PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 19/00 | A2

(11) International Publication Number: **WO 97/29447**

(43) International Publication Date: 14 August 1997 (14.08.97)

(21) International Application Number: PCT/US97/02104

(22) International Filing Date: 7 February 1997 (07.02.97)

(30) Priority Data:

08/599,275	9 February 1996 (09.02.96)	US
60/011,449	9 February 1996 (09.02.96)	US

(71) Applicant *(for all designated States except US)*: ADEZA BIOMEDICAL CORPORATION [US/US]; 1240 Elko Drive, Sunnyvale, CA 94089 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LAPOINTE, Jerome [US/US]; 3051 Revere Avenue, Oakland, CA 94605 (US). DESIENO, Duane, D. [US/US]; 2015 Olite Court, La Jolla, CA 92037 (US).

(74) Agent: SEIDMAN, Stephanie, L.; Brown Martin Haller & McClain, 1660 Union Street, San Diego, CA 92101-2926 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD FOR SELECTING MEDICAL AND BIOCHEMICAL DIAGNOSTIC TESTS USING NEURAL NETWORK-RELATED APPLICATIONS

(57) Abstract

Methods are provided for developing medical diagnostic tests using decision-support systems, such as neural networks. Patient data or information, typically patient history or clinical data, are analyzed by the decision-support systems to identify important or relevant variables and decision-support systems are trained on the patient data. Patient data are augmented by biochemical test data, or results, where available, to refine performance. The resulting decision-support systems are employed to evaluate specific observation values and test results, to guide the development of biochemical or other diagnostic tests, to assess a course of treatment, to identify new diagnostic tests and disease markers, to identify useful therapies, and to provide the decision-support functionality for the test. Methods for identification of important input variables for medical diagnostic tests for use in training the decision-support systems to guide the development of the tests, for improving the sensitivity and specificity of such tests, and for selecting diagnostic tests that improve overall diagnosis of, or potential for, a disease state and that permit the effectiveness of a selected therapeutic protocol to be assessed are provided. The methods for identification can be applied in any field in which statistics are used to determine outcomes. A method for evaluating the effectiveness of any given diagnostic test is also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

METHOD FOR SELECTING MEDICAL AND BIOCHEMICAL DIAGNOSTIC TESTS USING NEURAL NETWORK-RELATED APPLICATIONS

This application is a continuation-in-part of U.S. application Serial 08/599,275, entitled "METHOD FOR DEVELOPING MEDICAL AND BIOCHEMICAL DIAGNOSTIC TESTS USING NEURAL NETWORKS" to Jerome Lapointe and Duane DeSieno, filed February 9, 1996, and claims priority under 35 U.S.C. §119(e) to U.S. provisional application Serial No. 60/011,449, entitled "METHOD AND APPARATUS FOR AIDING IN THE DIAGNOSIS OF ENDOMETRIOSIS USING A PLURALITY OF PARAMETERS SUITED FOR ANALYSIS THROUGH A NEURAL NETWORK" to Jerome Lapointe and Duane DeSieno, filed February 9, 1996.

The subject matter of each of the above-noted application and provisional application is herein incorporated in its entirety by reference thereto.

MICROFICHE APPENDIX

Two computer Appendices containing computer program source code for programs described herein have been submitted concurrently with the filing of this application. The Computer Appendices will be converted to a Microfiche Appendices pursuant to 37 C.F.R. 1.96(b). The Computer Appendices, which are referred to hereafter as the "Microfiche Appendices, are each incorporated herein by reference in its entirety.

Thus, a portion of the disclosure of this patent document contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

FIELD OF THE INVENTION

This subject matter of the invention relates to the use of prediction technology, particularly nonlinear prediction technology, for the development of medical diagnostic aids. In particular, training techniques operative on neural networks and other expert systems with inputs from patient historical information for the development of medical diagnostic tools and methods of diagnosis are provided.

BACKGROUND OF THE INVENTION

Data Mining, decision support-systems and neural networks

A number of computer decision-support systems have the ability to classify information and identify patterns in input data, and are particularly useful in evaluating data sets having large quantities of variables and complex interactions between variables. These computer decision systems which are collectively identified as "data mining" or "knowledge discovery in databases" (and herein as decision-support systems) rely on similar basic hardware components, e.g., personal computers (PCS) with a processor, internal and peripheral devices, memory devices and input/output interfaces. The distinctions between the systems arise within the software, and more fundamentally, the paradigms upon which the software is based. Paradigms that provide decision-support functions include regression methods, decision trees, discriminant analysis, pattern recognition, Bayesian decision theory, and fuzzy logic. One of the more widely used decision-support computer systems is the artificial neural network.

Artificial neural networks or "neural nets" are parallel information processing tools in which individual processing elements called neurons are arrayed in layers and furnished with a large number of interconnections between elements in successive layers. The functioning of the processing elements are modeled to approximate biologic neurons where

the output of the processing element is determined by a typically non-linear transfer function. In a typical model for neural networks, the processing elements are arranged into an input layer for elements which receive inputs, an output layer containing one or more elements which generate an output, and one or more hidden layers of elements therebetween. The hidden layers provide the means by which non-linear problems may be solved. Within a processing element, the input signals to the element are weighted arithmetically according to a weight coefficient associated with each input. The resulting weighted sum is transformed by a selected non-linear transfer function, such as a sigmoid function, to produce an output, whose values range from 0 to 1, for each processing element. The learning process, called "training", is a trial-and-error process involving a series of iterative adjustments to the processing element weights so that a particular processing element provides an output which, when combined with the outputs of other processing elements, generates a result which minimizes the resulting error between the outputs of the neural network and the desired outputs as represented in the training data. Adjustment of the element weights are triggered by error signals. Training data are described as a number of training examples in which each example contains a set of input values to be presented to the neural network and an associated set of desired output values.

A common training method is backpropagation or "backprop", in which error signals are propagated backwards through the network. The error signal is used to determine how much any given element's weight is to be changed and the error gradient, with the goal being to converge to a global minimum of the mean squared error. The path toward convergence, i.e., the gradient descent, is taken in steps, each step being an adjustment of the input weights of the processing element. The size

of each step is determined by the learning rate. The slope of the gradient descent includes flat and steep regions with valleys that act as local minima, giving the false impression that convergence has been achieved, leading to an inaccurate result.

Some variants of backprop incorporate a momentum term in which a proportion of the previous weight-change value is added to the current value. This adds momentum to the algorithm's trajectory in its gradient descent, which may prevent it from becoming "trapped" in local minima. One backpropogation method which includes a momentum term is "Quickprop", in which the momentum rates are adaptive. The Quickprop variation is described by Fahlman (see,"Fast Learning Variations on Back-Propagation: An Empirical Study", Proceedings on the 1988 Connectionist Models Summer School, Pittsburgh, 1988, D. Touretzky, et al., eds., pp.38-51, Morgan Kaufmann, San Mateo, CA; and, with Lebriere, "The Cascade-Correlation Learning Architecture", Advances in Neural Information Processing Systems 2,(Denver, 1989), D. Touretzky, ed., pp. 524-32. Morgan Kaufmann, San Mateo, CA). The Quickprop algorithm is publicly accessible, and may be downloaded via the Internet, from the Artificial Intelligence Repository maintained by the School of Computer Science at Carnegie Mellon University. In Quickprop, a dynamic momentum rate is calculated based upon the slope of the gradient. If the slope is smaller but has the same sign as the slope following the immediately preceding weight adjustment, the weight change will accelerate. The acceleration rate is determined by the magnitude of successive differences between slope values. If the current slope is in the opposite direction from the previous slope, the weight change decelerates. The Quickprop method improves convergence speed, giving the steepest possible gradient descent, helping to prevent convergence to a local minimum.

When neural networks are trained on sufficient training data, the neural network acts as an associative memory that is able to generalize to a correct solution for sets of new input data that were not part of the training data. Neural networks have been shown to be able to operate even in the absence of complete data or in the presence of noise. It has also been observed that the performance of the network on new or test data tends to be lower than the performance on training data. The difference in the performance on test data indicates the extent to which the network was able to generalize from the training data. A neural network, however, can be retrained and thus learn from the new data, improving the overall performance of the network.

Neural nets, thus, have characteristics that make them well suited for a large number of different problems, including areas involving prediction, such as medical diagnosis.

Neural Nets and Diagnosis

In diagnosing and/or treating a patient, a physician will use patient condition, symptoms, and the results of applicable medical diagnostic tests to identify the disease state or condition of the patient. The physician must carefully determine the relevance of the symptoms and test results to the particular diagnosis and use judgement based on experience and intuition in making a particular diagnosis. Medical diagnosis involves integration of information from several sources including a medical history, a physical exam and biochemical tests. Based upon the results of the exam and tests and answers to the questions, the physician, using his or her training, experience and knowledge and expertise, formulates a diagnosis. A final diagnosis may require subsequent surgical procedures to verify or to formulate. Thus, the process of diagnosis involves a combination of decision-support,

intuition and experience. The validity of a physician's diagnosis is very dependent upon his/her experience and ability.

Because of the predictive and intuitive nature of medical diagnosis, attempts have been made to develop neural networks and other expert systems that aid in this process. The application of neural networks to medical diagnosis has been reported. For example, neural networks have been used to aid in the diagnosis of cardiovascular disorders (see, e.g., Baxt (1991) "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction," Annals of Internal Medicine 115:843; Baxt (1992) "Improving the Accuracy of an Artificial Neural Network Using Multiple Differently Trained Networks," Neural Computation 4:772; Baxt (1992), "Analysis of the clinical variables that drive decision in an artificial neural network trained to identify the presence of myocardial infarction," Annals of Emergency Medicine 21:1439; and Baxt (1994) "Complexity, chaos and human physiology: the justification for non-linear neural computational analysis," Cancer Letters 77:85). Other medical diagnostic applications include the use of neural networks for cancer diagnosis (see, e.g., Maclin, et al. (19910 "Using Neural Networks to Diagnose Cancer" Journal of Medical Systems 15:11-9; Rogers, et al. (1994) "Artificial Neural Networks for Early Detection and Diagnosis of Cancer" Cancer Letters 77:79-83; Wilding, et al. (1994) "Application of Backpropogation Neural Networks to Diagnosis of Breast and Ovarian Cancer" Cancer Letters 77:145-53), neuromuscular disorders (Pattichis, et al. (1995) "Neural Network Models in EMG Diagnosis", IEEE Transactions on Biomedical Engineering 42:5:486-495), and chronic fatigue syndrome (Solms, et al. (1996) "A Neural Network Diagnostic Tool for the Chronic Fatigue Syndrome", International Conference on Neural Networks, Paper No. 108). These methodologies, however, fail to address significant

issues relating to the development of practical diagnostic tests for a wide range of conditions and does not address the selection of input variables.

Computerized decision-support methods other than neural networks have been reported for their applications in medical diagnostics, including knowledge-based expert systems, including MYCIN (Davis, et al., "Production Systems as a Representation for a Knowledge-based Consultation Program", Artificial Intelligence, 1977; 8: 1: 15-45) and its progeny TEIRESIAS, EMYCIN, PUFF, CENTAUR, VM, GUIDON, SACON, ONCOCIN and ROGET. MYCIN is an interactive program that diagnoses certain infectious diseases and prescribes anti-microbial therapy. Such knowledge-based systems contain factual knowledge and rules or other methods for using that knowledge, with all of the information and rules being pre-programmed into the system's memory rather than the system developing its own procedure for reaching the desired result based upon input data, as in neural networks. Another computerized diagnosis method is the Bayesian network, also known as a belief or causal probabilistic network, which classifies patterns based on probability density functions from training patterns and *a priori* information. Bayesian decision systems are reported for uses in interpretation of mammograms for diagnosing breast cancer (Roberts, et al., "MammoNet: A Bayesian Network diagnosing Breast Cancer", Midwest Artificial Intelligence and Cognitive Science Society Conference, Carbondale, IL, April 1995) and hypertension (Blinowska, et al. (1993) "Diagnostica -- A Bayesian Decision-Aid System -- Applied to Hypertension Diagnosis", IEEE Transactions on Biomedical Engineering 40:230-35) Bayesian decision systems are somewhat limited in their reliance on linear relationships and in the number of input data points that can be handled, and may not be as well suited for decision-support involving non-linear relationships between variables. Implementation of Bayesian methods using the

processing elements of a neural network can overcome some of these limitations (see, e.g., Penny, et al. (1996) In "Neural Networks in Clinical Medicine", Medical Decision-support, 1996; 16:4: 386-98). These methods have been used, by mimicking the physician, to diagnose disorders in which important variables are input into the system. It, however, would be of interest to use these systems to improve upon existing diagnostic procedures.

