and dynamical properties of the three-dimensional space includes the geometric features of the alpha shape.

[0018] In some embodiments, the method includes generating a Delaunay triangle mesh of the three-dimensional point in said space, wherein the geometric features and dynamical properties of the three-dimensional space includes the geometric features of the Delaunay triangle mesh.

[0019] In some embodiments, the biopotential signal data comprises electrocardiogram (ECG) data.

[0020] In some embodiments, the machine learning analysis comprises using an artificial neural network algorithm or a regression random forest algorithm.

[0021] In another aspect, a method is disclosed of pre-processing data to extract variables for use in machine learning to diagnose a pathology. The method includes receiving biopotential signal data associated with a subject, said biopotential signal data being associated with

biopotential signals collected from two or more electrical leads; generating, via a processor, a fractional derivative signal data by numerically performing one or more fractional derivative operations of at least one of the biopotential signal data in a frequency domain and converting a result of the one or more fractional derivative operations to a time domain signal data, wherein the generated fractional derivative signal data comprises a same length and a same sampling frequency as the at least one of the biopotential signal data; and generating, via the processor, a three-dimensional space wherein each corresponding value of each of the biopotential signal data and the fractional derivative signal data forms a three-dimensional point in said space, wherein geometric features and dynamical properties of the three-dimensional space are used as variables representative of the subject in machine learning to detect one or more diagnosable pathology of the subject.

[0022] In some embodiments, each of the two or more electrical leads collected measurements of the biopotential signal at a location selected from the group consisting of a chest line of the subject, a waistline of the subject, a wrist of the subject, a pelvic line of the subject, a neck of the subject, an ankle of the subject, a forehead of the subject, and an arm line of the subject.

[0023] In another aspect, a method is disclosed of determining congenital heart defects (CHD) in a mammal. The method includes receiving biopotential recordings associated with the mammal, the biopotential recordings being recorded at predetermined intervals; developing variables associated with the biopotential recordings to create a dataset; and analyzing the dataset to determine if the mammal has a CHD.

[0024] In some embodiments, the biopotential recordings associated with the mammal are recorded using a measuring equipment comprising a single surface lead.

[0025] In some embodiments, a measuring equipment comprises an intracardiac electrogram instrument.

[0026] In some embodiments, the measuring equipment comprises a smart watch or fitness heart band.

[0027] In another aspect, a system is disclosed that includes that includes remote storage (e.g., storage area network) configured to receive biopotential data from a network-connected biopotential measuring apparatus; one or more processors; and a memory having instructions stored thereon, wherein the instructions, when executed by the processor, cause the processor to: generate phase space variables associated with the biopotential data; analyzing the phase space variables to determine if the mammal has a CHD.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] FIG. 1 is a diagram of a system for pre-processing data to extract variables for use in machine learning to diagnose a pathology, in accordance with an illustrative embodiment.

[0029] FIG. 2 is a diagram of a method for processing data to diagnose a pathology, in accordance with an illustrative embodiment.

[0030] FIG. 3 is a detailed diagram of a method of pre-processing the data as shown in FIG. 2, in accordance with an illustrative embodiment.

[0031] Fig. 4 is a detailed diagram of a method of signal data normalization as shown in FIG. 3, in accordance with an illustrative embodiment.

[0032] FIG. 5 is a detailed diagram of a method of virtual signal generation as shown in FIG. 3, in accordance with an illustrative embodiment.

[0033] FIG. 6 is a detailed diagram of a method of performing phase space analysis as shown in FIG. 2, in accordance with an illustrative embodiment.

[0034] FIG. 7 is a diagram illustrating postnatal VSD resolution, in accordance with an illustrative embodiment, in accordance with an embodiment, in accordance with an embodiment.

[0035] FIG. 8 is a diagram illustrating hypothesis of a clinical study, in accordance with an embodiment.

[0036] FIG. 9 is a diagram illustrating a design of experiment for the clinical study described in relation to FIG. 8, in accordance with an embodiment.

[0037] FIG. 10 is a diagram illustrating a method for conducting the experiment for the clinical study described in relation to FIG. 9 using machine learning process, in accordance with an embodiment.

[0038] FIG. 11 is a diagram illustrating a vectorcardiogram including virtual biopotential signals generated from an exemplified fractional derivative operation, in accordance with an embodiment.

[0039] FIG. 12 is a diagram illustrating exemplary inputs to the machine learning processes described in relation to FIG. 10, in accordance with an embodiment.

[0040] FIG. 13 is a diagram illustrating a phase space model of a vectorcardiogram of a specimen with a diagnosable disease.

[0041] FIG. 14 is a diagram illustrating an alpha-shape phase space model of a vectorcardiogram of a control specimen.

[0042] FIGS. 15 and 16 are diagrams illustrating results of the clinical study as described in relation to FIGS. 7-10, in accordance with an illustrative embodiment.

[0043] FIGS. 17A and 17B show the performance of heart rate variability as assessed using a receiver-operator characteristic curve in the prediction of DMO-exposed or chemical naïve status, using either the standard deviation method, or Poincare Pearson correlation method.

[0044] FIG. 18 show the performance of the artificial neural network on distinguishing the DMO-exposed data from the chemical naïve data as assessed using a receiver-operator characteristic curve.

[0045] FIG. 19 shows the performance of the random forest on distinguishing the DMO-exposed data from the chemical naïve data within the "leave-one-out" validation paradigm, as assessed by a receiver operator characteristic curve.

DETAILED DESCRIPTION

[0046] The components in the drawings are not necessarily to scale relative to each other and like reference numerals designate corresponding parts throughout the several views:

[0047] FIG. 1 is a diagram of a system 100 for pre-processing data to extract variables for use in machine learning to diagnose a pathology, in accordance with an illustrative embodiment. As shown in FIG. 1, the system 100 includes a biopotential measuring equipment 102 and an analysis subsystem 104. The biopotential measuring equipment 102 collects a biopotential signal via a single lead electrode 106 (and corresponding GND lead 124) that is attached to the surface of a subject 108 (e.g., the skin of a test animal or a person). The biopotential measuring equipment may be any device configured to capture electrophysiological signal. In some embodiments, a Holter ECG monitor is used as the biopotential measuring equipment 102 for measuring and recording the biopotential signal.

[0048] In some embodiments, the single lead electrode 106 comprise a surface electrode that is placed directly, or indirectly, on the surface of the skin or body tissue to record electrical activity. In some embodiments, the single lead electrode 106 comprises electrodes that are integrated into wearable devices to contact the skin when the wearable device is wore or attached to a patient or subject.

[0049] It should be appreciate that other wiring topology may be used without departing from the spirit of the disclosure. In some embodiments, the GND lead is a common-mode return. In other embodiments, the GND lead may serve as a differential mode signal with lead 106. In other embodiments, the GND lead is referenced to (or returned through) the earth, the chassis, or a shield.

[0050] Biopotential signals, in some embodiments, are electric potential that is measured between points on a tissue. Examples of biopotential signal includes an ECG (electrocardiogram) signal that is used to assess electrical and muscular functions of the heart.

[0051] In some embodiments, the biopotential measuring equipment is a wearable device that is configured to measure and to record the biopotential signal. In some embodiments, the

wearable device is configured to be placed at a chest line of a subject, a waistline of the subject, a wrist of the subject, a pelvic line of the subject, a neck of the subject, an ankle of the subject, a forehead of the subject, and an arm line of the subject and has electrodes positioned to be in proximal contact with the skin or surface of the wearer. The wearable device may have a housing in the form of watch, an arm band, a neck band, a leg band, a chest band, a head band and such. In other embodiments, the biopotential measuring equipment is a part of an exercise equipment (e.g., a handle bar), a weight scale, a mat, or any like device that contacts the skin or surface of a person.

[0052] Referring still to FIG. 1, the biopotential signals 110 are stored as biopotential signal data 112. The analysis system 104 receives the biopotential signal data 112, in some embodiments, over a network, from the biopotential measuring equipment 102. In some embodiments, the analysis system 104 receives the biopotential signal data 112 from a storage area network (SAN). In other embodiments, the analysis system 104 and biopotential measuring equipment 102 are located a single device, e.g., a wearable device. Other configurations may be used.

[0053] Referring still to FIG. 1, the analysis system 104 is configured to generate, from the source biopotential signal data 112 of a single signal, one or more additional biopotential signal data (shown as a first fractional derivative signal data 114a and a second fractional derivative signal data), via a fractional calculus operation 116, of the source biopotential signal data 112, where each of the first and second generated fractional derivative signal data 114a, 114b comprises a same length and a same sampling frequency as the biopotential signal data. In some embodiments, the numerical fractional derivative operation is performed to emphasize or deemphasize certain frequency components and such that there is an absence of orthogonality in the resulting vectors. The additional biopotential signal data are used in conjunction with the source biopotential signal data to generate a phase space map to be used in subsequent phase space analysis 118 later described herein. The output of the phase space analysis are then

evaluated using artificial neural networks 120 to assess parameters 122 associated with a presence of a disease or physiological characteristic. The output of the processor is then transmitted to a graphical user interface for visualization. The graphical user interface, in some embodiments, is included in a display unit configured to display parameters 122. In some embodiments, the graphical user interface displays intermediate parameters such as a 3D phase space plot representation of the biopotential signal data and virtual biopotential signal data.

[0054] FIG. 2 is a diagram of a method 200 of FIG. 1 for processing data to diagnose a pathology, in accordance with an illustrative embodiment. As shown in FIG. 2, the method 200 includes collecting 200 signal data and pre-processing 204 the signal data to generate a phase space dataset to be used in phase analysis 206, whereby features of the phase space dataset are extracted and evaluated via an artificial neural network analysis 208.

[0055] FIG. 3 is a detailed diagram of the pre-processing 204 of the data as shown in FIG. 2, in accordance with an illustrative embodiment. The pre-processing 204 includes, in some embodiments, signal data normalization 302 and virtual signal generation 304.

[0056] FIG. 4 is a detailed diagram of a method of signal data normalization as shown in FIG. 3, in accordance with an illustrative embodiment. As shown in FIG. 4, the signal normalization 302 includes an initial step of data channel removal 402 and Hurst exponent filtering 404 of the source biopotential signal data 112. Further detail of the data channel removal and Hurst exponent filtering is described in Matteo et al., "Scaling behaviors in differently developed markets," *Physica A, 324*, pg. 183-188 (2003). In some embodiments, the signal normalization 302 further includes selecting 406 a cleanest segment of the filtered signal, the segment having a minimized residue between a wavelet model, of the signal, that is designed to detect the presence of a non-biological noise and the signal itself. For example, a segment comprising a cleanest 5-second window of a 10 second recording interval may be selected. In some embodiments, the wavelet model is configured to decompose the signal into temporal levels in wherein one or more of the highest levels of decomposition of the temporal levels are

used in the subtraction of the wavelet model from the signal to determine the residue. In some embodiments, the signal normalization 302 further includes filtering 408 using a second wavelet model to remove undesired noise (e.g., any remaining noise) in the selected cleaned segment. The second wavelet model is configured decompose the signal into a number of temporal levels and one or more of the highest level of decomposition are maintained. It should be appreciated by one skilled in that art that other signal data normalization may be used.

[0057] FIG. 5 is a detailed diagram of a method of virtual signal generation as shown in FIG. 3, in accordance with an illustrative embodiment. As shown in FIG. 5, the virtual signal generation 304 includes creating one or more virtual signal data such that the data interact with the original signal data to create a valid phase space portrait in which the limit cycles of the input biopotential signal (e.g., cardiac cycle) are overlaid in 3-dimensional space and there was an absence of orthogonality in the resulting vector. For example, where a single source biopotential signal data is available, the virtual signal generation 304 may be used to generate two virtual biopotential signals where a valid phase space portrait is in 3 dimensional space. In another example, where two source biopotential signal data are available, the virtual signal generation 304 may be used to generate a virtual biopotential signal where a valid phase space portrait is in 3 dimensional space. In some embodiments, the valid phase space portrait may be in dimensional space greater than 3, such as 4, 5, 6, 7, 8, 9, 10, or greater.

[0058] In some embodiments, each of the one or more virtual signal data is generated by performing a Fast Fourier Transform 502 on the normalized signal data. A numerical fractional derivative is then performed 504 on each of the FFT signal data and an inversed Fast Fourier Transform (inversed FFT) is performed 506 on that output. Examples of the order of the fractional derivative include pi/2 or 0.5. In some embodiments, the order of the fractional derivative is a fraction of a real number or complex number. In some embodiments, the order of the fractional derivative is a fraction of an integer. In some embodiments, the output of the inversed FFT is further processed to remove 508 baseline wander.

[0059] In some embodiments, each of the one or more virtual signal data is generated by performing a numerical fractional differencing on the normalized signal data, implemented through the use of a convolution. A rational transfer function is defined to correspond to the specified order of the fractional derivative, which is then applied to the input data through the use of a digital filter configured to accept such input. Examples of the order of the fractional derivative include pi/2 or 0.5. In some embodiments, the order of the fractional derivative is a fraction of a real number or complex number. In some embodiments, the order of fractional derivative is a fraction of an integer. In some embodiments, the output of the convolution is further processed to remove 508 baseline wander.

[0060] FIG. 6 is a detailed diagram of the method of performing phase space analysis as shown in FIG. 2, in accordance with an illustrative embodiment. In some embodiments, the input to the phase space analysis is a point-cloud phase space map 601 of the biopotential signal data and corresponding virtual signal data as a vectorcardiogram. In some embodiments, the phase space analysis includes performing modified matching pursuit (MMP) algorithm 602. The MPP algorithm 602 may be used to generate a sparse mathematical model 604. Detail of the MMP algorithm is provided in Mallat *et al.*, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, Vol. 41 (12), Pages 3397-2415 (1993).

[0061] Characteristics of this model may be extracted, in a feature extraction operation 606, to determine geometric and dynamic properties of the model.

[0062] In some embodiments, the point-cloud phase space map 601 is encapsulated by alpha shape or a Delaunay triangulation. Features of the alpha shape and/or triangulation may be extracted, via feature extraction 610, to determine additional geometric and dynamic properties of the model.

[0063] In some embodiments, the extracted geometric and dynamic properties of the alpha shape and modified matching pursuit operations are use as variables to an artificial neural network analysis, a regression random forest analysis, or other machine learning analyses.

[0064] Experiment - Latent Teratogen-Induced Heart Deficits

[0065] Using an animal model described below, the exemplified system and method of using mathematical analysis and machine learning to diagnose disease is shown. Specifically, the exemplified system and method employs fractional calculus to increase the dimensionality of a single lead ECG via, a numerical method with the inverse FFT to be used in subsequent phase space analysis.

[0066] The exemplified system and method was shown to successfully extract meaningful variables from biopotential signals, in an animal model, specifically those from electrocardiogram data collected from an implanted radiotelemeter, at a time when teratogen-exposed test animals are clinically indistinguishable from controls. Machine learning was then leveraged to predict, within a robust validation framework, the presence of latent cardiovascular dysfunction. The test illustrates that the exemplified system and method can be used to analyze data (e.g., single time series data) from single lead measurements and to generate higher level of dimensional phase space data in a subsequent phase space analysis.

[0067] In the animal model, test rats were treated in such a way as to induce a high incidence of CHD (congenital heart defects) in offspring. Dams delivered naturally, and the heart structure and function were assessed in female pups using echocardiography on postnatal day (PND) 4, PND 21 and PND 56. At postnatal day (PND) 56, radiotelemetry units were implanted into 9 treated rats and 8 control rats. Two weeks post-surgery, telemeters were activated and ECG recordings were continuously collected every 10 seconds and every 12 seconds over a period of two weeks. 50,000 collected data points per rat were each transformed, from the single ECG recording, into a unique three-dimensional phase space dataset, and machine learning used to create predictive algorithms capable of identifying differences in heart function between control and treated rats and other mammals.

[0068] The results, as described in more detail below, demonstrate that a teratogen-induced CHD that resolves postnatally and results in hearts that appears normal by conventional

measures are, in fact, different from teratogen-naïve hearts. As equally important, the results reveal that fractional calculus may be used to increase the dimensionality of a single lead biopotential signal for use in phase space analysis.

[0069] Experiment Setup

[0070] FIG. 7 is a diagram illustrating postnatal ventricular septal defect (VSD) resolution, in accordance with an illustrative embodiment. As shown in FIG. 7, two cohorts of animals are selected with significant cardiovascular differences that decline to an absolute minimum at 8-10 weeks postnatal when these differences in the animals are clinically indistinguishable. The animals are used to verify that the exemplified system and method may be used to identify which animals were affected by drug administration (i.e., during this 8-10 week postnatal period when the cardiovascular differences in the animals are clinically indistinguishable).

[0071] FIG. 8 is a diagram illustrating hypothesis of a clinical study, in accordance with an embodiment. As shown in FIG. 8, the hypothesis is that exemplified mathematical analysis and machine learning will reveal functional deficits between the cohort of offspring with treated dams and the cohort of offspring with non-treated dams at a time when it is otherwise undetectable by conventional analysis (during the 8-10 week period) without the need for a conventional stressor.

[0072] FIG. 9 is a diagram illustrating a design of experiment for the clinical study described in relation to FIG. 8, in accordance with an embodiment. As shown in FIG. 9, the test rats were divided into two cohorts: exposed and unexposed to dimethadione (DMO). The test rates were allowed to come to natural parturition; the pups were evaluated over the first 8 weeks of life using echocardiography. Then, at 8 weeks, radiotelemeters were surgically implanted into the test rats to measure several physiological signals including ECG signals. The ECG signals were then used in machine learning to create a predictor to discriminate between the two cohorts.

[0073] FIG. 10 is a diagram illustrating a method for conducting the experiment for the clinical study described in relation to FIG. 9 using machine learning process, in accordance with an embodiment. As shown in FIG. 10, the exemplified experiment begins with test rats with known exposure paired to ECG data. Variables were then developed from the ECG data, and input to a machine learning process. The exemplified method and system is used to predict if the rat that generated that ECG was exposed or unexposed to dimethadione (DMO).

[0074] FIG. 11 is a diagram illustrating a vectorcardiogram including virtual biopotential signals generated from an exemplified fractional derivative operation, in accordance with an embodiment. As shown in FIG. 11, the exemplary inputs include the original measured ECG signal data 112 and the virtual ECG signal data 114a, 114b.

[0075] FIG. 12 is a diagram illustrating exemplary inputs to the machine learning processes described in relation to FIG. 11, in accordance with an embodiment. As shown in the FIG. 12, a phase space dataset 1202 is shown of the vectorcardiogram (VCG) that includes the measured biopotential signals 112 and the virtual biopotential signals 114a, 114b generated from the exemplified fractional derivative operation. The vectorcardiogram (VCG) is shown in a point-cloud phase space map 1202 in which the measured biopotential signals 112 and the virtual biopotential signals 114a, 114b are shown without time in a three axis coordinate system. The phase space dataset 1202 is quantified by being wrapped in a geometric shape, for example, an alpha shape or a Delaunay triangle. Detail of an alpha shape operation is described in Edelsbrunner *et al.*, "Three-dimensional alpha shapes," ACM Transactions on Graphics, Vol. 13 (1): 43-72 (1994).

[0076] FIG. 13 is a diagram illustrating an alpha-shape phase space model of a vectorcardiogram of a specimen with a diagnosable disease. FIG. 14 is a diagram illustrating an alpha-shape phase space model of a vectorcardiogram of a control specimen. As shown in FIGS. 13 and 14, there are visual differences between datasets of the exposed and unexposed test rats,

but these differences are subtle and are more conducive to be analyzed via machine learning to expose such differences.

[0077] In some embodiments, the machine learning analysis uses families of variables including data shape variables, fractional derivatives, signal modeling using MMP, phase space modeling, and dynamical system variables. In the exemplified experiment, with data for 17 test rats, each with 50,000 ECG signals, where each ECG signal has 250 variables, there are about two billion data points used in the analysis including an artificial neural network.

[0078] Discussion

[0079] Congenital heart defects (CHD) are the most common class of congenital anomalies with an incidence of approximately 1.9-7.5% of live births. The most prevalent birth defect is the ventricular septal defect (VSD) at 25-40% of all CHD, in which the septum fails to close between the left and the right ventricles. Postnatal persistence or surgical repair of VSD or other CHD significantly increases the lifetime risk of heart disease in these patients relative to unaffected cohorts, and therefore necessitates a vigilant lifetime of observation and potential intervention by a cardiologist specializing in the care of CHD. Interestingly, approximately 80% of VSDs that present clinically resolve spontaneously within the first year of life. While a clinical resolution of the structural damage is a favorable outcome, there is a dearth of information about the long-term functional consequences of resolved VSD.

[0080] To explore the potential deleterious long-term consequences of resolved CHDs, the exemplified rat model was developed that recapitulates many of the clinical presentations of CHD. To generate such a model, pregnant rats were treated with a chemical teratogen. It has been estimated that 80% of in utero trimethadione exposures resulted in embryo/fetal loss, or malformations including a high incidence of CHD. Without wishing to be bound to particular theory, it is thought that DMO teratogenicity is mediated by oxidative stress which might be the result of the result of hypoxia reperfusion injury caused by bradycardias induced by disruption of calcium and potassium ion channels in embryonic myocytes. It has been demonstrated in

outbred pregnant Sprague-Dawley rats that the incidence and severity of CHD in progeny are highly dependent upon the gestational window and total exposure to DMO. The administration of four 300 mg/kg doses every twelve hours beginning on the morning of gestation day (GD) 9 produces approximately a 50% incidence of CHD if fetuses are examined on GD 21, one day prior to natural parturition (*see, e.g.*, Weston et al., "Co-variation in frequency and severity of cardiovascular and skeletal defects in Sprague-Dawley rats after maternal administration of dimethadione, the N-demethylated metabolite of trimethadione. Birth defects research Part B," *Developmental and reproductive toxicology*, Vol. 92, Pages 206-15 (2011)). When similarly treated dams are allowed to deliver their progeny naturally and the cardiac structure and function of the pups is evaluated longitudinally by high-resolution echocardiography, a scenario reminiscent of the clinical presentation of CHD is revealed. For example, in infants, approximately 80% of VSD resolves by one year of age, and in test rats after DMO exposure, about 80% of the VSD resolve spontaneously by weaning. Other structural elements such as left ventricular (LV) mass, LV anterior wall thickness in systole (LVAW;s) and LVAW in diastole (LVAW;d) are all significantly different between control and DMO-exposed pups close to the time of parturition, but resolve over time such that by 10 weeks of age control and DMO animals are indistinguishable. In the rat pups exposed to DMO, cardiac dysfunction is pronounced close to parturition, but gradually resolves, such that by 10 weeks of age under basal, during unstressed conditions, cardiac output (CO), stroke volume (SV), ejection fraction (EF), fractional shortening (FS) and pulmonary artery regurgitation PA regurg. (mm/s) have all normalized. At this point, radiotelemeters capable of measuring continuous single channel electrocardiogram (ECG) were surgically implanted into the rats. After two weeks of surgical recovery, baseline heart function was obtained and the animals were mated. Pregnancy is a physiological challenge for the maternal cardiovascular system and clinically has been referred to as a cardiovascular "stress test". Pregnancy-induced changes to the mammalian maternal CV system include a 30-40% increase in blood volume, 30-60% increase in cardiac output (CO), transient cardiac

hypertrophy and uterine spiral arterial (SA) remodeling ref. Cardiac hypertrophy occurring in normal pregnancy is physiological and reversible, similar to exercise-induced hypertrophy.

[0081] Disconcertingly, the cardiovascular systems of test rats exposed to DMO *in-utero* adapted poorly to the burden of pregnancy. Manifestations included altered CO, SV, radial and longitudinal strain, and elevated mean arterial pressure at the time of spiral artery remodeling. Thus, the function of hearts with resolved CHD were indistinguishable from control hearts under basal unstressed conditions; however, under the stress of pregnancy profound cardiac and hemodynamic deficiencies emerged.

[0082] In the exemplified experiment, *in-utero* exposure to the heart teratogen DMO was used to generate a population of rats with resolved CHD that only presented with cardiac dysfunction during the burden of pregnancy. The exemplified method and system was used to predict the presence of latent teratogen-induced cardiac functional deficits prior to the onset of pregnancy, without the use of a cardiovascular stress test, and using the ECG data collected via the telemeter within the ten to twelve week period of the study previously described. There are no discernable differences between the cohorts during this ten to twelve week period.

[0083] Experiment Results

[0084] FIGS. 15 and 16 are diagrams illustrating results of the clinical study as described in relation to FIGS. 7-10, in accordance with an illustrative embodiment. Specifically, FIG. 15 is a plot of a predictive output of the neural network analysis for the two cohorts of test animals, which were by conventional measures, indistinguishable. The plot shows all the predictions on the treated and control cohorts test data (recall the 85%), where there is a clear visual separation between the baselines. For comparison, a heart-rate variability (HRV) analysis was run that uses landmarks in the ECG signal data to automatically detect high content data. As shown, HRV was not successful in identifying the cohorts, while the exemplified method and system (using the machine learning predictor) was found to be significantly better, showing that it is possible to discriminate between these two groups, and therefore they are different.

[0085] Specifically, the exemplified experiment was conducted with two cohorts in which five test rats were exposed to DMO *in utero* (Cohort #1), and eight control rats were chemical naïve (Cohort #2). Raw ECG signals were recorded from a single channel with a sampling frequency of 1000 Hz. Data acquisition was blocked into ten-second intervals from the test rats for a two week period between the ages of eleven to fourteen weeks, when the two cohorts were otherwise clinically indistinguishable using echocardiography or heart rate variability (HRV). The statistical evaluation of the predictors used to discriminate between the cohorts, such as HRV, primarily utilized area under the receiver operator characteristic curve (AUC). AUC examines the performance of a predictor against a binary target variable, which in this case, is the DMO-exposed or chemical naïve status of the rat. An AUC value of 0.5 represents a random prediction, an AUC of 1 is a perfect prediction, and an AUC of 0 is a perfect, but inverted, prediction (where all data points of one class are predicted to be the other class, and *vice versa*).

[0086] HRV reflects beat-to-beat changes in heart rate reflecting dynamic changes in autonomic tone. HRV is a clinically useful electrophysiological endpoint and is the foundation for risk stratification strategies such as REFINE, which seeks to determine whether non-invasive physiological parameters collected after the occurrence of a myocardial infarction (MI) predict the subsequent incidence of cardiac death or resuscitated cardiac arrest. In the REFINE study, HRV had an AUC of 0.62 and hazard ratio of 2.15 in the 10-14 week period post-MI, proving to be a somewhat useful technique for risk stratification, which led us to believe it might provide utility in the identification of DMO-exposed rats.

[0087] FIGS. 17A and 17B show the performance of heart rate variability as assessed using the receiver-operator characteristic curve in the prediction of DMO-exposed or chemical naïve status, using either the standard deviation method (FIG. 17A), or Poincare Pearson correlation method (FIG. 17B). FIGS. 17A and 17B were generated as follows. A five-second window that minimized non-biological noise was identified in a ten-second recording interval of

a given test rat. Heartbeats were identified during this five-second window, and a confidence score (CS) determined; five-second windows possessing identified heartbeats with a CS greater than 0.7 were included in statistical evaluation. After heartbeat identification, HRV was calculated using the most widely accepted approach in which the standard deviation of the R-R interval (referencing the standard PQRST notation to represent the five waveforms in the cardiac cycle) was calculated for each five-second window. Additionally, a Poincaré plot was created in which the length of the R-R interval and the length of the next R-R interval create a point in two-dimensional scatter plot and the Pearson correlation of this data is calculated. The performance statistics of these two HRV techniques are reported in Table 1, with a threshold applied for statistics and the corresponding receiver operator characteristic curves are plotted in FIGS. 17A and 17B.

[0088] Table 1 shows statistical performance of heart rate variability in the prediction of DMO-exposed or chemical naïve status. The AUCs of these two predictors demonstrate that quantifying HRV using either method has low or no predictive power. The poor performance is additionally reflected in the diagnostic odds ratio (DOR), which is only slightly above 1 (where 1 indicates there is no change in the relative odds of the rat being DMO-exposed, given a positive test result).

Table 1

Statistic	HRV: Standard Deviation	HRV: Poincare Pearson Correlation
AUC	0.51	0.59
Sensitivity	78%	2%
Specificity	23%	99%
Positive Predictive Value	39%	48%
Negative Predictive Value	62%	62%
Diagnostic Odds Ratio	1.05 (95% CI 1.03-1.07)	1.47 (95% CI 1.37-1.58)
Number of rats correctly classified	4/13	8/13
Percent of intervals correctly classified	44%	61%

[0089] FIG. 18 show the performance of the exemplified method suing artificial neural network on distinguishing the DMO-exposed data from the chemical naïve data, on the validation data when 15% of the data is allocated to the training set and the remaining 85% is allocated to validation set, as visualized by the receiver operator characteristic curve. In addition, Table 2 shows the statistical performance of the artificial neural network on distinguishing the DMO-exposed data from the chemical naïve data, on the validation data when 15% of the data is allocated to the training set and the remaining 85% is allocated to validation set.

Table 2

Statistic	**Value**
AUC	0.79
Sensitivity	70%
Specificity	73%
Positive Predictive Value	58%
Negative Predictive Value	82%
Diagnostic Odds Ratio	6.26 (95% CI 6.16-6.35)
Number of rats correctly classified	12/13
Number of intervals correctly classified	72%

[0090] ANNs are widely used in machine learning as they enable hitherto unachievable artificial intelligence benchmarks. A supervised ANN learns to predict a target using a vector of inputs in a way that mimics human neural processing. The ANN contains neurons arranged in a series of connected layers; firstly, the input layer that accepts the features, then one or more hidden layers to capture the non-linearity of the function being modeled, followed by the output layer that predicts the value of the target which corresponds to the input vector. The neurons in the ANN are stimulated by the input vector, and transmit those stimuli downstream to following

layers depending on the value of the features, the strength of the connection between the neurons, and the activation function found within the neuron. ANNs may be used as a universal functional approximator.