Endometriosis

Endometriosis is the growth of uterine-like tissue outside of the uterus. It affects about 15-30 percent of reproductive age women. The cause(s) of endometriosis are not known, but it may result from retrograde menstruation, the reflux of endometrial tissue and cells (menstrual debris) from the uterus into the peritoneal cavity. While retrograde menstruation is thought to occur in most or all women, it is unclear why some women develop endometriosis and others do not.

Not all women with endometriosis exhibit symptoms or suffer from the disease. The extent or severity of endometriosis does not correlate with symptoms. Some women with severe disease may be completely asymptomatic, whereas others with minimal disease may suffer from excruciating pain. Symptoms, such as infertility, pelvic pain, dysmenorrhea and past occurrence of endometriosis, that have been associated with endometriosis often occur in women who do not have endometriosis. In other instances, these symptoms may be present, and the women do have endometriosis. Although an association between these symptoms and endometriosis appears to exist, the correlation is far from perfect, the interplay with these and other factors are complex. Clinicians often perform diagnostic laparoscopies on patients whom they believe are excellent candidates for having endometriosis based a combination of the above indications. Endometriosis, however, is not

present in a significant proportion of these women. Thus, endometriosis represents an example of a disease state in which a physician must draw upon experience using a complex set of information to formulate a diagnosis. The validity of the diagnosis is related to the experience and ability of the physician.

As a result, determining if a woman has endometriosis from symptoms alone has not been possible. Within the medical community, the diagnosis of endometriosis is confirmed only by direct visualization of endometrial lesions during surgery. Many physicians often impose a further restriction and demand that the suspected lesions be verified as being endometrial-like (glands and stroma) using histology on endometrial biopsied tissue. Thus, a non-invasive diagnostic test for endometriosis would be of significant benefit.

Therefore, it is an object herein to provide a non-invasive diagnostic aid for endometriosis. It is also an object herein to provide methods to select important variables to be used in decision-support systems to aid in diagnosis of endometriosis and other disorders and conditions. It is also an object herein to identify new variables, identify new biochemical tests and markers for diseases and to design to new diagnostic tests that improve upon existing diagnostic methodologies.

SUMMARY OF THE INVENTION

Methods using decision-support systems for the diagnosis of and for aiding in the diagnosis of diseases, disorders and other medical conditions are provided. The methods provided herein, include a method of using patient history data and identification of important variables to develop a diagnostic test; a method for identification of important selected variables; a method of designing a diagnostic test; a method of evaluating the usefulness of diagnostic test; a method of expanding clinical utility of diagnostic test, and a method of selecting a course of

treatment by predicting the outcome of various possible treatments. Also provided are disease parameters or variables to aid in the diagnosis of disorders, including any disorders that are difficult to diagnose, such as endometriosis, predicting pregnancy related events, such as the likelihood of delivery within a particular time period, and other such disorders relevant to women's health. It is understood that although women's disorders are exemplified herein, the methods herein are applicable to any disorder or condition.

Also provided are means to use neural network training to guide the development of the tests to improve their sensitivity and specificity, and to select diagnostic tests that improve overall diagnosis of, or potential for, a disease state or medical condition. Finally, a method for evaluating the effectiveness of any given diagnostic test is described.

Thus, provided herein is a method for identifying variables or sets of variables that aid in the diagnosis of disorders or conditions. In the methods for identifying and selection of important variables and generating systems for diagnosis, patient data or information, typically patient history or clinical data are collected and variables based on this data are identified. For example, the data includes information for each patient regarding the number of pregnancies each patient has had. The extracted variable is, thus, number of pregnancies. The variables are analyzed by the decision-support systems, exemplified by neural networks, to identify important or relevant variables.

Methods are provided for developing medical diagnostic tests using computer-based decision-support systems, such as neural networks and other adaptive processing systems (collectively, "data mining tools"). The neural networks or other such systems are trained on the patient data and observations collected from a group of test patients in whom the condition is known or suspected; a subset or subsets of relevant variables

are identified through the use of a decision-support system or systems, such as a neural network or a consensus of neural networks; and another set of decision-support systems is trained on the identified subset(s) to produce a consensus decision-support system based test, such as a neural net-based test for the condition. The use of consensus systems, such as consensus neural networks, minimizes the negative effects of local minima in decision-support systems, such as neural network-based systems, thereby improving the accuracy of the system.

Also, to refine or improve performance, the patient data can be augmented by increasing the number of patients used. Also biochemical test data and other data may be included as part of additional examples or by using the data as dditional variables prior to the variable selection process.

The resulting systems are used as an aid in diagnosis. In addition, as the systems are used patient data can be stored and then used to further train the systems and to develop systems that are adapted for a particular genetic population. This inputting of additional data into the system may be implemented automatically or done manually. By doing so the systems continually learn and adapt to the particular environment in which they are used. The resulting systems have numerous uses in addition to diagnosis, which includes assessing the severity of a disease or disorder, predicting the outcome of a selected treatment protocol. The systems may also be used to assess the value of other data in a diagnostic procedure, such as biochemical test data and other such data, and to identify new tests that are useful for diagnosing a particular disease.

Thus, also provided are methods for improving upon existing biochemical tests, identifying relevant biochemical tests and for developing new biochemical tests to aid in diagnosis of disorders and conditions. These methods involve assessing the effect of a particular test or a potential new test on the performance of the decision-support system based test. If addition of information from the test improves performance, such test will have relevance in diagnosis.

The disorders and conditions that are of particular interest herein and to which the methods herein may be readily applied, are gyneco-logical conditions and other conditions that impact on fertility, including but not limited to endometriosis, infertility, prediction of pregnancy-related events, such as the likelihood of delivery within a particular time period, and pre-eclampsia. It is understood, however, that the methods herein are applicable to any disorder or condition.

The methods are exemplified with reference to neural networks, however, it is understood that other data mining tools, such as expert systems, fuzzy logic, decision trees, and other statistical decision-support systems which are generally non-linear, may be used. Although the variables provided herein are intended to be used with decision-support systems, once the variables are identified, then a person, typically a physician, armed with knowledge the important variables can use them to aid in diagnosis in the absence of a decision-support system or using a less complex linear system of analysis.

As shown herein, variables or combinations thereof that heretofore were not known to be important in aiding in diagnosis are identified. In addition, patient history data, without supplementing biochemical test data, can be used to diagnose or aid in diagnosing a disorder or condition when used with the decision-support systems, such as the neural nets provided herein. Furthermore, the accuracy of the diagnosis with or

without biochemical data may be sufficient to obviate the need for invasive surgical diagnostic procedures.

Also provided herein is a method of identifying and expanding clinical utility of diagnostic test. The results of a particular test, particular one that had heretofore not been considered of clinical utility with respect to the disorder or condition of interest, are combined with the variables and used with the decision-support system, such as a neural net. If the performance, the ability to correctly diagnose a disorder, of the system is improved by addition of the results of the test, then the test will have clinical utility or a new utility.

Similarly, the resulting systems can be used to identify new utilities for drugs or therapies and also to identify uses for particular drugs and therapies. For example, the systems can be used to select subpopulations of patients for whom a particular drug or therapy is effective. Thus, methods for expanding the indication for a drug or therapy and identifying new drugs and therapies are provided.

In specific embodiments, neural networks are employed to evaluate specific observation values and test results, to guide the development of biochemical or other diagnostic tests, and to provide the decision-support functionality for the test.

A method for identification of important variables (parameters) or sets thereof for use in the decision-support systems is also provided. This method, while exemplified herein with reference to medical diagnosis, has broad applicability in any field, such as financial analysis, in which important parameters or variables are selected from among a plurality.

In particular, a method for selecting effective combinations of variables is provided. The method involves: (1) providing a set of "n" candidate variables and a set of "selected important variables", which

initially is empty; (2) ranking all candidate variables based on a chi square and sensitivity analysis; (3) taking the highest "m" ranked variables one at a time, where m is from 1 up to n, and evaluating each by training a consensus of neural nets on that variable combined with the current set of important variables; (4) selecting the best of the m variables, where the best variable is the one that gives the highest performance, and if it improves performance in comparison to the performance of the selected important variables, adding it to the "selected important variable" set, removing it from the candidate set and continuing processing at step (3), otherwise going to step (5); (5) if all variables on the candidate set have been evaluated, the process is complete, otherwise continue taking the next highest "m" ranked variables one at a time, and evaluating each by training a consensus of neural nets on that variable combined with the current set of important selected variables and performing step (4). The final set of important selected variables will contain a plurality, typically more than three to five or more variables.

In a particular embodiment, the sensitivity analysis involves: (k) determining an average observation value for each of the variables in an observation data set; (l) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output; (m) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable; running the modified example in the decision-support system in the forward mode and recording the output as the modified output; (n) squaring the difference between the normal output and the modified output and accumulating it as a total for each variable, in which this total is designed the selected variable total for each variable; (o) repeat steps (m) and (n) for each variable in the example; (p) repeating steps (l)-(n) for each example in the

data set, where each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output. This total will be used to rank each variable according to its relative contribution to the determination of the decision-support system output.

As shown herein, computer-based decision-support systems such as neural networks reveal that certain input factors, which were not initially considered to be important, can influence an outcome. This ability of a neural network to reveal the relevant input factors permits its use in guiding the design of diagnostic tests. Thus a method of designing a diagnostic test, and a method of evaluating utility of diagnostic test are also provided. In each instance, the data from the test or possible test is added to the input of the decision-support system. If the results are improved when the data are included in the input, then the diagnostic test may have clinical utility. In this manner, tests that heretofore were not known to be of value in diagnosis of a particular disorder are identified, or new tests can be developed. Neural networks can add robustness to diagnostic tests by discounting the effects of spurious data points and by identifying other data points that might be substituted, if any.

Networks are trained on one set of variables and then clinical data from diagnostic or biochemical test data and/or additional patient information are added to the input data. Any variable that improves the results compared to their absence is (are) selected. As a result, particular tests that heretofore were of unknown value in diagnosing a particular disorder can be shown to have relevance. For example, the presence or absence of particular spots on a western blot of serum antibodies can be

correlated with a disease state. Based on the identity of particular spots (i.e., antigens) new diagnostic tests can be developed.

An example of the application of the prediction technology to aid in the diagnosis of disease and more particularly the use of neural network techniques with inputs from various information sources to aid in the diagnosis of the disease endometriosis is provided. A trained set of neural networks operative according to a consensus of networks in a computer system is employed to evaluate specific clinical associations, for example obtained by survey, some of which may not generally be associated with a disease condition. This is demonstrated with an exemplary disease condition endometriosis, and factors used to aid in the diagnosis of endometriosis are provided. The neural network training is based on correlations between answers to questions furnished by physicians of a significant number of clinical patients whose disease condition has been surgically verified, herein termed clinical data.

A plurality of factors, twelve to about sixteen, particularly a set of fourteen factors, in a specific trained neural network extracted from a collection of over forty clinical data factors have been identified as primary indicia for endometriosis. The following set of parameters: age, parity (number of births), gravidity (number of pregnancies), number abortions, smoking (packs/day), past history of endometriosis, dysmenorrhea, pelvic pain, abnormal pap/dysplasia, history pelvic surgery, medication history, pregnancy hypertension, genital warts and diabetes were identified as being significant. Other similar sets of parameters were also identified. Subsets of these variables also may be employed in diagnosing endometriosis.