[0091] To explore the possibility of creating a more robust prediction model, an ensemble machine-learning algorithm based on random forest was used which relies on the principle that a series of weak learners, or simple algorithms that describe a limited amount of complexity in the data, when used in combination have powerful generalization properties. Random forest is composed of decision trees as weak learners. Decision trees are tree structures where the nodes are decisions (or specific to this application, a threshold on a feature), and upon taking a decision the connection to that child node is followed, which reveals a sub-decision tree with the original child as the root. The process continues until arrival at a leaf node, which is the final prediction on the data. Each decision tree is exposed to a different partition of the both the recording intervals and the feature vector, and therefore tends to learn a distinct facet of the problem. A regression random forest algorithm with 100 decision trees was trained using the validation strategy just described, the results of which are shown in Table 3, and visualized in FIG. 19.

[0092] FIG. 19 shows the performance of the random forest on distinguishing the DMO-exposed data from the chemical naïve data within the "leave-one-out" validation paradigm, as assessed by the receiver operator characteristic curve. As shown in FIG. 19, the performance is diminished slightly as compared to that of the first validation strategy as described in relation to FIG. 18, but in the context of this robust validation strategy it is still highly predictive. The performance is the result of analyzing the predictions produced in amalgamation by thirteen random forests (one for each of the thirteen test animals), each predicting on only the intervals of the single rat it did not receive in the training phase. The threshold, for those statistics requiring it, was set using the boundary that provided the maximal classification accuracy. When

reserving a DMO-exposed rat for testing, the training set is composed of the 8 chemical naïve rats and the remaining 4 DMO-exposed rats, resulting in a ratio of 2:1.

[0093] Table 3 shows the statistical performance of the random forest on distinguishing the DMO-exposed data from the chemical naïve data within the "leave-one-out" validation paradigm.

Table 3

Statistic	**Random Forest**
AUC	0.73
Sensitivity	40%
Specificity	99%
Positive Predictive Value	97%
Negative Predictive Value	70%
Diagnostic Odds Ratio	80.4 (95% CI 77.0-84.0)
Number of rats correctly classified	10
Number of intervals correctly classified	74%

[0094] Experiment - Material and Methods

[0095] The analytical methodology described herein used pre-processing in the transition from data collection to feature extraction. Data channel removal and Hurst exponent filtering are the initial operations. The ideal threshold on which to accept data was observationally determined to be 0.7 on that exponent, which ranges from 0 to 1, through the visual inspection of a representative subset of the ECG signals along with the calculated exponents of those signals. Second, the cleanest 5-second window out of the 10-second recording interval was found by selecting the segment that minimizes the residue between a wavelet model of the signal, designed to detect the presence of non-biological noise, and the signal itself. The wavelet model is computed using the functionality of the MATLAB™ (MATHWORKS; Natick, MA) Wavelet Toolbox. Data outside of this 5-second segment is discarded.

[0096] In some embodiments, the filtering provides a clean signal that lend better to numerical operators (e.g., numerical fractional derivative operations). It should be appreciated by those skilled in the art that other filtering and signal cleaning operations may be used.

[0097] Following this, any remaining noise in this 5-second interval is removed using a second wavelet model, which is designed for noise removal rather than noise detection, but is otherwise similar to the model used for the selection of the 5-second segment. Both of the wavelet models decompose the signal into eight temporal levels, but the noise detection model only preserves the two highest levels of decomposition (resulting in efficient capture of noise when subtracting the wavelet model from the signal), while the noise removal model at least partially maintains the four highest levels of decomposition. The phase space reconstruction, and therefore the creation of the vVCG, was then performed through the creation of the two virtual signals. Transformations to create the two virtual signals were chosen such that the signals interact with the original signal to create a valid phase space portrait, where the limit cycles of the cardiac cycles were overlaid in 3-dimensional space and there was an absence of orthogonality in the resulting vectors. These virtual leads were created by taking derivatives of the acquired lead in such a way as to create a valid phase space portrait, or a shape in three-dimensional space where the values of each of the three leads at a given time form a three-dimensional point in that space. Fractional derivatives of order pi/2 and 0.5 were found to be suitable, as computed numerically through the conversion of the signal into the frequency domain using the Fast Fourier Transform (FFT), calculation of the required derivative within the frequency domain, and conversion of the derivative back to the time domain using the inverse FFT. The baseline wander was then extracted from each of the three dimensions through the use of a median filter with an order of 1500 ms, smoothed with a 1-Hz low-pass ideal filter, and subtracted from the signals. The bias was then removed from the resulting signals by subtracting estimations of the modes of the signals using the maximums of the probability densities

calculated with a kernel smoothing function. Finally, all signals were divided by their interquartile ranges to complete the normalization process.

[0098] The three-dimensional space construction is subsequently used to study geometrical and dynamical properties of the system. The pre-processed signal is subjected to a feature extraction process. The signal is modeled with the modified matching pursuit (MMP) algorithm to create a sparse mathematical model. Characteristics of the model, including residue quantification, were included in the feature set. The vVCG was further quantified by creating an encapsulating alpha shape, or a specific Delaunay triangulation. This triangulation has associated characteristics that composed the feature vector representing the recording interval.

[0099] Delaunay triangulations are triangulations on a set of points such that no point is within the circumcircle of any triangle in the triangulation, and the minimum angle of all the angles in each triangle in the triangulation is maximized.

[0100] An alpha shape adds a further constraint; this triangulation requires the specification of an alpha radius, and only pairs of points whose distance is less than the alpha radius may be connected by an edge. The alpha radius in this feature extraction process was determined observationally to be 0.6, which allowed for sufficient encapsulation of the vVCG while still creating appropriate sparsity in the triangulation in areas of reduced point density.

[0101] This feature extraction was performed using SHARCNET (a consortium of colleges, universities and research institutes operating a network of high-performance computer clusters across southwestern, central and northern Ontario, Canada).

[0102] At the conclusion of the feature creation process, there were 250 features to represent the 10-second recording interval.

[0103] After the feature extraction, the ANN and random forest algorithms were invoked to create the predictors. An ANN was trained using the labeled feature vectors from 15% of the recording intervals, where the relatively small training percentage was chosen to minimize the potential to overfit and allow for generalizability. Selecting a relatively small segment of the

available data for training, and performing well on the larger test set, requires that the ANN leverage overarching signatures in the data rather than patterns unique to the training set. The ANN had an input dimensionality of 250 neurons, a single hidden layer that expanded that dimensionality by a factor of three to a total of 750 neurons, and then a single neuron in the output layer to represent the prediction of DMO-exposed or chemical naïve. Each neuron in the ANN contained the hyperbolic tangent activation function. The ANN sought to minimize the root mean squared error between the prediction and the recording interval class on the training data using stochastic gradient descent. Low learning rate and momentum parameters were used to allow the ANN to evolve gradually, and input corruption (the addition of noise on the incoming features) as well as dropout (noise internal to the ANN) were used to control for overfitting, which occurs when the ANN is highly specific to the training data and is unlikely to generalize to novel data. The ANN was a custom implementation in the MATLAB language. The performance of this ANN when it was applied to the withheld 85% of the data (351,520 data points in total) is shown below in Table 2, and the corresponding ROC curve is plotted in FIG. 18.

[0104] With respect to the random forest, the MATLAB function TreeBagger was used in which the specification of the number of features to sample at each node in the tree was selected from all the features, thereby exploiting the varying levels of information present within the features, and yet gain the generalization benefit from the random sampling of the training data for each member of the random forest.

[0105] Heart rate variability was also calculated on the pre-processed signals. The detection of the R peaks was performed through the identification of high-confidence R peaks, or those that exist in the top decile of the signal, and the creation of templates based on these high-confidence peaks. A template matching procedure was then executed on the signal, identifying the peaks of segments of the signal with a high degree of similarity to the templates, resulting in an expanded set of R peaks. The confidence of these R peaks was then quantified by calculating

the absolute difference between the maximum R-R interval and the minimum R-R interval, and dividing by the maximum R-R interval in order to create a normalized peak confidence score. The threshold on the score was set to 0.7 (not to be confused with the generalized Hurst exponent threshold, also set to 0.7), in order to allow for the heart rate variability that is to be expected in any mammalian cardiovascular system, yet remove any peak detections that are likely to have missed a peak or identified an additional peak, and thus depressing the confidence score. HRV can then be calculated on R-R intervals derived from the identified peaks, using the methods already described.

[0106] Although the invention is described in terms of particular embodiments and applications, one of ordinary skill in the art, in light of this teaching, can generate additional embodiments and modifications without departing from the spirit of or exceeding the scope of the exampled invention. Accordingly, it is understood that the drawings and descriptions herein are proffered by way of example to facilitate comprehension of the invention and should not be construed to limit the scope thereof.

[0107] For example, in view of the exemplified method and system, a single electrode lead measurements may be used in analysis that conventionally use data from multiple lead measurements, e.g., to assess certain physiologic characteristics or disease.

[0108] In addition, if the current observations in rat are clinically translatable, there are important implications with respect to the long-term cardiac health after spontaneous post-partum resolution of CHD. For example, the ability to identify latent markers of future cardiac dysfunction using only ECG signals would be a significant cost-effective step forward for the identification of at-risk individuals who typically do not outwardly display an inherent cardiac risk.

[0109] In addition, the animal model upon which these exemplified experiments are based, demonstrate that there are important implications with respect to the long-term cardiac health after spontaneous post-partum resolution of CHD.

What is claimed is:

1. A method of pre-processing data to extract variables for use in machine learning to diagnose a pathology, the method comprising:

receiving a biopotential signal data associated with a subject, said biopotential signal data being associated with a biopotential signal collected from one or more electrical leads;

generating, via a processor, a first and a second fractional derivative signal data by numerically performing one or more fractional derivative operations of the biopotential signal data in a frequency domain and converting a result of the one or more fractional derivative operations to a time domain signal data, wherein each of the first and second generated fractional derivative signal data comprises a same length and a same sampling frequency as the biopotential signal data; and

generating, via the processor, a three-dimensional space wherein each corresponding value of the biopotential signal data, the first fractional derivative signal data, and the second fractional derivative signal data forms a three-dimensional point in said space, wherein geometric features and dynamical properties of the three-dimensional space are used as variables representative of the subject in machine learning to detect one or more diagnosable pathology of the subject.

2. The method of claim 1, wherein the first fractional derivative signal data is generated by performing a first numeric fractional derivative of a first order value on the biopotential signal data in the frequency domain and by performing an inversed transformation (e.g., inversed FFT) on the fractional derived signal data to convert the fractional derived signal data to a time domain signal data.

3. The method of claim 2, wherein the inversed transformation comprises an inversed Fast Fourier Transform (inversed FFT) operation.

4. The method of any one of claims 1-3, wherein the second fractional derivative signal data is generated by performing a second numeric fractional derivative of a second order value on the biopotential signal data and by performing an inversed transformation (e.g., inversed FFT) on the fractional derived signal data to convert the fractional derived signal data to a time domain signal data.

5. The method of any one of claims 1-4, wherein each of the first fractional derivative signal data and the second fractional derivative signal data comprises a time domain signal data.

6. The method of any one of claims 1-5, wherein the first fractional derivative signal data is generated by a fractional derivative of an order of pi/2.

7. The method of any one of claims 1-6, wherein the second fractional derivative signal data is generated by a fractional derivative of an order of 0.5.

8. The method of any one of claims 1-7, wherein the geometric features and dynamical properties of the three-dimensional space are generated by performing a MMP (modified matching pursuit) algorithm of the three-dimensional point in said space.

9. The method of any one of claims 1-8, wherein the biopotential signal data is associated with a biopotential signal collected from a single electrical lead.

10. The method of any one of claims 1-9, wherein the single electrical lead collected measurements of the biopotential signal at a location selected from the group consisting of a chest line of the subject, a waistline of the subject, a wrist of the subject, a pelvic line of the

subject, a neck of the subject, an ankle of the subject, a forehead of the subject, and an arm line of the subject.

11. The method of any one of claims 1-10, comprising:

generating, via a processor, an alpha shape of the three-dimensional point in said space, wherein the geometric features and dynamical properties of the three-dimensional space includes the geometric features of the alpha shape.

12. The method of any one of claims 1-11, comprising:

generating a Delaunay triangle mesh of the three-dimensional point in said space, wherein the geometric features and dynamical properties of the three-dimensional space includes the geometric features of the Delaunay triangle mesh.

13. The method of any one of claims 1-12, wherein the biopotential signal data comprises electrocardiogram (ECG) data.

14. The method of any one of claims 1-13, wherein the machine learning comprises an artificial neural network algorithm or a regression random forest algorithm.

15. A method of pre-processing data to extract variables for use in machine learning to diagnose a pathology, the method comprising:

receiving biopotential signal data associated with a subject, said biopotential signal data being associated with biopotential signals collected from two or more electrical leads;

generating, via a processor, a fractional derivative signal data by numerically performing one or more fractional derivative operations of at least one of the biopotential signal data in a frequency domain and converting a result of the one or more fractional derivative operations to a

time domain signal data, wherein the generated fractional derivative signal data comprises a same length and a same sampling frequency as the at least one of the biopotential signal data; and

generating, via the processor, a three-dimensional space wherein each corresponding value of each of the biopotential signal data and the fractional derivative signal data forms a three-dimensional point in said space, wherein geometric features and dynamical properties of the three-dimensional space are used as variables representative of the subject in machine learning to detect one or more diagnosable pathology of the subject.

16. The method of claim 15, wherein each of the two or more electrical leads collected measurements of the biopotential signal at a location selected from the group consisting of a chest line of the subject, a waistline of the subject, a wrist of the subject, a pelvic line of the subject, a neck of the subject, an ankle of the subject, a forehead of the subject, and an arm line of the subject.

17. A method of determining congenital heart defects (CHD) in a mammal, comprising:

receiving biopotential recordings associated with the mammal, the biopotential recordings being recorded at predetermined intervals;

developing variables associated with the biopotential recordings to create a dataset; and

analyzing the dataset to determine if the mammal has a CHD.

18. The method of claim 17, wherein the biopotential recordings associated with the mammal are recorded using a measuring equipment comprising a single surface lead.

19. The method of claim 17 or 18, wherein a measuring equipment comprises an intracardiac electrogram instrument.

20. The method of any one of claims 17-19, wherein a measuring equipment comprises a smart watch or fitness heart band.

21. A system comprising:

a storage device configured to receive biopotential data from a network-connected biopotential measuring apparatus;

one or more processors;

a memory having instructions stored thereon, wherein the instructions, when executed by the processor, cause the processor to:

generate phase space variables associated with the biopotential data;

analyzing the phase space variables to determine if the mammal has a CHD.