In particular, any subset of the selected set of parameters, particularly the set of fourteen variables, that contain one (or more) of the

following combinations of three variables can be used with a decision-support system for diagnosis of endometriosis:

a) number of births, history of endometriosis, history of pelvic surgery;

b) diabetes, pregnancy hypertension, smoking;

c) pregnancy hypertension, abnormal pap smear/dysplasia, history of endometriosis;

d) age, smoking, history of endometriosis;

e) smoking, history of endometriosis, dysmenorrhea;

f) age, diabetes, history of endometriosis;

g) pregnancy hypertension, number of births, history of endometriosis;

h) Smoking, number of births, history of endometriosis;

i) pregnancy hypertension, history endometriosis, history of pelvic surgery;

j) number of pregnancies, history of endometriosis, history of pelvic surgery;

k) number of births, abnormal PAP smear/dysplasia, history of endometriosis;

l) number of births, abnormal PAP smear/dysplasia, dysmenorrhea;

m) history of endometriosis, history of pelvic surgery, dysmenorrhea; and

n) number of pregnancies, history of endometriosis, dysmenorrhea.

Diagnostic software and exemplary neural networks that use the variables for diagnosis of endometriosis are also provided. The software generates a clinically useful endometriosis index.

In other embodiments, the performance of a diagnostic neural network system used to test for endometriosis is enhanced by including variables based on biochemical test results from a relevant biochemical test as part of the factors (herein termed biochemical test data, which includes tests from analyses and data such as vital signs, such as pulse rate and blood pressure) used for training the network. An exemplary network that results therefrom is an augmented neural network that employs fifteen input factors, including results of the biochemical test and the fourteen clinical parameters. The set of weights of the eight augmented neural networks differ from the set of weights of the eight clinical data neural networks. The exemplified biochemical test employs an immuno-diagnostic test format, such as the ELISA diagnostic test format.

The methodology applied to endometriosis as exemplified herein can be similarly applied and used to identify factors for other disorders, including, but not limited to gynecological disorders and female-associated disorders, such as, for example, infertility, prediction of pregnancy related events, such as the likelihood of delivery within a particular time period, and pre-eclampsia. Neural networks, thus, can be trained to predict the disease state based on the identification of factors important in predicting the disease state and combining them with biochemical data.

The resulting diagnostic systems may be adapted and used not only for diagnosing the presence of a condition or disorder, but also the severity of the disorder and as an aid in selecting a course of treatment.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a flow chart for developing a patient-history-based diagnostic test process.

FIGURE 2 is a flow chart for developing a biochemical diagnostic test.

FIGURE 3 is a flow chart of the process for isolating important variables.

FIGURE 4 is a flow chart on the process of training one or a set of neural networks involving a partitioning of variables.

FIGURE 5 is a flow chart for developing a biochemical diagnostic test.

FIGURE 6 is a flow chart for determining the effectiveness of a biochemical diagnostic test.

FIGURE 7 is a schematic diagram of a neural network trained on clinical data of the form used for the consensus network of a plurality of neural networks.

FIGURE 8 is a schematic diagram of a second embodiment of a neural network trained on clinical data augmented by test results data of the form used for the consensus of eight neural networks.

FIGURE 9 is a schematic diagram of a processing element at each node of the neural network.

FIGURE 10 is a schematic diagram of a consensus network of eight neural networks using either the first or second embodiment of the neural network.

FIGURE 11 is a depiction of an exemplary interface screen of the user interface in the diagnostic endometriosis index.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference.

As used herein, a decision-support system, also referred to as a "data mining system" or a "knowledge discovery in data system", is any system, typically a computer-based system, that can be trained on data to classify the input data and then subsequently used with new input data to make decisions based on the training data. These systems include, but are not limited, expert systems, fuzzy logic, non-linear regression analysis, multivariate analysis, decision tree classifiers, Bayesian belief networks and, as exemplified herein, neural networks.

As used herein, an adaptive machine learning process refers to any system whereby data are used to generate a predictive solution. Such processes include those effected by expert systems, neural networks, and fuzzy logic.

As used herein, expert system is a computer-based problem solving and decision-support system based on knowledge of its task and logical rules or procedures for using the knowledge. Both the knowledge and the logic are entered into the computer from the experience of human specialists in the area of expertise.

As used herein, a neural network, or neural net, is a parallel computational model comprised of densely interconnected adaptive processing elements. In the neural network, the processing elements are configured into an input layer, an output layer and at least one hidden layer. Suitable neural networks are known to those of skill in this art (see, e.g., U.S. Patents 5,251,626; 5,473,537; and 5,331,550, Baxt

(1991) "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction," Annals of Internal Medicine 115:843; Baxt (1992) "Improving the Accuracy of an Artificial Neural Network Using Multiple Differently Trained Networks," Neural Computation 4:772; Baxt (1992) "Analysis of the clinical variables that drive decision in an artificial neural network trained to identify the presence of myocardial infarction," Annals of Emergency Medicine 21:1439; and Baxt (1994) "Complexity, chaos and human physiology: the justification for non-linear neural computational analysis," Cancer Letters 77:85).

As used herein, a processing element, which may also be known as a perceptron or an artificial neuron, is a computational unit which maps input data from a plurality of inputs into a single binary output in accordance with a transfer function. Each processing element has an input weight corresponding to each input which is multiplied with the signal received at that input to produce a weighted input value. The processing element sums the weighted inputs values of each of the inputs to generate a weighted sum which is then compared to the threshold defined by the transfer function.

As used herein, transfer function, also known as a threshold function or an activation function, is a mathematical function which creates a curve defining two distinct categories. Transfer functions may be linear, but, as used in neural networks, are more typically non-linear, including quadratic, polynomial, or sigmoid functions.

As used herein, backpropogation, also known as backprop, is a training method for neural networks for correcting errors between the target output and the actual output. The error signal is fed back through the processing layer of the neural network, causing changes in the weights of the processing elements to bring the actual output closer to the target output.

As used herein, Quickprop is a backpropogation method that was proposed, developed and reported by Fahlman ("Fast Learning Variations on Back-Propagation: An Empirical Study", Proceedings on the 1988 Connectionist Models Summer School, Pittsburgh, 1988, D. Touretzky, et al., eds., pp.38-51, Morgan Kaufmann, San Mateo, CA; and, with Lebriere, "The Cascade-Correlation Learning Architecture", Advances in Neural Information Processing Systems 2,(Denver, 1989), D. Touretzky, ed., pp. 524-32. Morgan Kaufmann, San Mateo, CA).

As used herein, diagnosis refers to a predictive process in which the presence, absence, severity or course of treatment of a disease, disorder or other medical condition is assessed. For purposes herein, diagnosis will also include predictive processes for determining the outcome resulting from a treatment.

As used herein, biochemical test data refers to the results of any analytical methods, which include, but are not limited to:, immunoassays, bioassays, chromatography, data from monitors, and imagers; measurements and also includes data related to vital signs and body function, such as pulse rate, temperature, blood pressure, the results of, for example, EKG, ECG and EEG, biorhythm monitors and other such information. The analysis can assess for example, analytes, serum markers, antibodies, and other such material obtained from the patient through a sample.

As used herein, patient historical data refers to data obtained from a patient, such as by questionnaire format, but typically does not include biochemical test data as used herein, except to the extent such data is historical, a desired solution is one that generates a number or result whereby a diagnosis of a disorder can be generated.

As used herein, wherein a training example includes the observation data for a single diagnosis, typically the observation data related to one patient.

As used herein, the parameters identified from patient historical data are herein termed observation factors or values or variables. For example, patient data will include information with respect to individual patient's smoking habits. The variable associated with that will be smoking.

As used herein, partition means to select a portion of the data, such as 80%, and use it for training a neural net and to use the remaining portion as test data. Thus, the network is trained on all but one portion of the data. The process can then be repeated and a second network trained. The process is repeated until all partitions are used as used as test data and training data.

As used herein, the method of training by partitioning the available data into a plurality of subsets is generally referred to as the "holdout method" of training. The holdout method is particularly useful when the data available for network training is limited.

As used herein, training refers to the process in which input data are used to generate a decision-support system. In particularly, with reference to neural nets, training is a trial-and-error process involving a series of iterative adjustments to the processing element weights so that a particular processing element provides an output which, when combined with the outputs of other processing elements, generates a result which minimizes the resulting error between the outputs of the

neural network and the desired outputs as represented in the training data.

As used herein, a variable selection process is a systematic method whereby combinations of variables that yield predictive results are selected from any available set. Selection is effected by maximizing predictive performance of subsets such that addition of additional variables does not improve the result. The preferred methods provided herein advantageously permit selection of variables without considering all possible combinations.

As used herein, a candidate variable is a selected item from collected observations from a group of test patients for the diagnostic embodiments or other records, such as financial records, that can be used with the decision-support system. Candidate variables will be obtained by collecting data, such as patient data, and categorizing the observations as a set of variables.

As used herein, important selected variables refer to variables that enhance the network performance of the task at hand. Inclusion of all available variables does not result in the optimal neural network; some variables, when included in network training, lower the network performance. Networks that are trained only with relevant parameters result in increased network performance. These variables are also referred to herein as a subset of relevant variables.

As used herein, ranking refers to a process in which variables are listed in an order for selection. Ranking may be arbitrary or, preferably, is ordered. Ordering may be effected, for example, by a statistical analysis that ranks the variables in order of importance with respect to the task, such as diagnosis, or by a decision-support system based analysis.

Ranking may also be effected, for example, by human experts, by rule based systems, or any combination of any of these methods.

As used herein, a consensus of neural networks refers to the linear combination of outputs from a plurality of neural networks where the weight on each is outputs is determined arbitrarily or set to an equal value.

As used herein, a greedy algorithm is a method for optimizing a data set by determining whether to include or exclude a point from a given data set. The set begins with no elements and sequentially selects an element from the feasible set of remaining elements by myopic optimization, in which, given any partial solution, another value that improves the object the most is selected.

As used herein, a genetic algorithm is a method that begins with an initial population of randomly generated neural networks which are run through a training cycle and ranked according to their performance in reaching the desired target. The poor-performing networks are removed from the population, with the fitter networks being retained and selected for the crossover process to offspring that retain the desirable characteristics of the parent networks.

As used herein, performance of a system is said to be improved or higher when the results more accurately predict or deterimine a particular outcome. It is also to be understood that the performance of a system will typically be better as more training examples are used. Thus, the systems herein will improve over time as they are used and more patient data is accumulated and then added to the systems as training data.

As used herein, sensitivity = TP/(TP + FN); specificity is TN/(TN + FP), where TP = true positives; TN = true negatives; FP = false positives; and FN = false negative. Clinical sensitivity measures how well a test detects patients with the disease; clinical specificity measures how

well a test correctly identifies those patients who do not have the disease.

As used herein, positive predictive value (PPV) is TP/(TP+FP); and negative predictive value (NPV) is TN/(TN+FN). Positive predictive value is the likelihood that a patient with a positive test actually has the disease, and negative predictive value is the likelihood that a patient with a negative test result does not have the disease.

As used herein, fuzzy logic is an approach to deal with systems that cannot be described precisely. Membership functions (membership in a data set) are not binary in fuzzy logic systems; instead membership function may take on fractional values. Therefore, an element can be simultaneously contained in two contradictory sets, albeit with different coefficients of set membership. Thus, this type of approach is useful for answering questions in which there is no yes or answer. Thus, this type of logic is suitable for categorizing responses from patient historical questionnaires, in which the answer is often one of degree.

1. **General considerations and general methodology**

It has been determined that a number of techniques can be used to train neural networks for analyzing observation values such as patient history and/or biochemical information. Depending upon the characteristics of the available data and the problem to be analyzed, different neural network training techniques can be used. For example, where large amounts of training inputs are available, methodology may be adopted to eliminate redundant training information.

As shown herein, neural networks may also reveal that certain input factors that were not initially considered to be important can influence an outcome, as well as reveal that presumably important factors are not outcome determinative. The ability of neural networks to reveal the relevant and irrelevant input factors permit their use in guiding the

design of a diagnostic test. As shown herein, neural networks, and other such data mining tools, are a valuable advance in diagnostics, providing an opportunity to increase the sensitivity and specificity of a diagnostic test. As shown herein, care must be taken to avoid the potential of poor-accuracy answer due to the phenomenon of local minima. The methods herein provide a means to avoid this problem or at least minimize it.

In developing the developing diagnostic procedures, and in particular diagnostic tests that are based solely or in part on patient information, a number of problems have been solved. For example, there is generally a limited amount of data because there is a limited number of patients where training data are available. To solve this, as described below, the patient information is partitioned when training the network. Also, there is generally a large number of input observation factors available for use in connection with the available data, so methods for ranking and selecting observations were developed.

Also, there are generally large number of binary (true/false) input factors in the available patient data, but these factors are generally sparse in nature (values that are positive or negative in only a small percentage of cases of the binary input factors in the available patient data). Also there is a high degree of overlap between the positive and negative factors of the condition being diagnosed.

These characteristics and others impact the choice of procedures and methods used to develop a diagnostic test. These problems are addressed and solved herein.

2. Development of patient history diagnostic test

Diagnostic test

Methods for diagnosis based solely on patient history data are provided. As demonstrated herein, it is possible to provide decision-support system that rely only on patient history information but that aid in

diagnosis. Consequently, the resulting systems can then be used to improve the predictive ability of biochemical test data, to identify new disease markers, to develop biochemical tests, to identify tests that heretofore were not thought to be predictive of a particular disorder.

The methods may also be used to select an appropriate course of treatment by predicting the result of selected course of treatment and to predict status following therapy. The input variables for training would be derived from, for example, electronic patient records, that indicate diagnoses and other available data, including selected treatments and outcomes. The resulting decision-support system would then be used with all available data to, for example, categorize women into different classes that will respond to different treatments and predict the outcome of a particular treatment. This permits selection of a treatment or protocol most likely to be successful.

Similarly, the systems can be used to identify new utilities for drugs or therapies and also to identify uses for particular drugs and therapies. For example, the systems can be used to select subpopulations of patients for whom a particular drug or therapy is effective. Thus, methods for expanding the indication for a drug or therapy and identifying new drugs and therapies are provided.

Collection of patient data, generation of variables and overview

To exemplify the methods herein, Fig. 1 sets forth a flow chart for developing a patient-history-based diagnostic test process. The process begins with collection of patient history data (Step A). Patient history data or observation values are obtained from patient questionnaires, clinical results, in some instances diagnostic test results, and patient medical records and supplied in computer-readable form to a system operating on a computer. In the digital computer, the patient history data are categorized into a set of variables of two forms: binary (such as

true/false) values and quantitative (continuous) values. A binary-valued variable might include the answer to the question, "Do you smoke?" A quantitative-valued variable might be the answer to the question, "How many packs per day do you smoke?" Other values, such as membership functions, may also be useful as input vehicles.

The patient history data will also include a target or desired outcome variable that would be assumed to be indicative of the presence, absence, or severity of the medical condition to be diagnosed. This desired outcome information is useful for neural network training. The selection of data to be included in the training data can be made with the knowledge or assumption of the presence, severity, or absence of the medical condition to be diagnosed. As noted herein, diagnosis may also include assessment of the progress and/or effectiveness of a therapeutic treatment.

The number of variables, which can be defined and thus generated, can be unwieldy. Binary variables are typically sparse in that the number of positive (or negative) responses is often a small percentage of the overall number of responses. Thus, in instances in which there is a large number of variables and a small number of patient cases available in a typical training data environment, steps are taken to isolate from the available variables a subset of variables important to the diagnosis (Step B). The specific choice of the subset of variables from among the available variables will affect the diagnostic performance of the neural network.

The process outlined herein has been found to produce a subset of variables which is comparable or superior in sensitivity and reliability to the subset of variables typically chosen by a trained human expert, such as a physician. In some instances, the variables are prioritized or placed in order of rank or relevance.

Thereafter, the final neural networks to be used in the diagnostic procedure are trained (Step C). In preferred embodiments, a consensus (i.e. a plurality) of networks are trained. The resulting networks form the decision-support functionality for the completed patient history diagnostic test (Step D).

Method for isolation of important variables

A method for isolation of important variables is provided herein. The method permits sets of effective variables to be selected without comparing every possible combination of variables. The important variables may be used as the inputs for the decision-support systems.

Isolation of important or relevant variables -ranking the variables

Figure 3 provides a flow chart of the process for isolating the important or relevant variables (Step E) within a diagnostic test. Such a process is typically conducted using a digital computer system to which potentially relevant information has been provided. This procedure ranks the variables in order of importance using two independent methods, then selects a subset of the available variables from the uppermost of the ranking. As noted above, other ranking methods can be used by those of skill in the art in place of chi square or sensitivity analysis. Also, if x is set to N (the total number of candidate variables), then ranking can be arbitrary.

The system trains a plurality of neural networks on the available data (Step I), as explained hereinafter, then generates a sensitivity analysis over all trained networks to determine to what extent each input variable was used in the network to perform the diagnosis (Step J). A consensus sensitivity analysis of each input variable is determined by averaging the individual sensitivity analysis results for each of the networks

trained. Based upon sensitivity, a ranking order for each of the variables available from the patient history information is determined (Step K).

Ranking the variables

In preferred embodiments, the variables are ranked using a statistical analysis, such as a chi square analysis, and/or a decision-support system-based analysis, such as a sensitivity analysis. A sensitivity analysis and chi square analysis are used, in the exemplary embodiment to rank variables. Other statistical methods and/or decision-support system-based , including but not limited to regression analysis, discriminant analysis and other methods known to those of skill in the art, may be used. The ranked variables may be used to train the networks, or, preferably, used in the method of variable selection provided herein.

The method employs a sensitivity analysis in which each input is varied and the corresponding change in output is measured (see, also, Modai, et al., (1993) "Clinical Decisions for Psychiatric Inpatients and Their Evaluation by Trained Neural Networks", Methods of Information in Medicine 32:396-99; Wilding et al. (1994) "Application of Backpropogation Neural Networks to Diagnosis of Breast and Ovarian Cancer", Cancer Letters 77:145-53; Ruck et al. (1990) "Feature Selection in Feed-Forward Neural Networks", Neural Network Computing 20:40-48; and Utans, et al. (1993) "Selecting Neural Network Architectures Via the Prediction Risk: Application to Corporate Bond Rating Prediction"; Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street. Washington, D.C., IEEE Computer Society Press. pp. 35-41; Penny et al. (1996) In "Neural Networks in Clinical Medicine", Medical Decision-support 4:386-398). Such methods, which have heretofore not been used to select important variables, as described herein. For example, sensitivity analysis has bee reported to be used to develop a statistical approach to determine the relationships between the

variables, but not for selection of important variables (see, Baxt et al. (1995) "Bootstrapping Confidence Intervals for Clinical Input Variable Effects in a Network Trained to Identify the Presence of Myocardial Infarction," Neural Computation 7: 624-38). Any such sensitivity analyses may be used as described herein as part of the selection of important variables as an aid to diagnosis.

Step K, Fig. 3, provides an outline of the sensitivity analysis. Each network or a plurality of trained neural networks (networks N_1 through N_n) is run in the forward mode (no training) for each training example S_x (input data group for which true output is known or suspected; there must be at least two training examples), where "x" is the number of training examples. The output of each network N_1-N_n for each training example S_x is recorded, i.e., stored in memory. A new training example is defined containing the average value for each input variable within all training examples. One at a time, each input variable within each original training example S_x is replaced with its corresponding average value $V_{1(avg)}$ through $V_{y(avg)}$, where "y" is the number of variables, and the modified training example S_x' is again executed through the multiple networks to produce a modified output for each network for each variable. The differences between the output from the original training example S_x and the modified output for each input variable are the squared and summed (accumulated) to obtain individual sums corresponding to each input variable. To provide an illustration, for example, for 10 separate neural networks N_1-N_{10} and 5 different training examples S_1-S_5, each having 15 variables V_1-V_{15}, each of the 5 training examples would be run through the 10 networks to produce 50 total outputs. Taking variable V_1 from each of the training examples, an average value $V_{1(avg)}$ is calculated. This averaged variable $V_{1(avg)}$ is substituted into each of the 5 training examples to create modified training examples S_1'-S_5' and they are again

run through the 10 networks. Fifty modified output values are generated by the networks N_1-N_{10} for the 5 training examples, the modification being the result of using the average value variable $V_{1(avg)}$. The difference between each of the fifty original and modified output values is calculated, i.e., the original output from training S_4 in network N_6: $OUT(S_4N_6)$ is subtracted from the modified output from training example S_4 in network N_6, $OUT(S_4'N_6)$. That difference value is squared $[OUT(S_4'N_6)-OUT(S_4N_6)]^2_{V1}$. This value is summed with the squared difference values for all combinations of networks and training examples for the iteration in which variable V_1 was substituted with its average value $V_{1(avg)}$, i.e.,

$$\sum_{x=1}^{5}\sum_{n=1}^{10}[OUT(S_x' N_n)-OUT(S_x N_n)]^2_{V1}.$$

Next, the process is repeated for variable #2, finding the differences between the original and modified outputs for each combination of network and training example, squaring, then summing the differences. This process is repeated for each variable until all 15 variables have been completed.

Each of the resultant sums is then normalized so that if all variables contributed equally to the single resultant output, the normalized value would be 1.0. Following the preceding example, the summed squared differences for each variable are summed to obtain a total summed squared difference for all variables. The value for each variable is divided by the total summed square difference to normalize the contribution from each variable. From this information, the normalized value for each variable can be ranked in order of importance, with higher relative numbers indicating that the corresponding variable has a greater influence on the output. The sensitivity analysis of the input variables is used to

indicate which variables played the greatest roles in generating the network output.

It has been found herein that using consensus networks to perform sensitivity analysis improves the variable selection process. For example, if two variables are highly correlated, a single neural network trained on the data might use only one of the two variables to produce the diagnosis. Since the variables are highly correlated, little is gained by including both, and the choice of which to include is dependent on the initial starting conditions of the network being trained. Sensitivity analysis using a single network might show that only one, or the other, is important. Sensitivity analysis derived from a consensus of multiple networks, each trained using different initial conditions, may reveal that both of the highly correlated variables are important. By averaging the sensitivity analysis over a set of neural networks, a consensus is formed that minimizes the effects of the initial conditions.

Chi-square contingency table

When dealing with sparse binary data, a positive response on a given variable might be highly correlated to the condition being diagnosed, but occur so infrequently in the training data that the importance of the variable, as indicated by the neural network sensitivity analysis, might be very low. In order to catch these occurrences, the Chi-square contingency table is used as a secondary ranking process. A 2X2 contingency table Chi-square test on the binary variables, where each cell of the table is the observed frequency for the combination of the two variables (Fig. 3, Step F) is performed. A 2X2 contingency table Chi-square test is performed on the continuous variables using optimal thresholds (which might be empirically-determined) (Step G)> The binary and continuous variables that have been based on Chi-square analysis are ranked (Step H).

The standard Chi-square 2X2 contingency table operative on the binary variables (Step F) is used to determine the significance of the relationship between a specific binary input variable and the desired output (as determined by comparing the training data with the known single output result). Variables that have a low Chi-square value are typically unrelated to the desired output.

For variables that have continuous values, a 2X2 contingency table can be constructed (Step G) by comparing the continuous variable to a threshold value. The threshold value is modified experimentally to yield the highest possible Chi-square value.

The Chi-square values of the continuous variables and of the binary variables can then be combined for common ranking (Step H). A second level of ranking can then be performed that combines the Chi-square-ranked variables with the sensitivity-analysis-ranked variables (Step L). This combining of rankings allows variables that are significantly related to the output but that are sparse (i.e, values that are positive or negative in only a small percentage of cases) to be included in the set of important variables. Otherwise, important information in such a non-linear system could easily be overlooked.

Selection of important variables from among the ranked variables

As noted above, important variables are selected from among the identified variables. Preferably the selection is effected after ranking the variables at which time a second level ranking process is invoked. A method for identification of important variables (parameters) or sets thereof for use in the decision-support systems is also provided. This method, while exemplified herein with reference to medical diagnosis, has broad applicability in any field, such as financial analysis and other

endeavors that involve statistically-based prediction, in which important parameters or variables are selected from among a plurality.