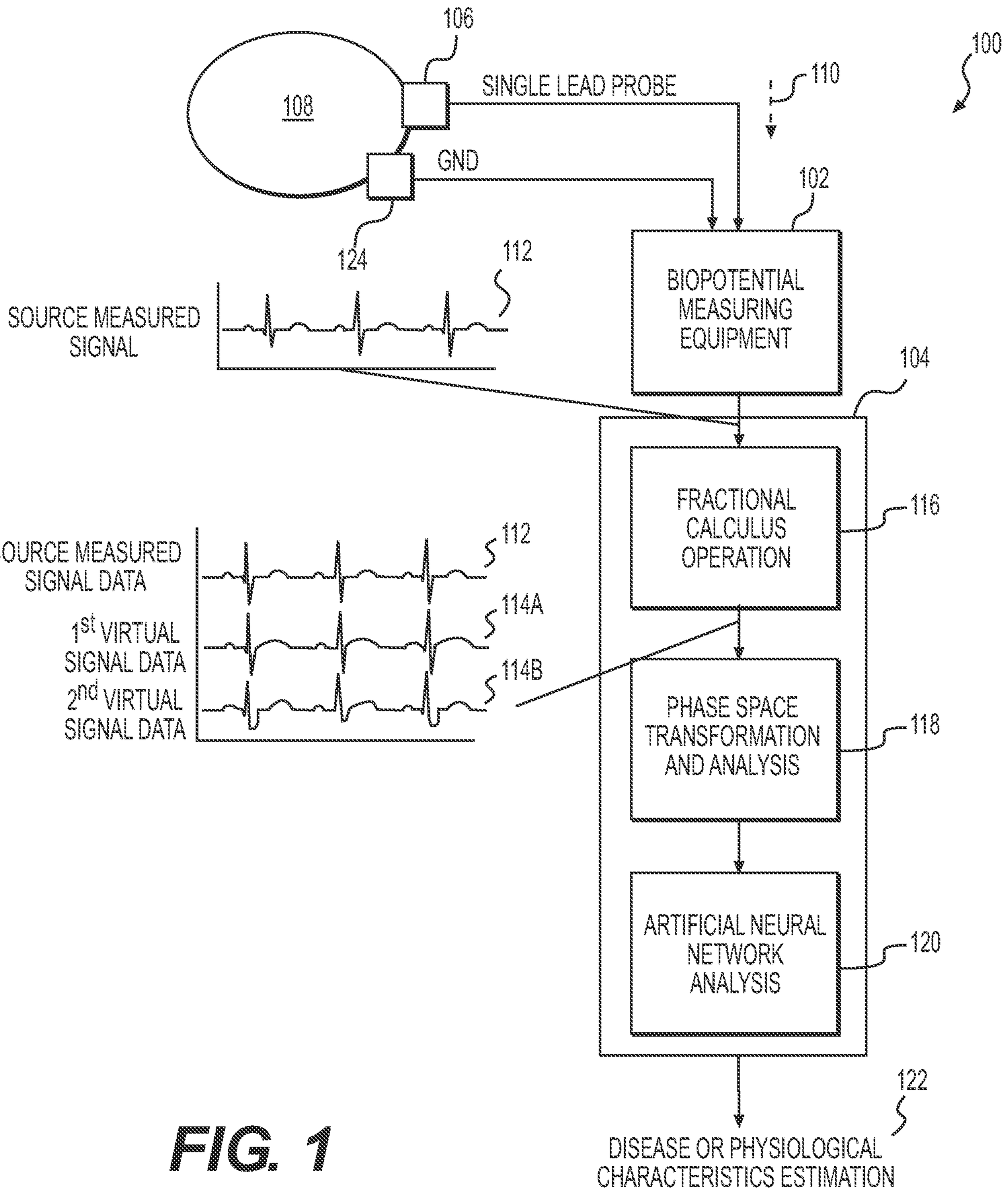
106
100
SINGLE LEAD PROBE
110
108
GND
102
124
112
SOURCE MEASURED SIGNAL
BIOPOTENTIAL MEASURING EQUIPMENT
104
FRACTIONAL CALCULUS OPERATION
116
SOURCE MEASURED SIGNAL DATA
112
1st VIRTUAL SIGNAL DATA
114A
2nd VIRTUAL SIGNAL DATA
114B
PHASE SPACE TRANSFORMATION AND ANALYSIS
118
ARTIFICIAL NEURAL NETWORK ANALYSIS
120
122
DISEASE OR PHYSIOLOGICAL CHARACTERISTICS ESTIMATION

FIG. 1

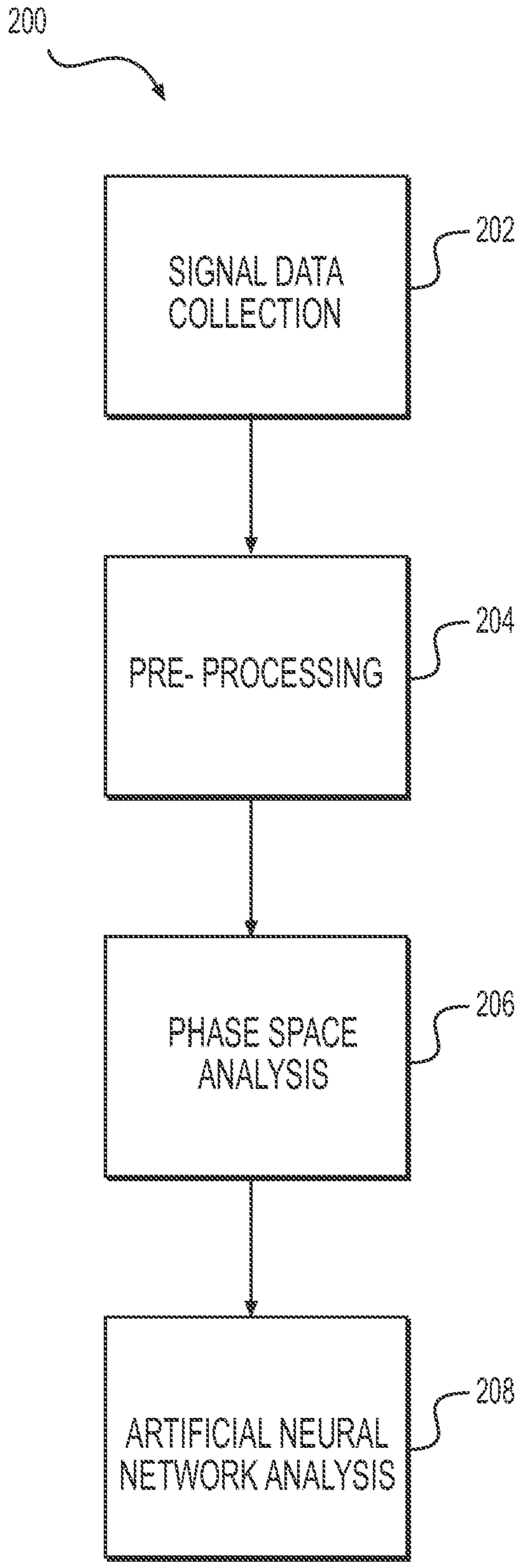
200
SIGNAL DATA COLLECTION
202
PRE- PROCESSING
204
PHASE SPACE ANALYSIS
206
ARTIFICIAL NEURAL NETWORK ANALYSIS
208

FIG. 2

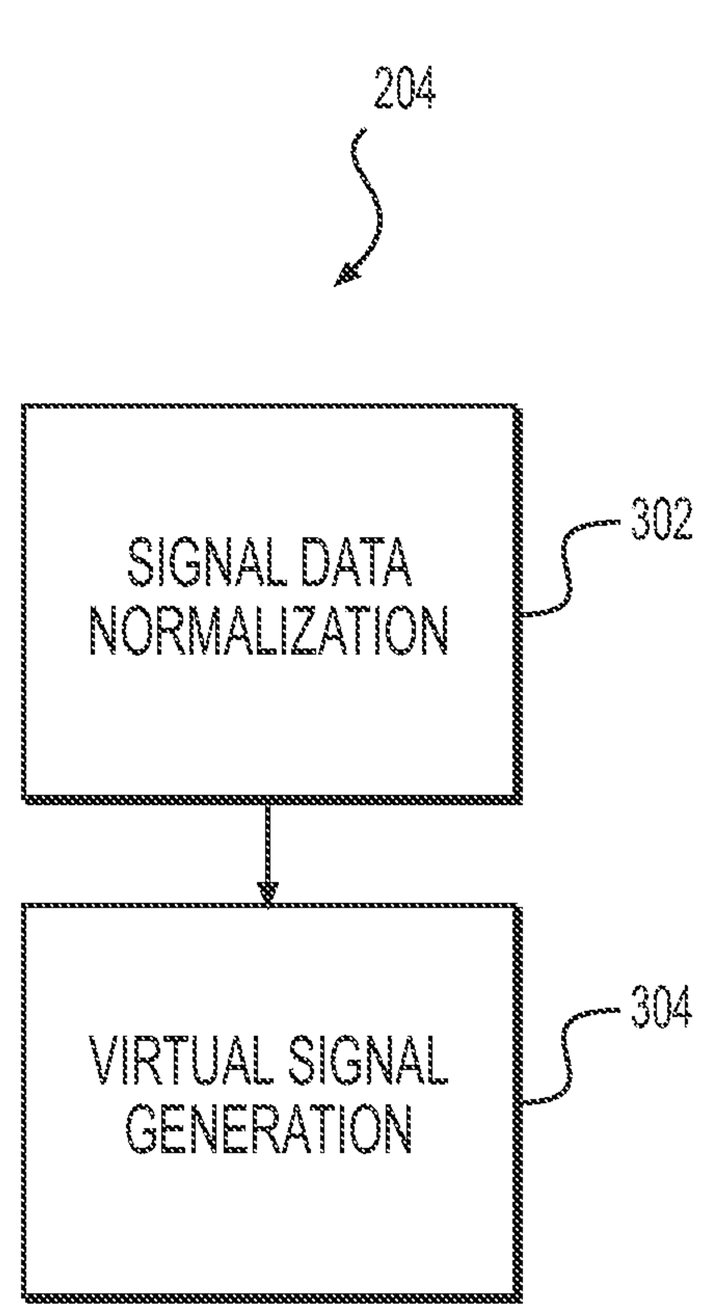
204
SIGNAL DATA NORMALIZATION
302
VIRTUAL SIGNAL GENERATION
304

FIG. 3

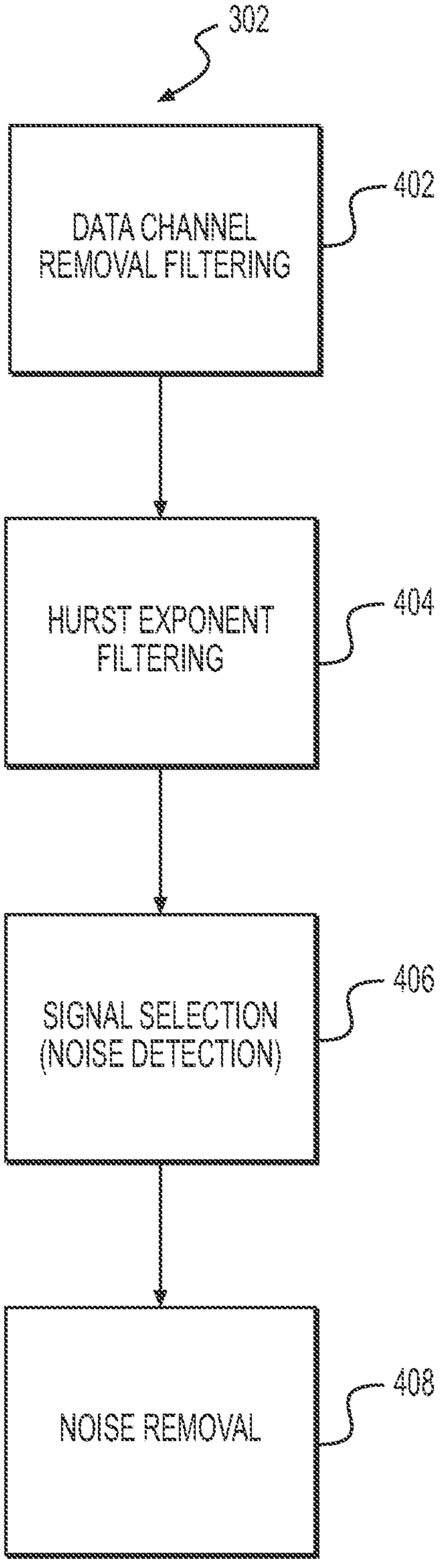
302
DATA CHANNEL REMOVAL FILTERING
402
HURST EXPONENT FILTERING
404
SIGNAL SELECTION (NOISE DETECTION)
406
NOISE REMOVAL
408

FIG. 4

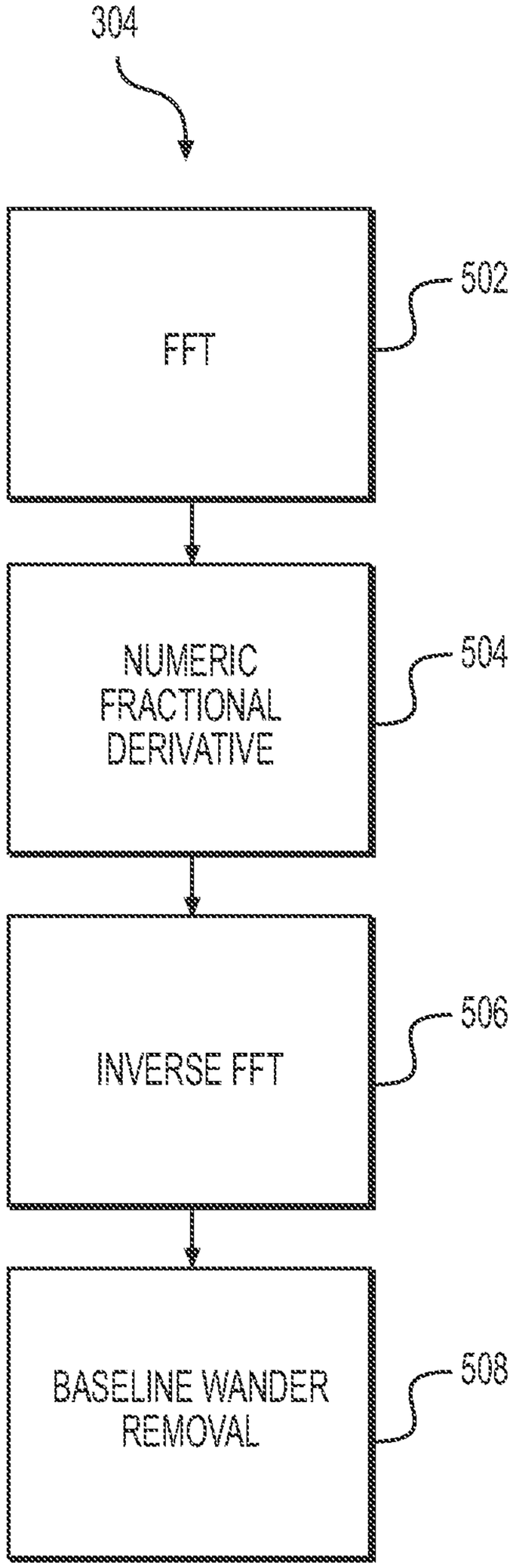
304
FFT
502
NUMERIC FRACTIONAL DERIVATIVE
504
INVERSE FFT
506
BASELINE WANDER REMOVAL
508

FIG. 5A

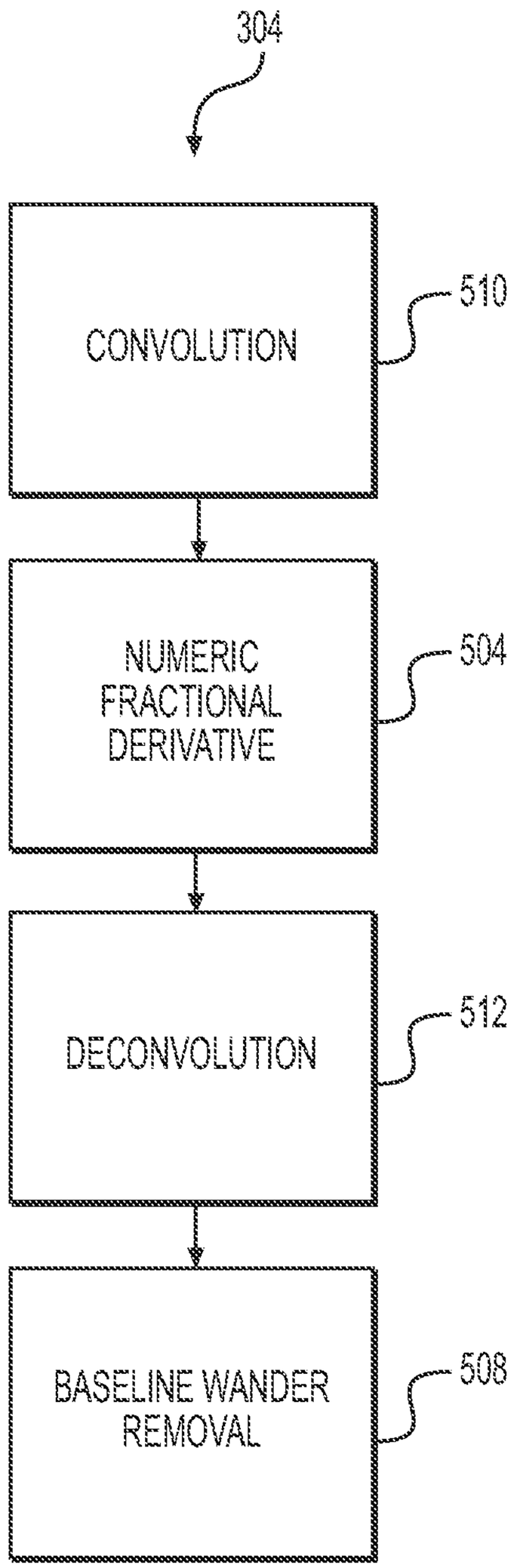
304
CONVOLUTION
510
NUMERIC FRACTIONAL DERIVATIVE
504
DECONVOLUTION
512
BASELINE WANDER REMOVAL
508

FIG. 5B

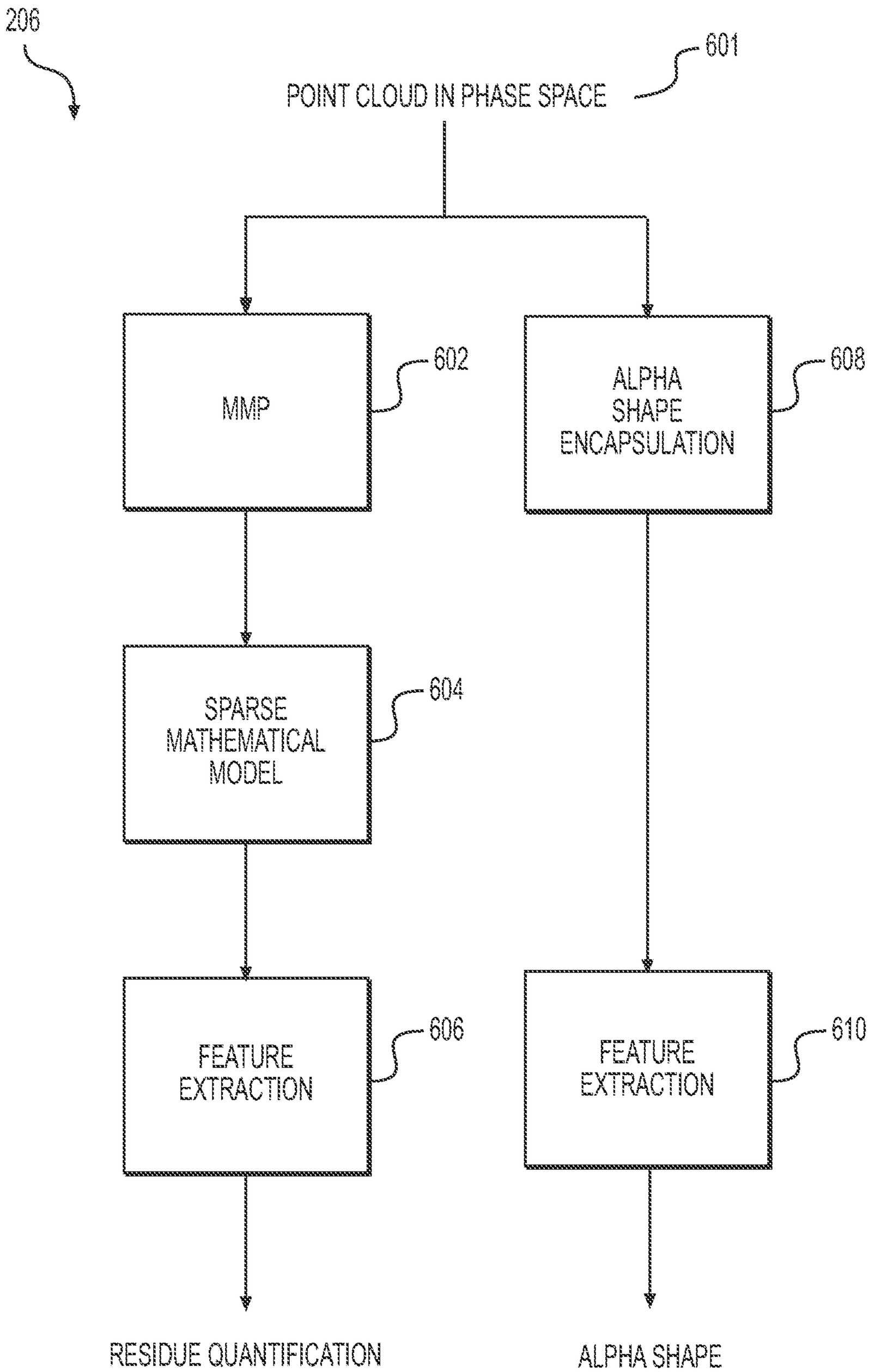
206
POINT CLOUD IN PHASE SPACE
601
MMP
602
ALPHA SHAPE ENCAPSULATION
608
SPARSE MATHEMATICAL MODEL
604
FEATURE EXTRACTION
606
FEATURE EXTRACTION
610
RESIDUE QUANTIFICATION
ALPHA SHAPE

FIG. 6

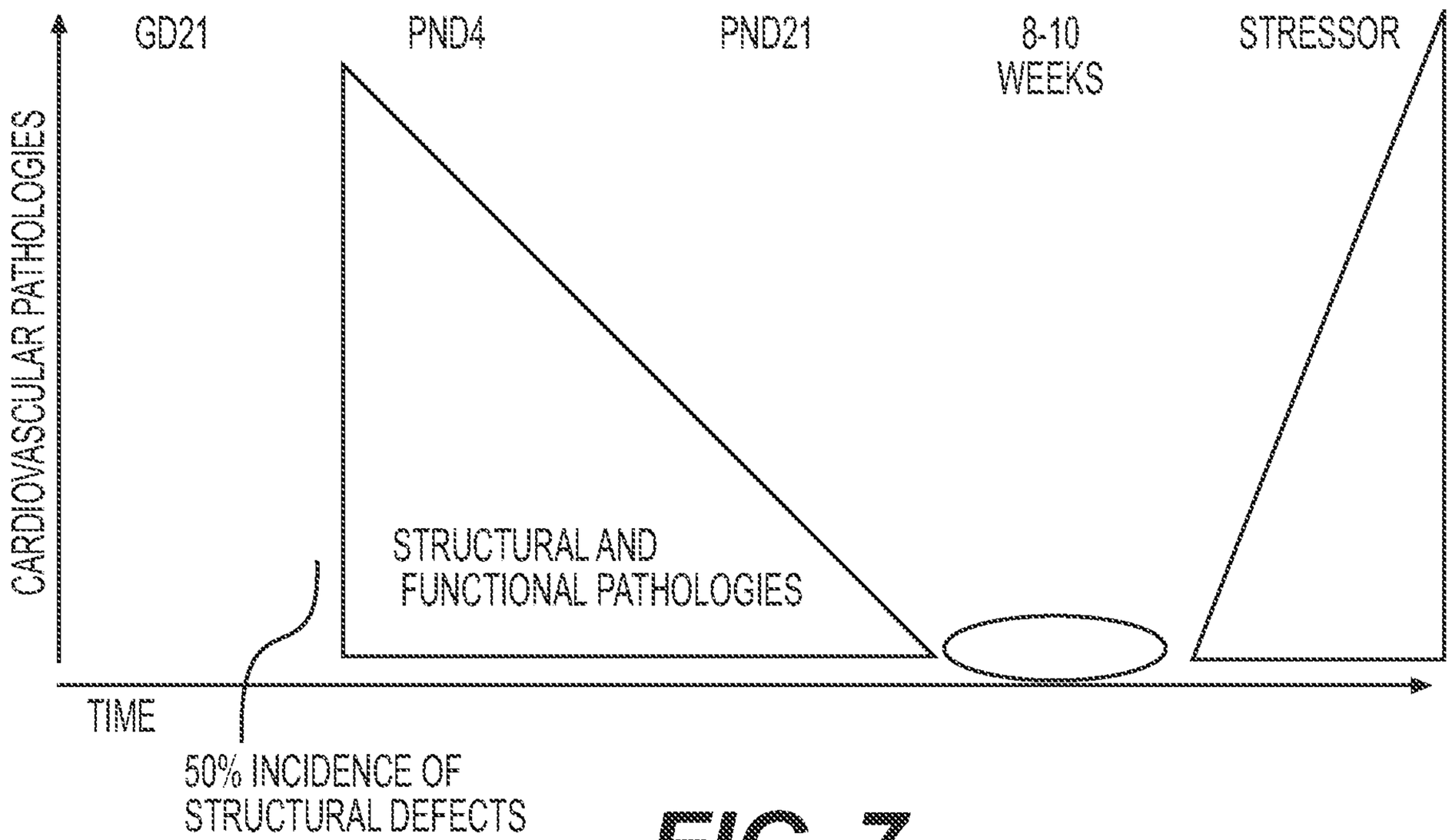
GD21
PND4
PND21
8-10 WEEKS
STRESSOR
CARDIOVASCULAR PATHOLOGIES
STRUCTURAL AND FUNCTIONAL PATHOLOGIES
TIME
50% INCIDENCE OF STRUCTURAL DEFECTS

FIG. 7

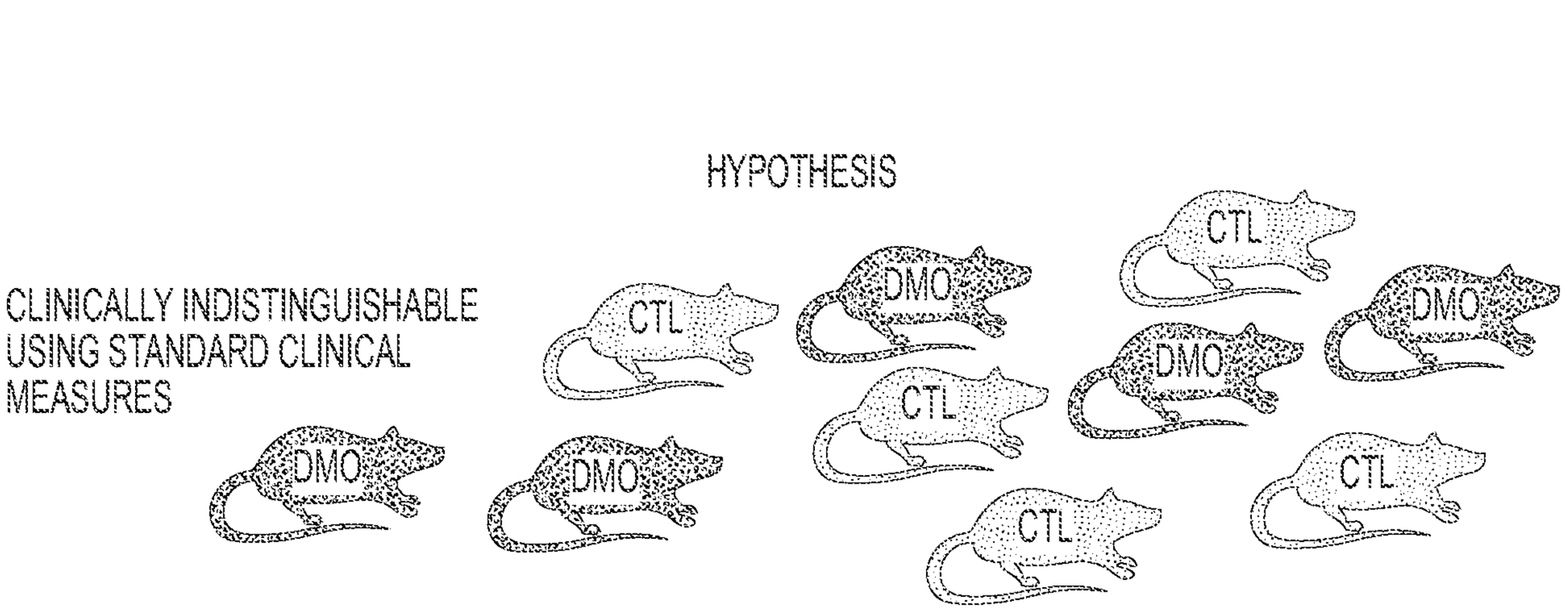
HYPOTHESIS
CLINICALLY INDISTINGUISHABLE USING STANDARD CLINICAL MEASURES
CTL
DMO
MACHINE LEARNING ANALYSIS