In particular, a method for selecting effective combinations of variables is provided. After (1) providing a set of "n" candidate variables and a set of "selected important variables", which initially is empty; and (2) ranking all candidate variables based on a chi square and sensitivity analysis, as described above, the method involves: (3) taking the highest "m" ranked variables one at a time, where m is from 1 up to n, and evaluating each by training a consensus of neural nets on that variable combined with the current set of important variables; (4) selecting the best of the m variables, where the best variable is the one that most improves performance, and if it improves performance, adding it to the "selected important variable" set, removing it from the candidate set and continuing processing at step (3) otherwise continuing by going to step (5); (5) if all variables on the candidate set have been evaluated, the process is complete, otherwise continue taking the next highest "m" ranked variables one at a time, and evaluating each by training a consensus of neural nets on that variable combined with the current set of important selected variables and performing step (4).

In particular, the second level ranking process (Step L) starts by adding the highest ranked variable from the sensitivity analysis (Step K) to the set of important variables (Step H). Alternatively, the second level ranking process could be started with an empty set and then testing the top several (x) variables from each of the two sets of ranking. This second level ranking process uses the network training procedure (Step I) on a currently selected partition or subset of variables from the available data to train a set of neural networks. The ranking process is a network training procedure using the current set of "important" variables (which generally will initially be empty) plus the current variable being ranked or

tested for ranking, and uses a greedy algorithm to optimize the set of input variables by myopically optimizing the input set based upon the previously identified important variable(s), to identify the remaining variable(s) which improve the output the most.

This training process is illustrated in Fig. 4. The number of inputs used by the neural network is controlled by excluding inputs which are found to not contribute significantly to the desired output, i.e., the known target output of the training data. A commercial computer program, such as ThinksPro™ neural networks for Windows™ (or TrainDos™ the DOS version) by Logical Designs Consulting, Inc, La Jolla, California, or any other such program that one of skill in the art can develop may be used to vary the inputs and train the networks.

A number of other commercially available neural network computer programs may be used to perform any of the above operations, including Brainmaker™, which is available from California Scientific Software Co., Nevada Adaptive Solutions, Beaverton, OR; Neural Network Utility/2™, from NeuralWare, Inc., Pittsburgh, PA; NeuroShell™ and Neuro-Windows™, from Ward Systems Group, Inc., Frederick, MD. Other types of data mining tools, i.e., decision-support systems, that will provide the function of variable selection and network optimization may be designed or other commercially available systems may be used. For example, NeuroGenetic Optimizer™ from BioComp Systems, Inc., Redmond, WA; and Neuro Forecaster/GENETICA, from New Wave Intelligent Business Systems (NIBS) Pte Ltd., Republic of Singapore, use genetic algorithms that are modelled on natural selection to eliminate poor-performing nodes within network population while passing on the best performing rates to offspring nodes to "grow" an optimized network and to eliminate input variables which do not contribute significantly to the outcome. Networks based on genetic algorithms use mutation to avoid trapping in local

minima and use crossover processes to introduce new structures into the population.

Knowledge discovery in data (KDD) is another data mining tool, decision-support system, designed to identify significant relationship is that exist among variables, and are useful when there are many possible relationships. A number of KDD systems are commercially available including Darwin™, from Thinking Machines, Bedford, MA; Mineset™, from Silicon Graphics, Mountain View, CA, and Eikoplex™ from Ultragem Data Mining Company, San Francisco, CA. (Eikoplex™ has been used to provide classification rules for determining the probability of the presence of heart disease.) Others may be developed by those of skill in the art.

Proceeding with the ranking procedure, if, for example, x is set to 2, then the top two variables from each of the two ranking sets will be tested by the process (Fig. 3, Steps L, S), and results are checked to see if the test results show improvement (Step T). If there is an improvement, the single best performing variable is added to the set of "important" variables, and then that variable is removed from the two rankings (Fig. 3, Step U) for further testing (Step S). If there is no improvement, then the process is repeated with the next x variables from each set until an improvement is found or all of the variables from the two sets have been tested. This process is repeated until either the source sets are empty, i.e., all relevant or important variables have been included in the final network, or all of the remaining variables in the sets being tested are found to be below the performance of the current list of important variables. This process of elimination greatly reduces the number of subsets of the available variables which must be tested in order to determine the set of important variables. Even in the worst case, with ten available variables, the process would test only 34 subsets where $x = 2$ and only 19 subsets of the 1024 possible combinations if

$x = 1$. Thus, where there are 100 available variables, only 394 subsets would be tested where $x = 2$. The variables from the network with the best test performance are thus identified for use (Fig. 3, Step V).

Then the final set of networks is trained to perform the diagnosis (Fig. 4, Steps M, N, Q, R). Typically, a number of final neural networks are trained to perform the diagnosis. It is this set of neural networks (a that can form the basis of a deliverable product to the end user. Since different initial conditions (initial weights) can produce differing outputs for a given network, it is useful to seek a consensus. (The different initial weights are used to avoid error from trapping in local minima.) The consensus is formed by averaging the outputs of each of the trained networks which then becomes the single output of the diagnostic test.

Training a consensus of networks

Fig. 4 illustrates the procedure for the training of a consensus of neural networks. It is first determined whether the current training cycle is the final training step (Step M). If yes, then all available data are placed into the training data set (i.e., $P = 1$) (Step N). If no, then the available data are divided into P equal-sized partitions, randomly selecting the data for each partition (Step O). In an exemplary embodiment, for example five partitions, e.g., P_1-P_5, are created from the full set of available training data. Then two constructions are undertaken (Step P). First, one or more of the partitions are copied to a test file and the remaining partitions are copied to a training file. Continuing the exemplary embodiment of five partitions, one of the partitions, e.g., P_1,, representing 20% of the total data set, is copied to the test file. The remaining four files, P_2-P_4, are identified as training data. A group of N neural networks is trained using the training partitions, each network having different starting weights (Step Q). Thus, in the exemplary embodiment, there will be 20 networks ($N = 20$) with starting weights

selected randomly using 20 different random number seeds. Following completion of training for each of the 20 networks, the output values of all 20 networks are averaged to provide the average performance on the test data for the trained networks. The data in the test file (partition P_1) is then run through the trained networks to provide an estimate of the performance of the trained networks. The performance is typically determined as the mean squared error of prediction, or misclassification rate. A final performance estimate is generated by averaging the individual performance estimates of each network to produce a completed consensus network (Step R). This method of training by partitioning the available data into a plurality of subsets is generally referred to as the "holdout method" of training. The holdout method is particularly useful when the data available for network training is limited.

Test set performance can be empirically maximized by performing various experiments that identify network parameters that maximize test set performance. The parameters that can be modified in this set of experiments are 1) the number of hidden processing elements, 2) the amount of noise added to the inputs, 3) the amount of error tolerance, 4) the choice of learning algorithm, 5) the amount of weights decay, and 6) the number of variables. A complete search of all possible combinations is typically not practical, due to the amount of processing time that is required. Accordingly, test networks are trained with training parameters chosen empirically via a computer program, such as ThinksPro™ or a user developed program, or from the results of existing test results generated by others who are working in the field of interest. Once a "best" configuration is determined, a final set of networks can be trained on the complete data set.

3. **Development of biochemical diagnostic test**

A similar technique for isolating variables may be used to build or validate a biochemical diagnostic test, and also to combine a biochemical diagnostic test data with the patient history diagnostic test to enhance the reliability of a medical diagnosis.

The selected biochemical test can include any test from which useful diagnostic information may be obtained in association with a patient and/or patient's condition. The test can be instrument or non-instrument based and can include the analysis of a biological specimen, a patient symptom, a patient indication, a patient status, and/or any change in these factors. Any of a number of analytical methods can be employed and can include, but are not limited to, immunoassays, bioassays, chromatography, monitors, and imagers. The analysis can assess analytes, serum markers, antibodies, and the like obtained from the patient through a sample. Further, information concerning the patient can be supplied in conjunction with the test. Such information includes, but is not limited to, age, weight, blood pressure, genetic history, and the other such parameters or variables.

The exemplary biochemical test developed in this embodiment employs a standardized test format, such as the Enzyme Linked Immunosorbent Assay or ELISA test, although the information provided herein may apply to the development of other biochemical or diagnostic tests and is not limited to the development of an ELISA test (see, e.g., Molecular Immunology: A Textbook, edited by Atassi et al. Marcel Dekker Inc., New York and Basel 1984, for a description of ELISA tests). Information important to the development of the ELISA test can be found in the Western Blot test, a test format that determines antibody reactivity to proteins in order to characterize antibody profiles and extract their properties.

A Western Blot is a technique used to identify, for example, particular antigens in a mixture by separating these antigens on polyacrylamide gels, blotting onto nitrocellulose, and detecting with labeled antibodies as probes. (See, for example, Basic and Clinical Immunology, Seventh Edition, edited by Stites and Terr, Appleton and Large 1991, for information on Western Blots.) It is, however, sometimes undesirable to employ the Western Blot test as a diagnostic tool. If instead, ranges of molecular weight that contain relevant information to the diagnosis can be pre-identified then this information can be "coded" into an equivalent ELISA test.

In this example, the development of an effective biochemical diagnostic test is dependent upon the availability of Western Blot data for the patients for which the disease condition is known or suspected. Referring to Fig. 5, Western Blot data are used as a source (Step W), and the first step in processing the Western Blot data are to pre-process the Western Blot data for use by the neural network (Step X). Images are digitized and converted to fixed dimension training records by using a computer to perform the spline interpolation and image normalization. It is necessary to align images on a given gel based only on information in the image in order to use data from multiple Western Blot tests. Each input of a neural network needs to represent a specific molecular weight or range of molecular weights accurately. Normally, each gel produced contains a standards image for calibration, wherein the proteins contained are of a known molecular weight, so that the standards image can also be used for alignment of images contained within the same Western Blot. For example, a standard curve can be used to estimate the molecular weight range of other images on the same Western Blot and thereby align the nitrocellulose strips.

The process for alignment of images is cubic spline interpolation. This is a method which guarantees smooth transitions at the data points represented by the standards. To avoid possible performance problems due to extrapolation, termination conditions are set so that extrapolation is linear. This alignment step minimizes the variations in the estimates of molecular weight for a given band on the output of the Western Blot.

The resultant scanned image is then processed to normalize the density of the image by scaling the density so that the darkest band has a scaled density of 1.0 and the lightest band is scaled to 0.0. The image is then processed into a fixed length vector of numbers which become the inputs to a neural network, which at the outset must be trained as hereinafter explained.

A training example is built in a process similar to that previously described where the results generated from the processing of the Western Blot data are trained (Step Y). To minimize the recognized problems of dependency on starting weights, redundancy among interdependent variables, and desensitivity resulting from overtraining a network, it is helpful to train a set of neural networks (consensus) on the data by the partitioning method discussed previously.

From the sensitivity analysis of the training runs on the processed Western Blot data, regions of significantly contributing molecular weights (MW) can be determined and identified (Step AA). As part of the isolation step, inputs in contiguous regions are preferably combined into "bins" as long as the sign of the correlation between the input and the desired output is the same. This process reduces the typical 100-plus inputs produced by the Western Blot, plus the other inputs, to a much more manageable number of inputs of less than about twenty.

In a particular embodiment, it may be found that several ranges of molecular weight may correlate with the desired output, indicative of the

condition being diagnosed. A correlation may be either positive or negative. A reduced input representation may be produced by using a Gaussian region centered on each of the peaks found in the Western Blot training, with a standard deviation determined so that the value of the Gaussian was below 0.5 at the edges of the region.

In a specific embodiment, the basic operation to generate the neural network input is to perform a convolution between the Gaussian and the Western Blot image, using the log of the molecular weight for calculation.

The data may be tested using the holdout method, as previously described. For example, five partitions might be used where, in each partition, 80% of the data are used for training and 20% of the data are used for testing. The data are shuffled so that each of the partitions is likely to have examples from each of the gels.

Once the molecular weight regions important to diagnosis have been identified (Step AA), one or more tests for the selected region or regions of molecular weight may be built (Step AB). The ELISA biochemical test is one example. The selected region or regions of molecular weight identified as important to the diagnosis may then be physically identified and used as a component of the ELISA biochemical test. Whereas regions of the same correlation sign may, or may not, be combined into a single ELISA test, regions of differing correlation signs should not be combined into a single test. The value of such a biochemical test may then be determined by comparing the biochemical test result with the known or suspected medical condition.