FIG. 8

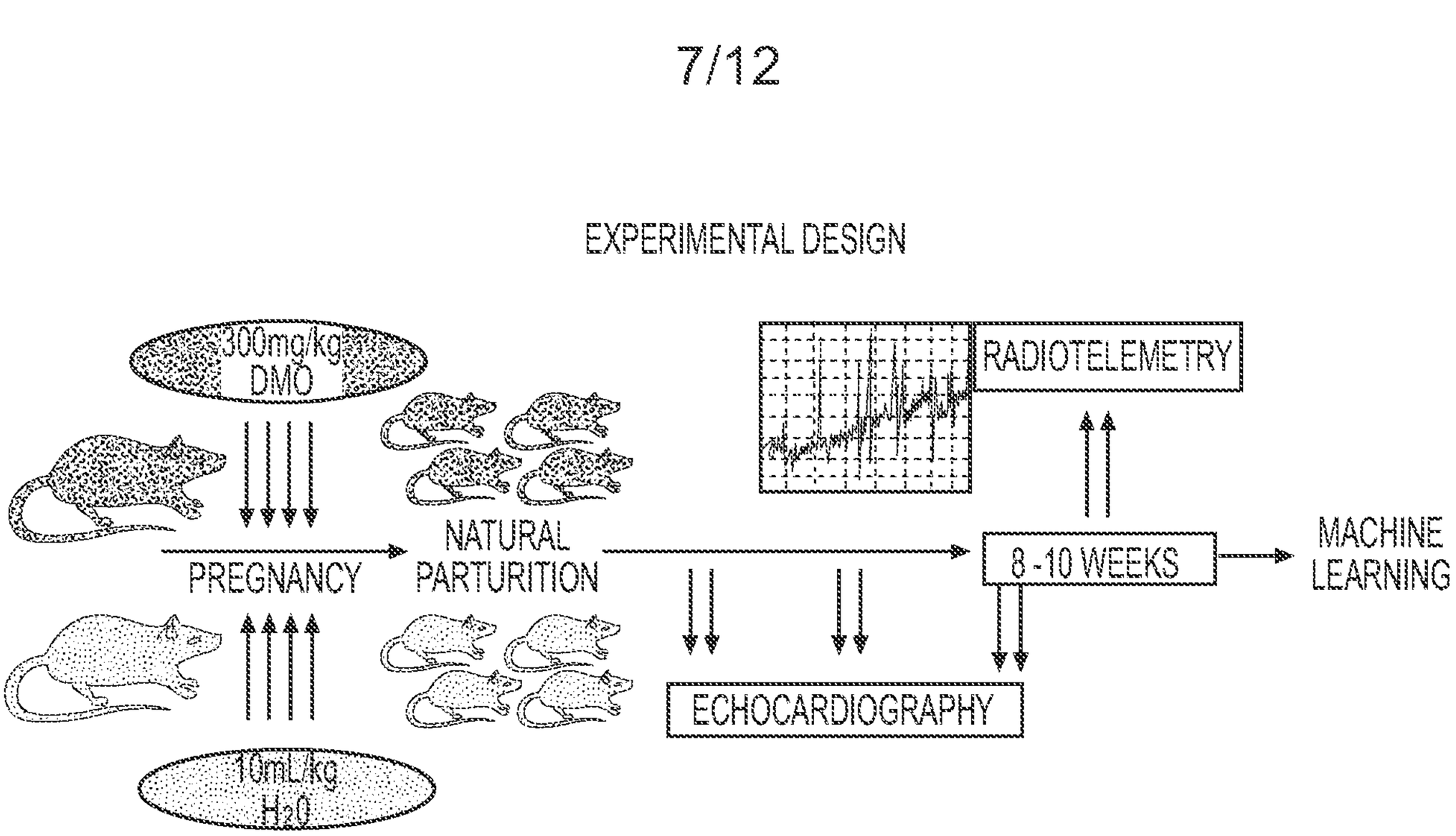
EXPERIMENTAL DESIGN
300mg/kg
DMO
RADIOTELEMETRY
NATURAL
PARTURITION
PREGNANCY
8 -10 WEEKS
MACHINE
LEARNING
ECHOCARDIOGRAPHY
10mL/kg
H20

FIG. 9

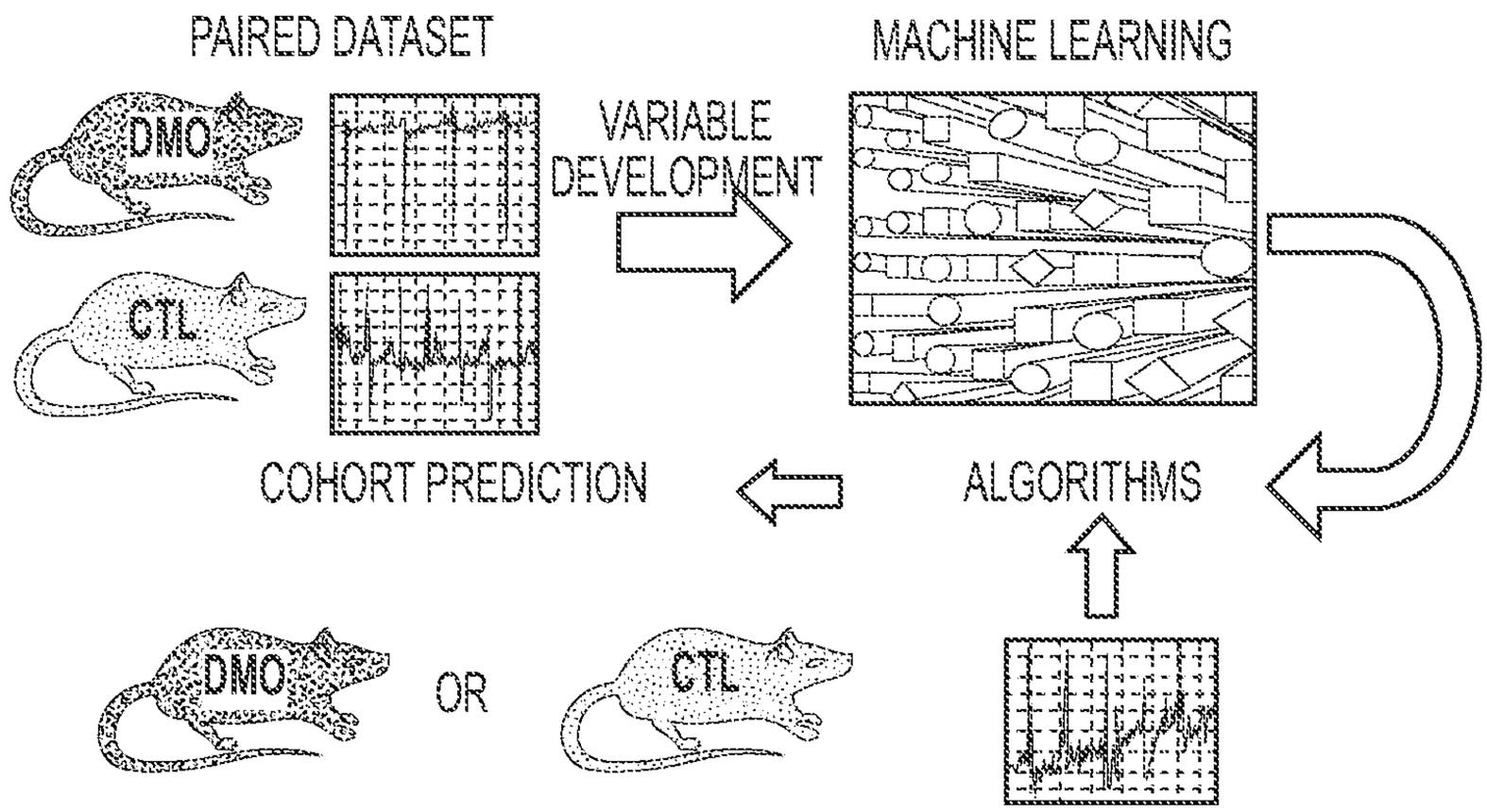
METHODOLOGY OF MACHINE LEARNING
PAIRED DATASET
MACHINE LEARNING
DMO
CTL
VARIABLE
DEVELOPMENT
COHORT PREDICTION
ALGORITHMS
DMO
OR
CTL

FIG. 10

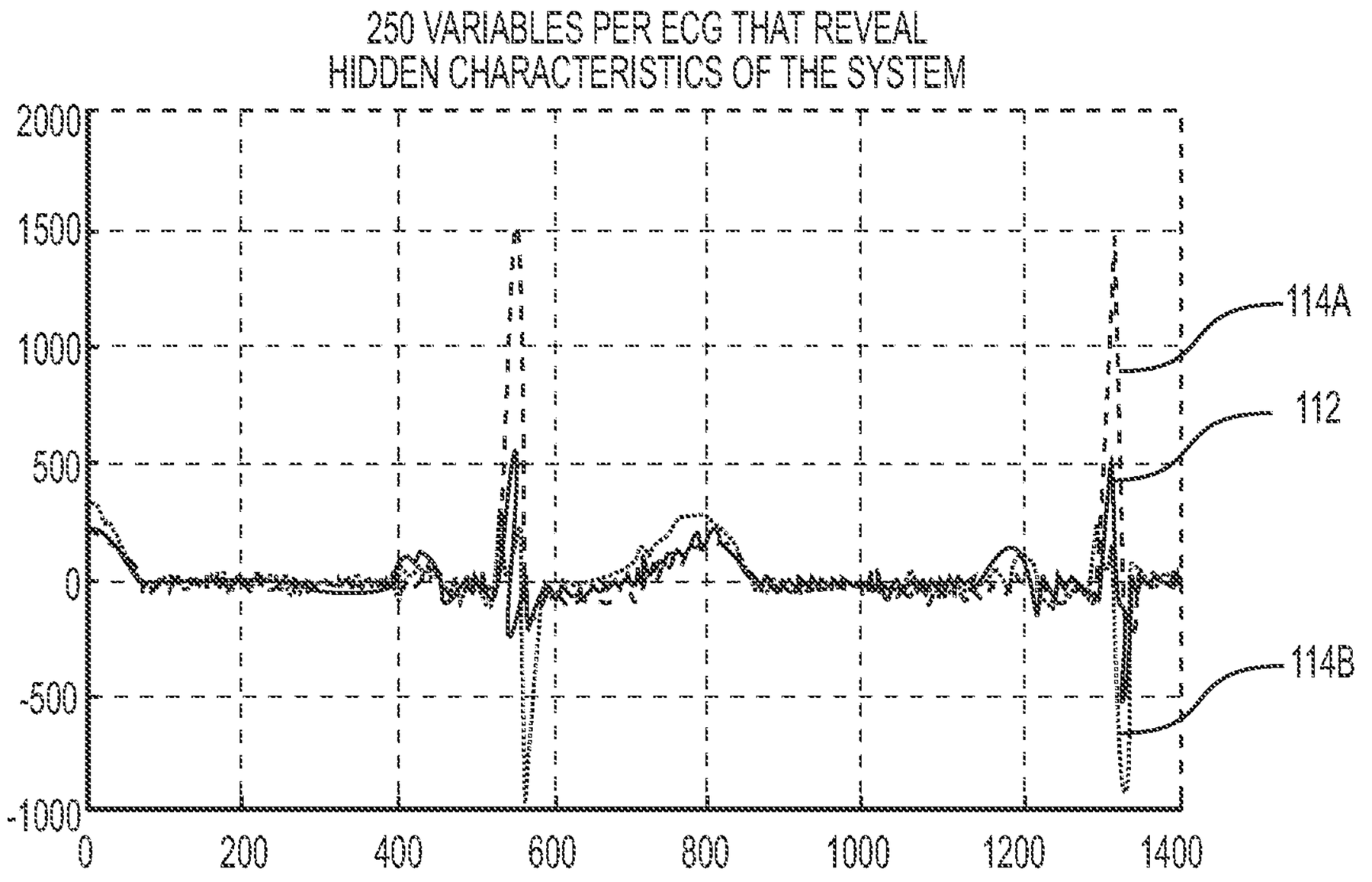
250 VARIABLES PER ECG THAT REVEAL
HIDDEN CHARACTERISTICS OF THE SYSTEM
2000
1500
1000
500
0
-500
-1000
0
200
400
600
800
1000
1200
1400
114A
112
114B

FIG. 11

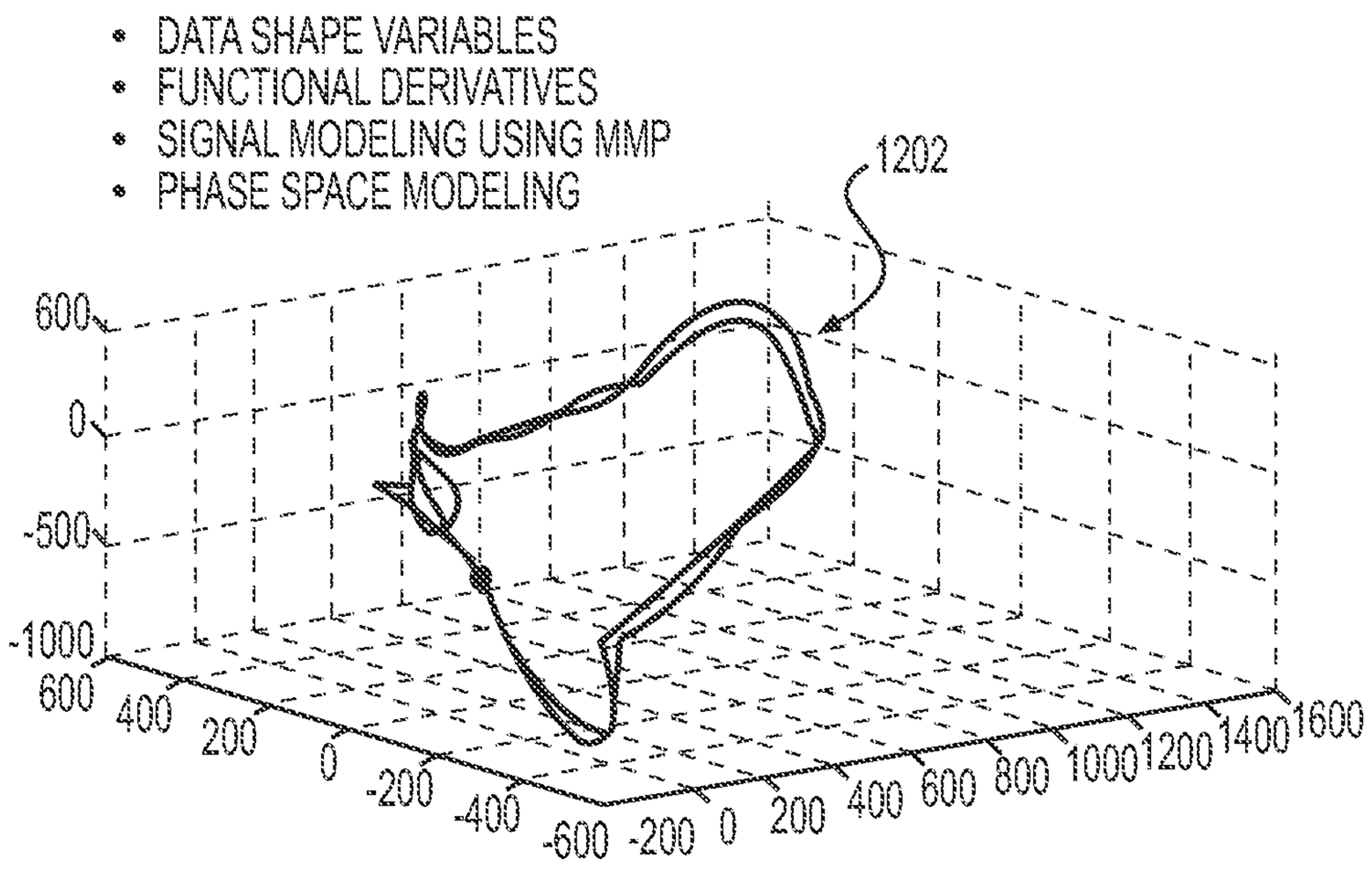
• DATA SHAPE VARIABLES
• FUNCTIONAL DERIVATIVES
• SIGNAL MODELING USING MMP
• PHASE SPACE MODELING
1202
600
0
-500
-1000
600
400
200
0
-200
-400
-600
-200
0
200
400
600
800
1000
1200
1400
1600

FIG. 12

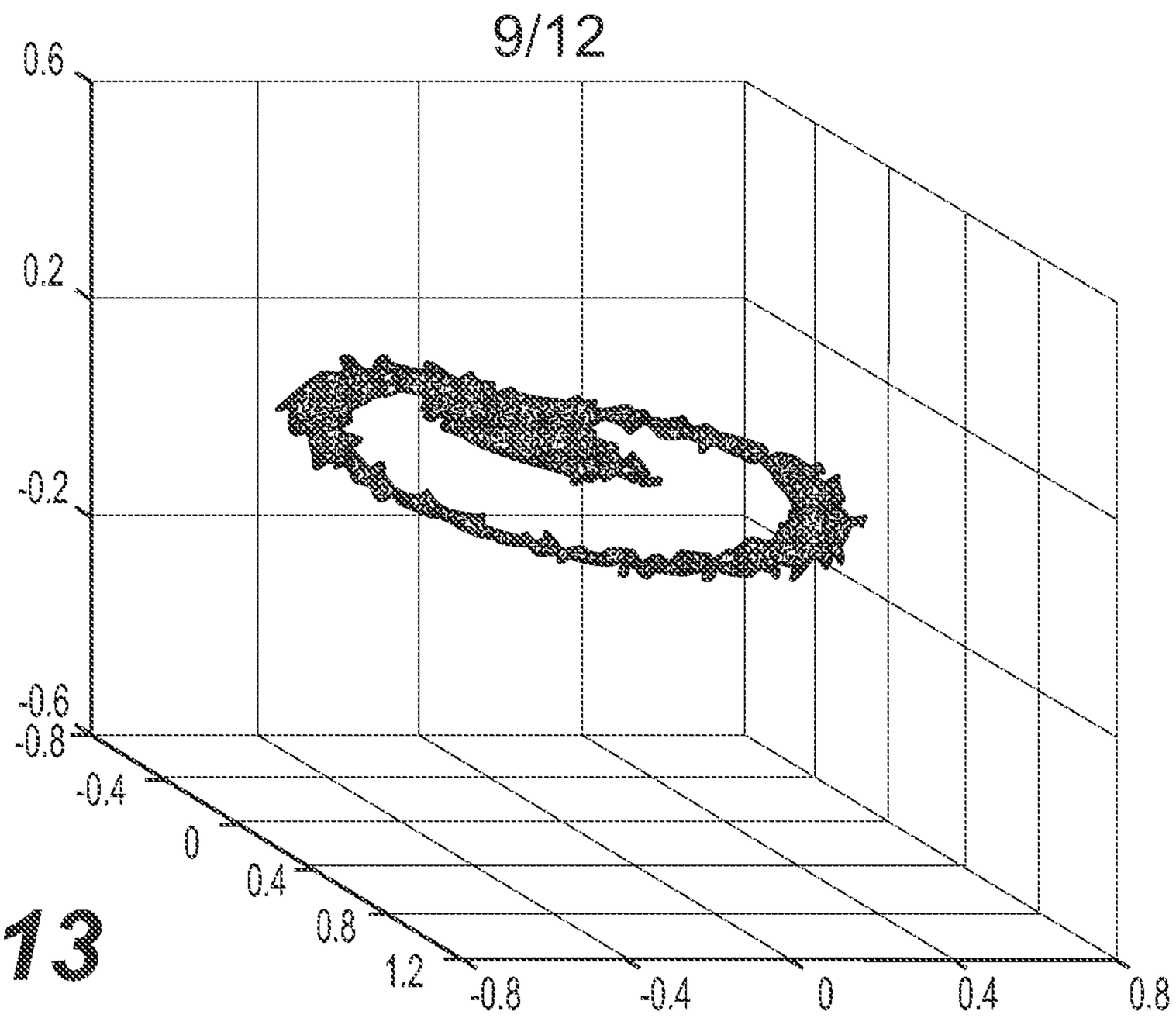
0.6
0.2
-0.2
-0.6
-0.8
-0.4
0
0.4
0.8
1.2
-0.8
-0.4
0
0.4
0.8

FIG. 13

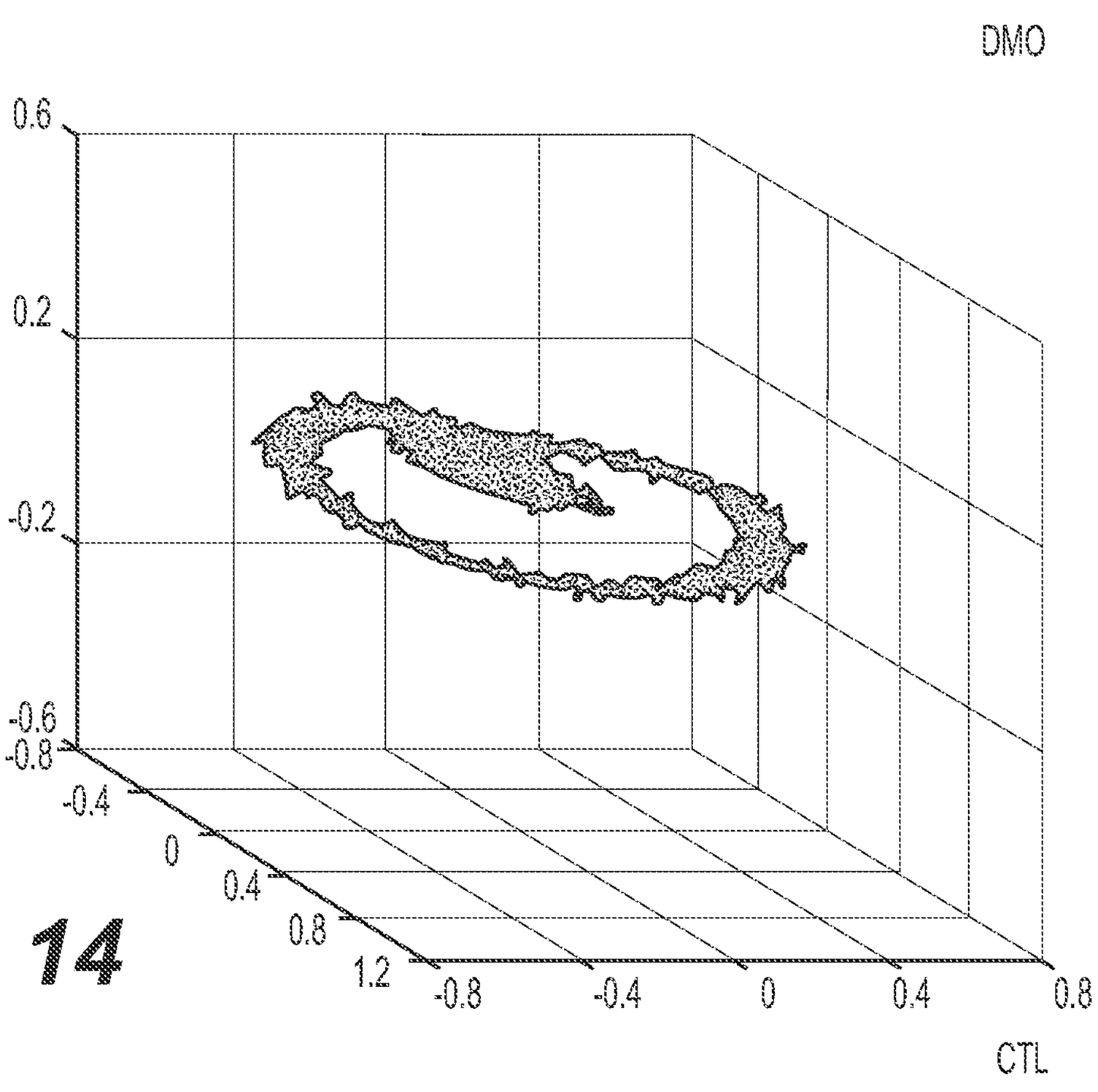
DMO
0.6
0.2
-0.2
-0.6
-0.8
-0.4
0
0.4
0.8
1.2
-0.8
-0.4
0
0.4
0.8
CTL

FIG. 14

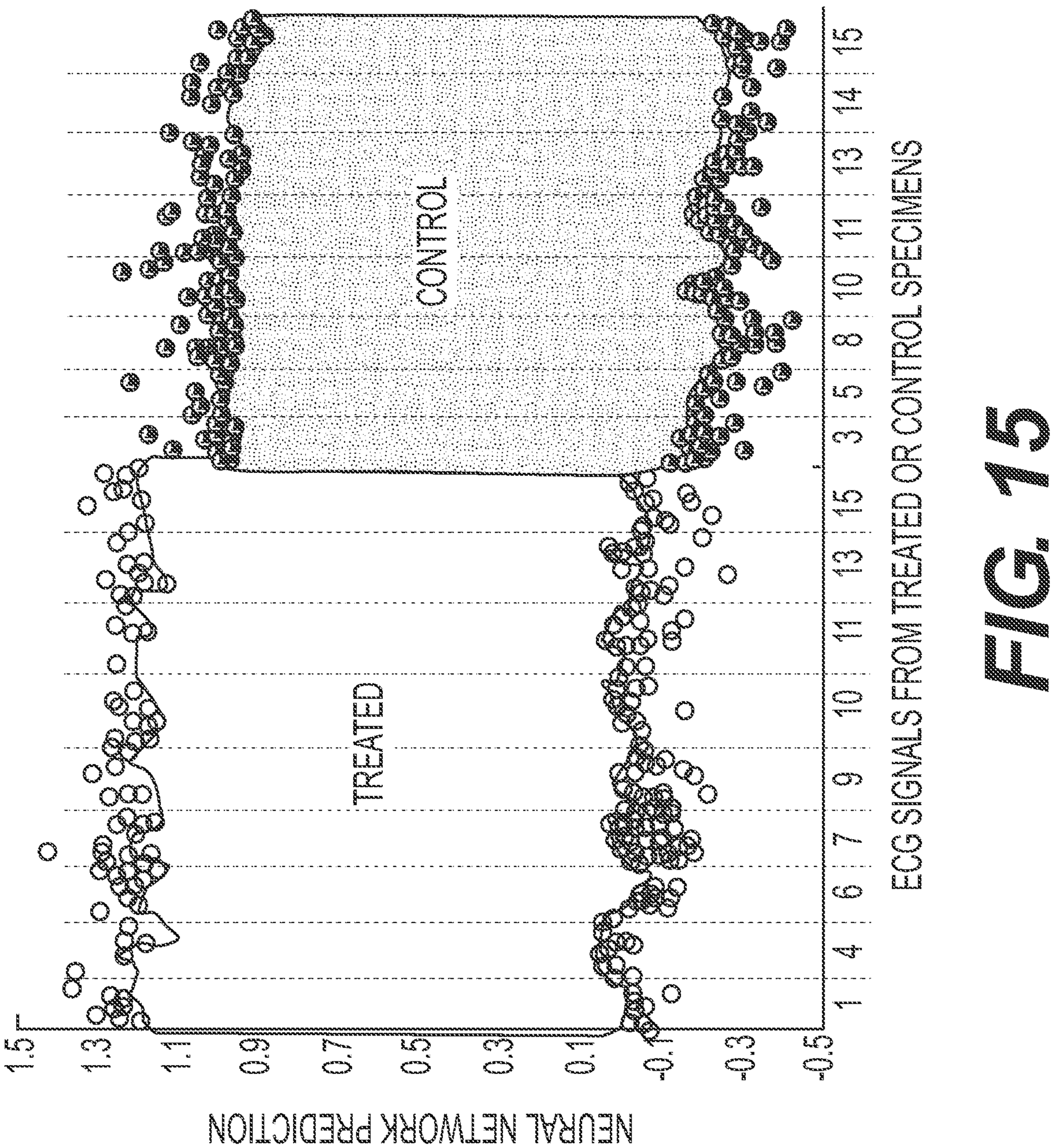
CONTROL
TREATED
1.5
1.3
1.1
0.9
0.7
0.5
0.3
0.1
-0.1
-0.3
-0.5
1 3 5 8 10 11 13 14 15
1 4 6 7 9 10 11 13 15
ECG SIGNALS FROM TREATED OR CONTROL SPECIMENS
NEURAL NETWORK PREDICTION