In this example, the development of a biochemical diagnostic test may be enhanced by combining patient data and biochemical data in a process shown in Fig. 2. Under these conditions, the patient history diagnostic test is the basis for the biochemical diagnostic test. As

explained herein, the variables that are identified as important variables are combined with data derived from the Western Blot data in order to train a set of neural networks to be used to identify molecular weight regions that are important to a diagnosis.

Referring to Fig. 2, Western Blot data are used as a source (Step W) and pre-processed for use by the neural network as described previously (Step X). A training example is built in a process similar to that previously described wherein the important variables from the patient history data and the results generated from the processing of the Western Blot data are combined and are trained using the combined data (Step Y). In parallel, networks are trained on patient history data, as described above (Step Z).

To minimize the recognized problems of dependency on starting weights, redundancy among interdependent variables, and desensitivity resulting from overtraining a network, it was found that it was preferable to train a set of neural networks (consensus set) on the data by the partitioning method. From the sensitivity analysis of the training runs on patient history data alone and on combined data, regions of significantly contributing molecular weights can be determined and identified as previously described (Step AA). As a further step in the isolation process, a set of networks is thereafter trained using as inputs the combined patient history and bin information in order to isolate the important bins for the Western Blot data. The "important bins" represent the important regions of molecular weight related to the diagnosis considering the contribution of patient history information. These bins are either positively or negatively correlated with the desired output of the diagnosis.

Once the molecular weight regions important to diagnosis have been identified (Step AA), one or more tests for the selected region or

regions may be built and validated as previously described (Step AB). The designed ELISA tests are then produced and used to generate ELISA data for each patient in the database (Step AC). Using ELISA data and the important patient history data as input, a set of networks is trained using the partition approach as described above (Step AE). The partition approach can be used to obtain an estimate of the lower bound of the biochemical test. The final training (Step AE) of a set of networks, i.e., the networks to be used as a deliverable product, is made using all available data as part of the training data. If desired, new data may be used to validate the performance of the diagnostic test (Step AF). The performance on all the training data becomes the upper bound on the performance estimate for the biochemical test. The consensus of the networks represents the intended diagnostic test output (AG). This final set of neural networks can then be used for diagnosis.

4. Improvement of neural network performance

An important feature of the decision-support systems, as exemplified with the neural networks, and methods provided herein is the ability to improve performance. The training methodology outlined above may be repeated as more information becomes available. During operation, all input and output variables are recorded and augment the training data in future training sessions. In this way, the diagnostic neural network may adapt to individual populations and to gradual changes in population characteristics.

If the trained neural network is contained within an apparatus that allows the user to enter the required information and outputs to the user the neural network score, then the process of improving performance through use may be automated. Each entry and corresponding output is retained in memory. Since the steps for retraining the network can be

encoded into the apparatus, the network can be re-trained at any time with data that are specific to the population.

5. Method for evaluating the effectiveness of a diagnostic test course of treatment

Typically, the effectiveness or usefulness of a diagnostic test is determined by comparing the diagnostic test result with the patient medical condition that is either known or suspected. A diagnostic test is considered to be of value if there is good correlation between the diagnostic test result and the patient medical condition; the better the correlation between the diagnostic test result and the patient medical condition, the higher the value placed on the effectiveness of the diagnostic test. In the absence of such a correlation, a diagnostic test is considered to be of lesser value. The systems provided herein, provide a means to assess the effectiveness of a biochemical test by determining whether the variable that corresponds to that test is an important selected variable. Any test that yields data that improves the performance of the system is identified.

A method by which the effectiveness of a diagnostic test may be determined, independent of the correlation between the diagnostic test result and the patient medical condition (Fig. 6) is described below. A similar method may be used to assess the effectiveness of a particular treatment.

In one embodiment, the method compares the performance of a patient history diagnostic neural network trained on patient data alone, with the performance of a combined neural network trained on the combination of patient historical data and biochemical test data, such as ELISA data. Patient history data are used to isolate important variables for the diagnosis (Step AH), and final neural networks are trained (Step AJ), all as previously described. In parallel, biochemical test results are

provided for all or a subset of the patients for whom the patient data are known (Step AK), and a diagnostic neural network is trained on the combined patient and biochemical data by first isolating important variables for the diagnosis (Step AL), and subsequently training the final neural networks (Step AM), all as previously described.

The performance of the patient history diagnostic neural network derived from Step AJ is then compared with the performance of the combined diagnostic neural network derived from Step AM, in Step AN. The performance of a diagnostic neural network may be measured by any number of means. In one example, the correlations between each diagnostic neural network output to the known or suspected medical condition of the patient are compared. Performance can then be measured as a function of this correlation. There are many other ways to measure performance. In this example, any increase in the performance of the diagnostic neural network derived from Step AM over that derived from Step AJ is used as a measure of th effectiveness of the biochemical test.

A biochemical test in this example, and any diagnostic test in general, that lacks sufficient correlation between that test result and the known or suspected medical condition, is traditionally considered to be of limited utility. Such a test may be shown to have some use through the method described above, thereby enhancing the effectiveness of that test which otherwise might be considered uninformative. The method described herein serves two functions: it provides a means of evaluating the usefulness of a diagnostic test, and also provides a means of enhancing the effectiveness of a diagnostic test.

6. Application of the methods to identification of variables for diagnosis and development of diagnostic tests

The methods and networks provided herein provide a means to, for example, identify important variables, improve upon existing biochemical tests, develop new tests, assess therapeutic progress, and identify new disease markers. To demonstrate these advantages, the methods provided have been applied to endometriosis and to pregnancy related events, such as the likelihood of labor and delivery during a particular period.

Endometriosis

The methods described herein, have provided a means to develop a non-invasive methodology for the diagnosis of endometriosis. In addition, the methods herein provide means to develop biochemical tests that provide data indicative of endometriosis, and also to identify and develop new biochemical tests.

The methodology for variable selection and use of decision-support systems, has been applied to endometriosis. A decision-support system, in this instance, a consensus of neural networks, has been developed for the diagnosis of endometriosis. In the course of this development, which is detailed in the EXAMPLES, it was found that it was possible to develop neural networks capable of aiding in the diagnosis of endometriosis that only rely on patient historical data, i.e., data that can be obtained from a patient by questionnaire format. It was found that biochemical test data could be used to enhance the performance of a particular network, but it was not essential to its value as a diagnostic tool. The variable selection protocol and neural nets provide a means to select sets of variables that can be inputted into the decision-support system to provide a means to diagnose endometriosis. While some of the identified variables, include those that have traditionally been associated with endometriosis, others

of the variables have not. In addition, as noted above, variables, such as pelvic pain and dysmenorrhea that have been associated with endometriosis are not linearly correlated with it to permit diagnosis.

Exemplary decision-support system are described in the Example. For example, one neural net, designated pat07 herein, is described in Example 14. Comparison of the output of the pat07 network output with the probability of having endometriosis yields a positive correlation (see Table 1). The pat07 network can predict the likelihood of a woman having endometriosis based on her pat07 score. For example, if a woman has a pat07 score of 0.6, then she has a 90% probability of having endometriosis; if her pat07 score is 0.4, then she has a 10% probability of having endometriosis. The dynamic range of pat07 output when applied to the database was about 0.3 to about 0.7. Theoretically, the output values can range from 0 to 1, but values below 0.3 or above 0.7 were not observed. Over 800 women have been evaluated using the pat07 network, and its performance can be summarized as follows:

TABLE 1

Pat07 Score	Endometriosis (% of Total)
< 0.40	10
0.40 - 0.45	30
0.45 - 0.55	50
0.55 - 0.60	70
> 0.60	90

The pat07 network score is interpreted as the likelihood of having endometriosis, and not whether or not a woman is diagnosed with endometriosis. The likelihood is based on the relative incidence of endometriosis found in each score group. For example, in the group of women with pat07 network score of 0.6 or greater, 90% of these women

had endometriosis, and 10% of these women did not. This likelihood relates to the population of women at infertility clinics.

Software programs have been developed that contain the pat07 network.

One program, referred to as adezacrf.exe, provides a single screen windows interface that allows the user to obtain the pat07 network score for a woman. User enters values for all 14 variables, pat07 network score is calculated following every keystroke. Another program, designated adzcrf2.exe, is almost exactly the same as adezacrf.exe, except that it allows for one additional input: the value of an ELISA test.

This program and network is a specific example of a method of expanding clinical utility of a diagnostic test. The ELISA test results did not correlate with endometriosis. By itself, the ELISA test does not have clinical utility. As another input parameter, the ELISA test improved network performance, so that one may assert that incorporating the ELISA result as an input for network analysis expanded the clinical utility of that ELISA test. Another program (provided herein in the Appendix II, designated adzcrf2.exe, provides a multiple screen windows interface that allows the user to obtain the pat07 network score for a woman. The multiple data entry screens guides the user to enter all patient historical data, and not just those parameters required as inputs for pat07. Pat07 score is calculated after all data are entered and accepted as correct by the user. This program also saves the data entered in *.fdb files, can import data, calculate pat07 scores on imported data, and export data. The user can edit previously entered data. All three of the above programs serve as specific examples of the diagnostic software for endometriosis.

Figure 11 illustrates an exemplary interface screen used in the diagnostic software. The display 1100, which is provided as a MicroSoft Window™-type display, provides a template for entry of numerical values for each of the important variables which have been determined for

diagnosis of endometriosis. Input of data to perform a test is accomplished using a conventional keyboard alone, or in combination with a computer mouse, a trackball or joystick. For purposes of this description, a mouse/keyboard combination will be used. Each of the text boxes 1101-1106 is for entry of numerical values representative of the important variables Age (box 1101); Number of Pregnancies (box 1102); Number of Births (box 1103); Number of Abortions (box 1104); Number of Packs of Cigarettes Smoked per Day (box 1105); and ELISA test results (box 1106). To enter the subject patient's age, the user moves the mouse so that the pointer on the screen is in box 1101, then clicks on that location. Entry of the number representative of the patient's age is done using the keyboard. The remaining boxes are accessed by pointing and clicking at the selected box.

Boxes 1107-1115 are important selected variables for which the data are binary, i.e., either "yes" or "no". The boxes and the variables are correlated as follows:

BOX	VARIABLE
1107	Past History of Endometriosis
1108	Dysmenorrhea
1109	Hypertension During Pregnancy
1110	Pelvic Pain
1111	Abnormal PAP/Dysplasia
1112	History of Pelvic Surgery
1113	Medication History
1114	Genital Warts
1115	Diabetes

A "yes" to any one of these variables can be indicated by pointing at the corresponding box and clicking the mouse button to indicate an "X" within the box.

The network automatically processes the data after every keystroke, so changes will be seen in the output values displayed in text boxes 1118-1120 after every entry into the template 1100. Text box 1118, labelled "Endo" provides consensus network output for the presence of endometriosis; text box 1119, labelled "No Endo" provides consensus network output for the absence of endometriosis; and text box 1120 provides a relative score indicative of whether or not the patient has endometriosis. It is noted that the score in the text box 1120 is an artificial number derived from boxes 1118 and 1119 that makes it easier for the physician to interpret results. As presently set, a value in this box in the positive range up to 25 is indicative of having endometriosis, and a value in the negative range down to -25 will be indicative of not having endometriosis. The selected transformations permits the physician to readily interpret the pat07 output more readily.

As described in the Examples, the pat07 is not the only network that is predictive of endometriosis. Other networks, designated pat08 through pat23a have been developed. These are also predictive of endometriosis. All these networks perform very similarly, and can readily be used in place of pat07. Thus, by following the methodology used to develop pat07, other similarly functioning neural nets can be and have been developed. Pat08 and pat09 are the most similar to pat07: these networks were developed by following the protocol outlined above, and were allowed to select important variables from the same set as that used for development of pat07.

It was found that the initial weighting of variables can have effects on the outcome of the variable selection procedure, but not in the ultimate diagnostic result. Pat08 and pat09 used the same database of patient data as pat07 to derive the disease relevant parameters. Pat10 through pat23a were training runs originally designed to elucidate the

importance of certain parameters: history of endometriosis, history of pelvic surgery, dysmenorrhea and pelvic pain. For development of these, the importance of a variable was assessed by withholding that variable from the variable selection process. It was found that the variable selection process and training the final consensus networks, network performance did not significantly deteriorate.

Thus, although a particular variable, or set of variables, may have appeared to be significant in predicting endometriosis, networks trained in the absence of such variables do not have a markedly reduced ability to predict endometriosis. These results to demonstrate (1) the effectiveness of the methodology for variable selection and consensus network training;, and (2) the adaptability of networks in general. In the absence of one type of data, the network found other variable(s) from which to extract that information. In the absence of one variable, the network selected different variables in its place and maintained performance.

Patients suspected of having endometriosis typically must undergo exploratory surgery to diagnose the disease. The ability to diagnose this disorder reliably using patient history information and optionally biochemical test data, such as western blot data, provides a highly desirable alternative to surgery. The methods herein and identified variables provide a means to do so.

Data related to the diagnosis of the disease of endometriosis has been gathered. The data includes, patient history data, western blot data and ELISA data. Application of the methodology herein, as shown in the EXAMPLES, demonstrated that patient history data alone can be predictive of endometriosis.

To assess the performance of the variable selection protocol and to ascertain where the 14 variable network (pat07) ranked (in performance) compared to all possible combinations of the 14 variables, networks were

trained on every possible combination of the variables (16,384 combinations). Also, the variable selection protocol was applied to the set of 14 variables. From among the 14, 5 variables were selected. These are pregnancy hypertension, number of births, abnormal PAP/dysplasia, history of endo and history of pelvic surgery. This combination ranked as the 68th best performing combination out the 16,384 (99.6 percentile) possible combinations, thereby demonstrating the effectiveness of variable selection protocol. Also, the combination that includes all 14 variables was ranked 718th out the 16,384 possible combinations (95.6 percentile).

These results also show that subsets of the 14 variables are useful. In particular, any subset of the selected set of parameters, particularly the set of fourteen variables, that contain one (or more) of the following combinations of three variables can be used with a decision-support system for diagnosis of endometriosis:

a) number of births, history of endometriosis, history of pelvic surgery;

b) diabetes, pregnancy hypertension, smoking;

c) pregnancy hypertension, abnormal pap smear/dysplasia, history of endometriosis;

d) age, smoking, history of endometriosis;

e) smoking, history of endometriosis, dysmenorrhea;

f) age, diabetes, history of endometriosis;

g) pregnancy hypertension, number of births, history of endometriosis;

h) Smoking, number of births, history of endometriosis;

I) pregnancy hypertension, history endometriosis, history of pelvic surgery;

j) number of pregnancies, history of endometriosis, history of pelvic surgery;

k) number of births, abnormal PAP smear/dysplasia, history of endometriosis;

l) number of births, abnormal PAP smear/dysplasia, dysmenorrhea;

m) history of endometriosis, history of pelvic surgery, dysmenorrhea; and

n) number of pregnancies, history of endometriosis, dysmenorrhea.

As shown in the examples, other sets of important selected variables that perform similarly to the enumerated 14 variables can be obtained. Other smaller subsets thereof may be also be identified.

Predicting pregnancy related events, such as the likelihood of delivery within a particular time period

The methods herein may be applied to any disorder or condition, and are particularly suitable for conditions in which no diagnostic test can be adequately correlated or for which no biochemical test or convenient biochemical test is available. For example, the methods herein have been applied to predicting pregnancy related events, such as the likelihood of delivery within a particular time period.

Determination of impending birth is of importance, example, for increasing neonatal survival of infants born before 34 weeks. The presence of fetal fibronectin in secretion samples from the vaginal cavity or the cervical canal from a pregnant patient after week 20 of pregnancy is associated with a risk of labor and delivery before 34 weeks. Methods and kits for screening for fetal fibronectin in secretion samples from the vaginal cavity or the cervical canal for a pregnant patient after week 20 of pregnancy are available (see, U.S. Patent Nos. 5,516,702, 5,468,619,

and 5,281,522, and 5,096,830; see, also U.S. Patent Nos. 5,236,846, 5,223,440, 5,185,270).

The correlation between the presence of fetal fibronectin in these secretions and the labor and delivery before 34 weeks is not perfect; there are significant false-negative and false-positive rates. Consequently, to address the need for methods to assess the likelihood of labor and delivery before 34 weeks and to improve the predictability of the available tests, the methods herein have been applied to development of a decision-support system that assesses the likelihood of certain pregnancy related events. In particular, neural nets for predicting delivery before (and after) 34 weeks of gestation have been developed. Neural networks and other decision-support systems developed as described herein can improve the performance of the fetal fibronectin (fFN) test by lowering the number of false positives. The results, which are shown in EXAMPLE 13, demonstrate that use of the methods herein can improve the diagnostic utility of existing tests by improving predictive performance.

As described above, these methods may be used to identify tests previously not thought to be relevant to a disease, condition or disorder, and to design new tests and identify new disease markers.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

Evaluation of Patient History Data for Relevant Variables

This examples demonstrates selection of the candidate variables.

Requirements

Evaluation of the patient history to determine which variables are relevant to the diagnosis. This example is performed by performing a sensitivity analysis on each of the variables to be used in the diagnosis.

Two methods can be used to perform this analysis. The first is to train a network on all the information and determine from the network weights, the influence of each input on the network output. The second method is to compare the performance of two networks, one trained with the variable included and the second trained with the variable eliminated. This training would be performed for each of the suspected relevant variables. Those that did not contribute to the performance will be eliminated. These operations are performed to lower the dimension of the inputs to the network. When training with limited amounts of data, a lower input dimension will increase the generalization capabilities of the network.

Analysis of Data

The data used for this example included 510 patient histories. Each record contained 120 text and numeric fields. Of these fields, 45 were identified as being known before surgery and always containing information. These fields were used as the basic available variables for the analysis and training of networks. A summary of the variables used in this example was as follows:

1.	age (preproc)- Preprossed to normalize the age to fall between 0 and 1	24.	Uterine/Tubal Anomalies26. Ectopic Preg
2.	Diabetes	25.	Fibroids
3.	Preg Induced DM	26.	Ectopic Preg.
4.	Hypertension	27.	Dysfunctional Uterine Bld.
5.	Preg hyperplasia	28.	Ovarian Cyst
6.	Autoimmune Disease	29.	Polycystic Ovarian Synd
7.	Transplant	30.	Abnormal PAP/Dysplasia
8.	Packs/Day	31.	Gyn Cancer

9.	Drug Use	32.	Other HX
10.	#Preg	33.	Past HX of Endo
11.	#Birth	34.	Hist of Pelvic Surgery
12.	#Abort	35.	Medication History
13.	Hist of Infertility	36.	Current Exog. Hormone
14.	Ovulatory	37.	Pelvic Pain
15.	Anovulatory	38.	Abnormal Pain
16.	Unknown	39.	Menstrual Abnormalities
17.	Oligoovulatory	40.	Dysmenorrhea
18.	Hormone Induced	41.	Dyspaarunia
19.	Herpes	42.	Infertility
20.	Genital Warts	43.	Adnneaxal Masses/Thickening
21.	Other sexually transmitted diseases (STD)	44.	Undetermined
22.	Vag Infections	45.	Other Symptoms
23.	Pelvic inflammatory disease (PID)		

Methodology Used

The most commonly used method for determining the importance of variables is to train a neural network on the data with all the variables included. Using the trained network as the basis, a sensitivity analysis is performed on the network and the training data. For each training example the network is run in the forward mode (no training). The network outputs were recorded. The for each input variable, the network is rerun with the variable replaced by it's average value over the training example. The difference in output values is squared and accumulated. This process is repeated for each training example. The resulting sums are then normalized so that the sum of the normalized values equals the

number of variables. In this way, if all variables contribute equally to the output, their normalized value would be 1.0. The normalized value can then be ranked in order of importance.

There are several problems with the above approach. First, it is dependent on the neural network solution found. If different network starting weights are used, a different ranking might be found. Secondly, if two variables are highly correlated, the use of either would contain sufficient information. Depending on the network training run, only one of the variables might be identified as important. The third problem is that an overtrained network can distort the true importance of a variable.

To minimize the effects of the above problems, several networks were trained on the data. The training process was refined to produce the best possible test set performance so that the networks had learned the underlying relationship between the inputs and the desired outputs. By the end of this process, both a good set of networks would be available and the training configuration for the final trained networks would be established. The sensitivity analysis was performed on each of the networks trained and the normalized values were averaged. For this example, a training run included 15 networks trained on five partitions of the available data using a holdout method.

Once the ranking of variables has been established, test runs are made to determine the effects of eliminating variables has on the test set performance. Eliminating a variable that has a small contribution should lower the test set performance. When overtraining is an issue, due to limited training data, eliminating variables can actually improve the test set performance. To save on processing time, groups of variables can be eliminated in a test based on the ranking.

Results

The rankings or variables are as follows and are reported for networks trained in run pat05.

	01.	35	Medication History
5	02.	33.	Past history of Endo
	03.	11	#Birth
	04.	37.	Pelvic Pain
	05.	40.	Dysmenorrhea
	06.	34.	Hist of Pelvic Surgery
10	07.	1.	Age(preproc)
	08.	13.	Hist of Infert
	09.	8.	Packs/Day
	10.	36.	Current Exog. Hormones
	11.	42.	Infertility
15	12.	18.	Hormone Induced
	13.	15.	Anovulatory
	14.	14.	Ovulatory
	15.	43.	Adnnexal Masses/Thickening
	16.	45.	Other Symptoms
20	17.	30	Abnormal PAP/Dysplasia
	18.	26.	Ectopic Preg.
	19.	19.	Herpes
	20.	39.	Menstrual Abnormalities
	21.	12.	#Abort
25	22.	41.	Dyspaarunia
	23.	24.	Uterine/Tubal Anomalies
	24.	31.	Gyn Cancer
	25.	32.	Other history
	26.	10.	#Preg
30	27.	28.	Ovarian Cyst
	28.	25.	Fibroids
	29.	22.	Vag Infections
	30.	16.	Unknown

31. 27. Dysfunctional Uterine Bld.
32. 38. Abdominal Pain
33. 5. Preg hyperplasia
34. 9. Drug Use
35. 20. Genital Warts
36. 3. Preg Induced DM
37. 4. Hypertension
38. 21. Other STD
39. 23. PID
40. 44. Undetermined
41. 2. Diabetes
42. 17. Oligoovulatory
43. 6. Autoimmune Disease
44. 29. Polycystic Ovarian Synd
45. 7. Transplant

Subsets of the variables were tested and the final set of 14 variables to be used to train the pat07 networks [see, Examples 13 and 14]. Some variables were used that were not in the above top 14. This occurred to improve the test set performance. The rankings for the pat07 networks are as follows:

01. 10. Past history of Endo
02. 6. #Birth
03. 14. Dysmenorrhea
04. 1. Age(preproc)
05. 13. Pelvic Pain
06. 11. Hist of Pelvic Surgery
07. 4. Packs/Day
08. 12. Medication History
09. 5. #Preg
10. 7. #Abort
11. 9. Abnormal PAP/Dysplasia
12. 3. Preg hyperplasia

13. 8. Genital Warts
14. 2. Diabetes

Conclusions

The set of variables identified in this example appear to be reasonable based on the testing and information.

EXAMPLE 2

Train Networks on Patient History Data

This example, using the above 14 variables, demonstrates methods for setting and optimizing various parameters.

Requirements

At the completion of the above examples, train a set of networks on the reduced patient history and record their performance. Experiments were run to determine the best configuration and parameters for training of the networks. An analysis of the performance was performed to determine the number of false positive and false negatives, to see if a given subset of the patients can be reliably diagnosed. Since there was limited data, the estimated performance was determined by leaving out small portions of the database (25%) for testing and training on the remaining data. This method was repeated until all of the data has been used as test data in one of the networks. The combined results on the test data then become the performance estimate. A final network was be trained using all of the available data as training data.