FIG. 15

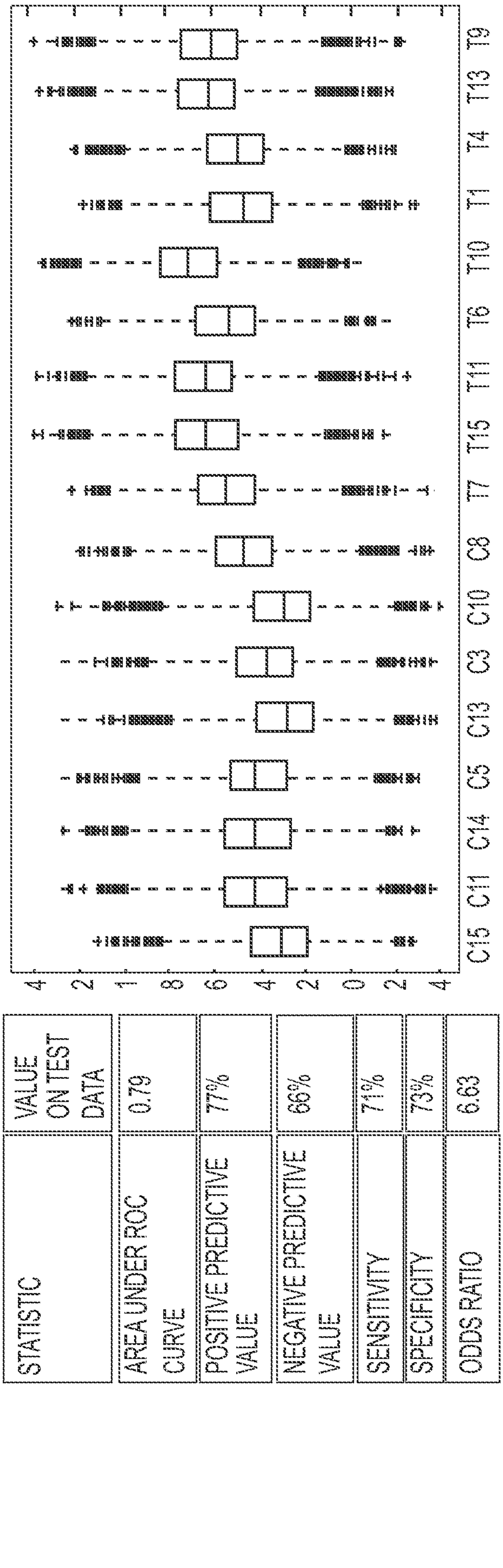

FIG. 16

STATISTIC	VALUE ON TEST DATA
AREA UNDER ROC CURVE	0.79
POSITIVE PREDICTIVE VALUE	77%
NEGATIVE PREDICTIVE VALUE	66%
SENSITIVITY	71%
SPECIFICITY	73%
ODDS RATIO	6.63

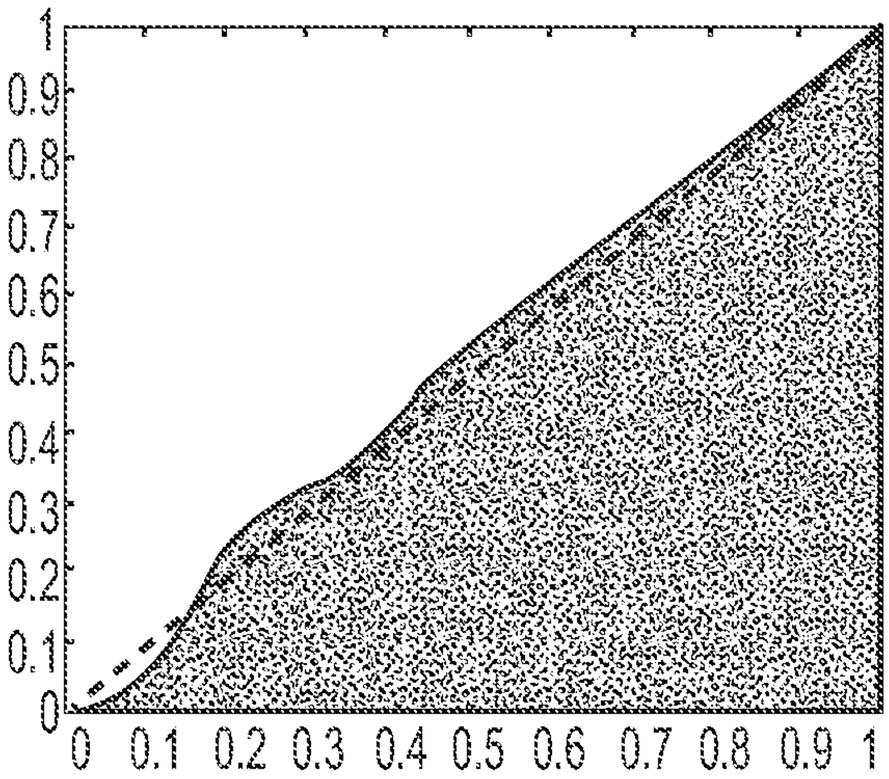

FIG. 17A

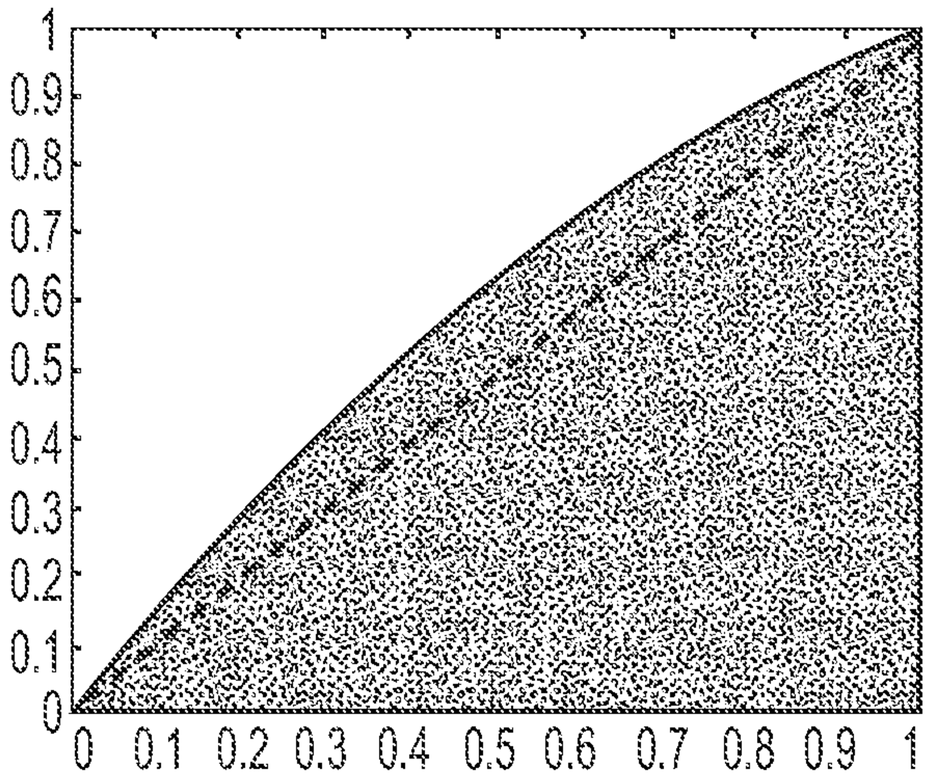

FIG. 17B

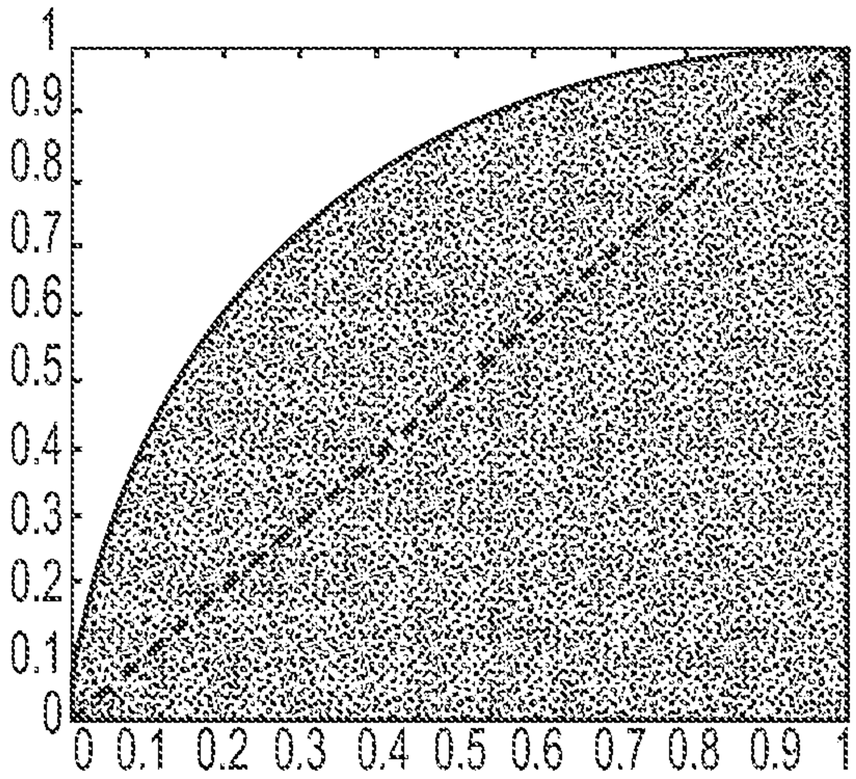

FIG. 18

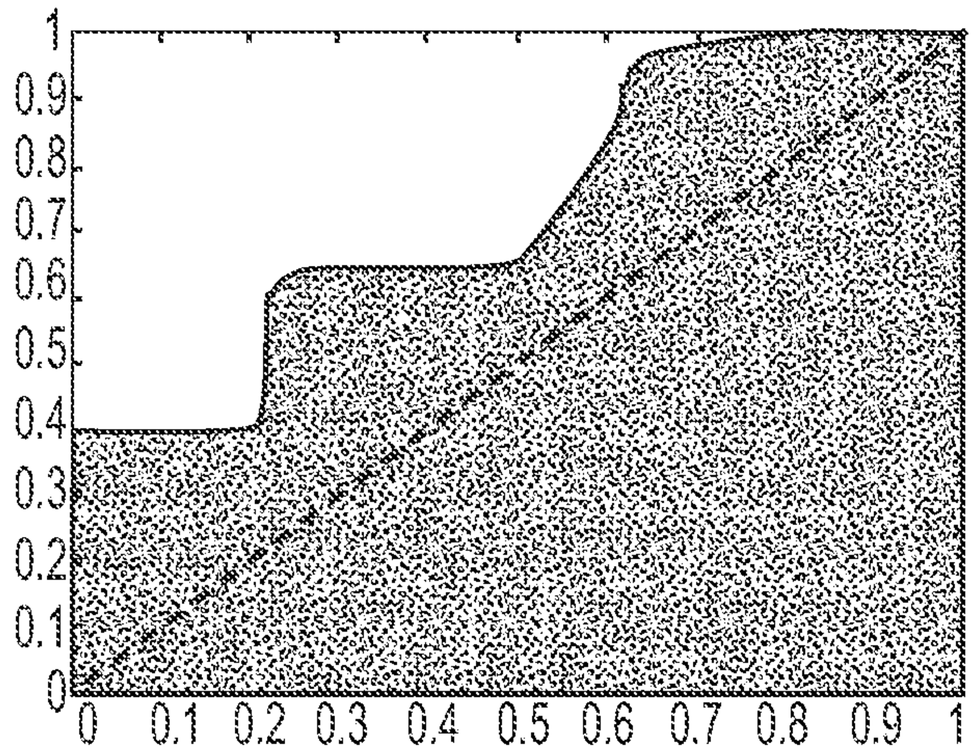

FIG. 19

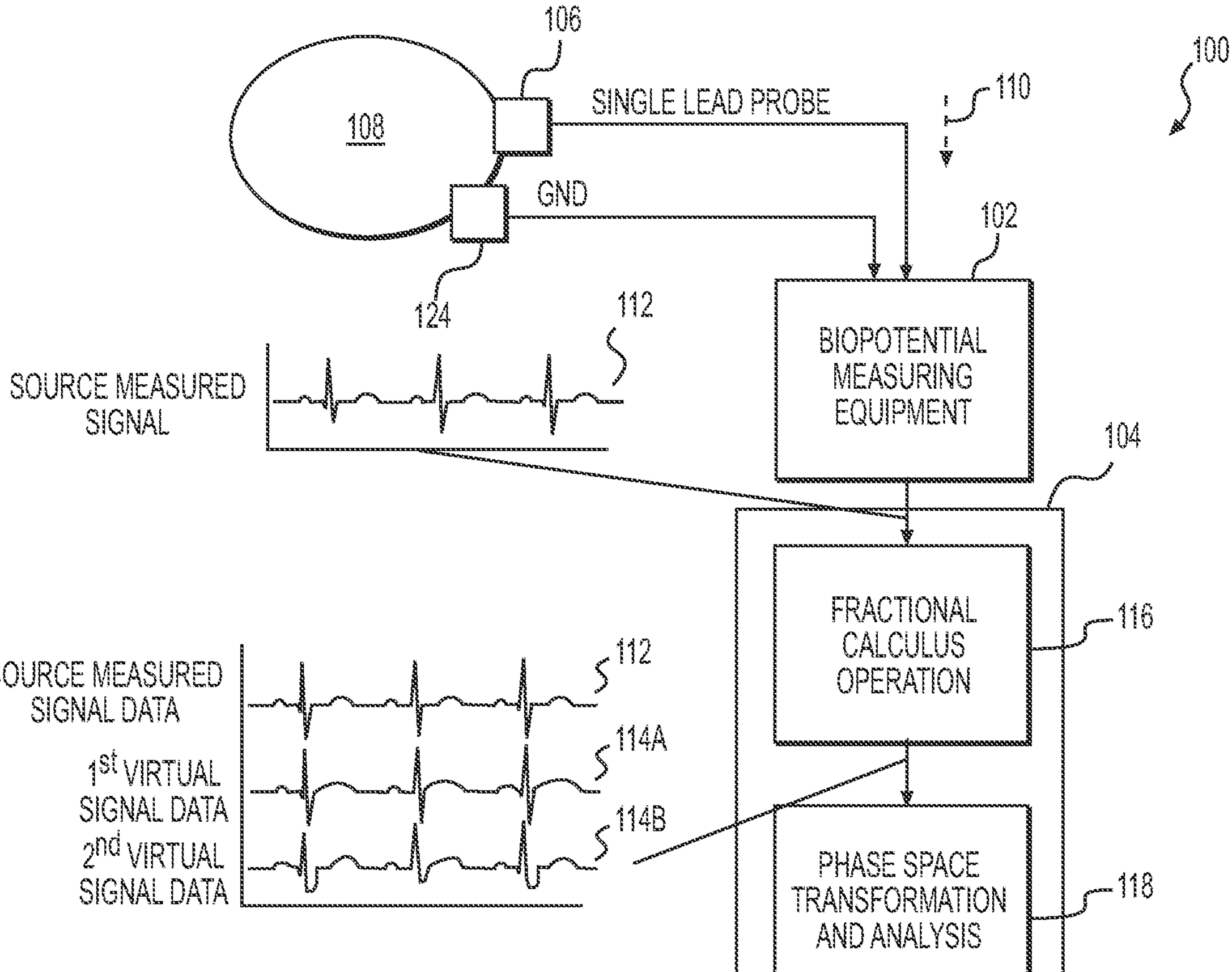
106
108
SINGLE LEAD PROBE
110
100
GND
102
124
112
SOURCE MEASURED SIGNAL
BIOPOTENTIAL MEASURING EQUIPMENT
104
SOURCE MEASURED SIGNAL DATA
112
1st VIRTUAL SIGNAL DATA
114A
2nd VIRTUAL SIGNAL DATA
114B
FRACTIONAL CALCULUS OPERATION
116
PHASE SPACE TRANSFORMATION AND ANALYSIS
118
ARTIFICIAL NEURAL NETWORK ANALYSIS
120
122
DISEASE OR PHYSIOLOGICAL CHARACTERISTICS ESTIMATION

FIG. 1