Methodology Used

When dealing with small training examples, the holdout method is effective in providing test information useful in determining network configuration and parameter settings. In order to maximize the data available for training without a big increase in processing time a 20% holdout was used instead of the proposed 25%. This produced 5

partitions of the data instead of 4 and made 80% of the data for training in each partition.

To minimize the effects of the random starting weights, several networks were trained in the full training runs. In theses runs three networks were trained in each of the five partitions of data, each from a different random start. The outputs of the networks were averaged to form a consensus result that has a lower variance than could be obtained from a single network.

A number of experiments were performed to find network parameters that would maximize the test set performance. The parameters modified in this process were as follows;

1. Number of hidden processing elements.
2. Amount of Noise added to the inputs.
3. Amount of Error tolerance.
4. Learning algorithm used.
5. Amount of Weights decay used.
6. Number of input variables used.

A complete search of all possible combinations of 45 variables was not feasible because of the amount of CPU time needed for the tests. Test networks were trained with parameters chosen from based on parameters known by those of skill in the art to be important in this area and based on results of prior tests. Other sets of variables would also have been suitable. Also, as shown elsewhere herein, combinations of all of the selected 14 variables have been tested. Once the best configuration was determined, a final set of networks was trained on the complete data set of 510 patients. In the final set of networks, a consensus of eight networks was made to produce the final statistics.

Results

The final holdout training run was pat06 with 14 variables. The performance on the test data was 68.23%. The full training run was pat07 with the same network configuration as pat06. The performance on the training data was 72.9%. Statistics were generated on the last training run based on the use of a cutoff in the network output values. If the network output was below the cutoff, the example was rejected from consideration. The following table is a summary of the results for the consensus of eight networks in the pat07. A test program named adzcrf was produced to demonstrate this final training.

Cutoff	0.00	0.02	0.075	0.075	0.10
Sensitivity	.828179	.835821	.872247	.943750	.956522
Specificity	.598174	.627660	.630137	.638298	.745098
PPV	.732523	.761905	.785714	.816216	.871287
NPV	.723757	.728395	.760331	.869565	.904762
Odds Ratio	2.695	3.000	3.494	4.907	7.412
% Rejected	0	10.58	26.86	50.20	71.96

PPV = positive predictive value;, NPV = negative predictive value;

EXAMPLE 3

Preprocessing and input Western Blot Data Requirements

The antigen data, from Western Blots, for the patients that was originally delivered to Logical Designs provided information on only the peak molecular weights and their associated intensities. Analysis of this data and of the original images from which the data was taken, suggests that it may be possible to use the original image digitized in a way that could provide more information to the neural network. In examining the

original images for two experiments, it preprocessing of the image data decreases the variability of the position of a specific molecular weight in the image. This preprocessing will use a polynomial fit through the standards image to produce a modified image. Preprocessing of the images will also include steps to normalize the background level and contrast of the images.

Once the preprocessing is complete, the image data could be used as is, or the peak molecular weights could be extracted. From resulting images, inputs to the neural network will be generated. As a typical image is about 1000 pixels long, methods to reduce the number of inputs will be investigated. As the image would be coded directly into the network inputs using the full or reduced dimension (resolution) images, a neural network will be trained with supervised learning to aid in the determination of the ranges of molecular weights that are related to the determination of the disease. This Example focuses on using the image as a whole in the input to the network.

Methodology Used

Using a correlation technique, similar features on images of Western blots were matched and a correlation plot was produced. From those plots, it was concluded that there was too much variation in matches on the correlation plots of two samples to accurately align the samples. Since each input of the network needed to accurately represent a molecular weight value, it was decided that only information from the standards image would be used for alignment of images.

A quadratic fit was performed on the standards image to generate a means to translate relative mobility information to molecular weight. After plotting a curve of relative mobility to the log of the molecular weight and examining the RSQR values, it was concluded that the quadratic fit was not accurate enough for performing this translation. The

calculated weight for a standard molecule varied from gel to gel using the quadratic fit.

Several methods were tried to improve the translation of relative mobility to molecular weight. Cubic spline interpolation was selected. This method guarantees smooth transitions at the data points and is rapidly calculated. The only concern is how the method performs on values of relative mobility outside the intervals covered by the standards. If termination conditions are set properly, the extrapolation problem seems to be avoided. This was the selected method.

Using spline interpolation, the images were converted to fixed dimension training records. At this point, image intensity normalization had to be considered. Two alternatives were considered. The first was to perform no normalization. The second was to process the images so that the maximum value across the image was set to 1.0 and the minimum value was set to 0.0. Networks were trained on each alternative and the results were compared. With no noise added to the inputs, the preprocessed image network had a training example performance of 97% while the performance for no preprocessing was 79%. When noise was added, the two alternative gave similar results. The choice was made to use the preprocessed images for further training runs. This choice insured that a given network input would consistently be associated with a specific molecular weight within the tolerances achievable using the Western Blot method.

Using the above choices, a series of eight neural networks were trained to provide information on the importance of different molecular weights on the prediction of the Endo present variable. In order to permit an analysis of the direction of correlation, only a single hidden processing element was used in the training. A sensitivity analysis was performed

on each of the networks and the resulting consensus was plotted using Excel.

The weights of the network were then averaged together to generate a consensus value for each weight. Since the interconnection weight from the hidden clement to the output could be either positive or negative. The weights were transformed so that all the output connections had the same sign. The weights were then averaged and the results plotted using Excel.

Results

The analysis of the Western Blot data was performed using a cubic spline interpolation for image alignment to the network inputs and Max/Min image preprocessing. Given that a certain amount of variability can be expected in the accuracy of alignment of the images, due to the Western Blot methodology, this approach is believed to give better results that the polynomial fit originally used.

The plot of the sensitivity analysis and of the weights for the final consensus networks indicated that there are regions on the Western Blot that can aid in the prediction and diagnosis of the disease. The width of the regions of positive and negative correlation, as seen in the network weights, also indicates that the results shown are significant. If the peaks had been very narrow, one would have to conclude that the peaks were artifacts of the training process, similar to overtraining, and not form the underlying process being learned. The regions that appear important are as follows:

Positive Correlation:

31503.98 - 34452.12

62548.87 - 65735.97

84279.36 - 89458.49

Negative Correlation:

19165.9 - 20142.47

50263.36 - 53352.14

67725.77 - 78614.77

While there were a number of positive and negative peaks, these appeared to be the most likely regions for inclusion into two ELISA tests. One test will focus on the positive regions and the other on the negative regions. The two resulting values can then be combined with the patient history data as input to the neural networks.

Conclusions

The neural network was able to find regions on the Western Blot that correlate with the presence of the disease.

EXAMPLE 4

Investigate Fixed Input Dimension for Western Blot Data

Requirements

Using peak molecular weights extracted from the preprocessed image, methods to reduce the varying dimension of the western blot data for a patient to a fixed dimension for the neural network were investigated. This approach is desirable in that it will have substantially fewer network inputs than the full image approach. The basic problem is that the test yields a varying number of molecular weights that might by interrelated. Comparison of results from Example and this example will indicate that patterns of molecular weights exist or if the weights are unrelated. Since there is some variability in the weight data the approach to the processing of this data will be similar to a fuzzy membership function, even though the classification will be performed with a neural network.

Additional Requirements

Fractions are identified from the Western blot data. Since production of these fractions is reproducible, the effectiveness of the use of this information is determined by processing the Western blot image data into bins corresponding to the molecular weights of the fractions.

Methodology Used

From the results of Example 4, several ranges in molecular weight were determined to correlate with the disease. A reduced input representation was produced by using a gaussian region centered on each of the peaks found in Example 5. The standard deviation of the gaussian was determined so that the value of the gaussian was below 0.5 at the edges of the region. The basic operation performed to generate the neural network input was to perform a convolution between the gaussian and the Western blot image. The calculations were all performed using the log of the molecular weight.

A separate software program was produced. The program performed the convolution on the normalized images with respect to molecular weight and intensity. The parameters for calculation of the network inputs are contained in a table in the binproc program. In binproc the mean and std. deviation are stored in the table. The program is recompiled when the table values are changed. The program has a test mode that produces an output file that allows the gaussians to be plotted an compared to the Western Blot images, using Excel. Plots of regions are included in the documentation.

When working with 36 fractions, binproc.c was again modified to translate the positions of the fractions into table values for binproc. This modified program is called fproc.d. It's purpose is to perform the spline interpolation required to normalize the molecular weight value based on the standards. Binproc2.c was produced from binproc, replacing the

mean and std. deviation tables with min. and max. tables which correspond to the endpoints of the fractions in the files supplied.

In order to test any of the data files produced by the above programs, the holdout methods was used with 80% of the data being used for training and the remaining 20% to be used for testing. Once the training data are produced form the Western Blot data, a random number column and the Patient ID column was added in the Excel spreadsheet. The data was then sorted on the random number column. This in effect shuffles the data. In this way, it is likely that each partition has examples from each of the gels. With these percentages, five separate training and test files were produced os as to allow a network performance to be estimated from the combined test set results.

Using ThinksPro™ the number of inputs used by the network could be varied by excluding inputs. Excluded inputs are not presented to the network during training. Using the sensitivity analysis as a guide, unimportant inputs are eliminated. Decreasing the dimension of the input space becomes even more important when the number of training examples is small. This method is the same as was used for eliminating variables in the patient history training runs. At the current time, this process is performed manually.

Results

In Example 5, networks trained on all the data were used to determine what ranges of molecular weights were important to the classification process. In this Example, the holdout method was used to train networks so that the test set performance could be estimated. The first set of test were based on regions identified in Example 5. The second set of tests were made using the fractions identified in the four ishgel files.

The initial consensus runs based on the top six regions found in Example 5 yielded poor performance (50%). Analysis of the input data generated, indicated that the regions used to generate the input data were too narrow to capture the important information from the image data. The regions were widened and the top ten regions from Example 5 were included instead of the top six. A test on the ten wider regions indicated slightly better performance. Using sensitivity analysis, three of the ten regions were eliminated and a complete test as run. The performance on the six of the ten wider regions improved to 54.5%.

As the number of inputs to the network was further decreased, the test set performance (estimated with the holdout method) continued to increase. The best performance was achieved from using only one of the regions, with molecular weight ranging from 66392.65 to 78614.74. The performance estimate was 58.5% on test data using the holdout method.

This process was applied again using, as a start, 36 regions based on identified fractions. There was a great deal of overlap in the 36 fractions. The top 7 fractions were determined from the 36 using sensitivity analysis. Similar performance of 58% was achieved using the subset of the fractions.

Conclusions

None of the tests yielded results that were very high. The primary reason for this is likely to be the limited amount of training data available for this Example. Results from previous Examples indicated that as the number of patients in the training sample decreased, that the performance on validation data also decreased. This relationship is illustrated in the following table.

Network	Patients	Estimated Performance
Patient History	510	68.23 (pat06)
Elisa using patient history data only	350	62.76
Western Blot	200	58.0

Better results were achieved on the ELISA/patient historical data when including the Elisa variables, even with the reduced number of patients. This showed the value of the ELISA variable.

It appears that a number of regions can be determined as being important to the classification of the disease. Substantially different sets of regions has yielded similar results, indicating that there may be patterns in the Western Blot data that indicate the presence of the disease. With a small database of patients, it is more difficult to isolate these patterns.

It is clear that increasing the size of the database for Western Blot data will improve the performance of the networks trained on this data. When combining Western Blot data with patient historical data, the input dimension of the network will increase. The increase in input dimension usually requires more training examples to ensure generalization.

EXAMPLE 5

Train Networks Using Western Blot Data

The purpose of this example was to train a set of networks to determine the performance estimate for the diagnosis using only western blot data. Experiments were run to determine the best configuration and parameters for training of the networks. The method described in Example 2 above was be used for this performance estimate. A final network was trained using all of the available data as training data. The

output of this trained network (antigen index) was used as an input to the network generated in the combined data phase.

Methodology Used

Several methods were used to find the best performing set of inputs for the available training data. From previous Examples, the use of the sensitivity analysis was shown to produce good results in identifying the importance of each of the inputs variables. The number of networks were trained on combinations of variables manually determined from the sensitivity analysis.

In preparing an automated procedure, the use of a 2x2 contingency table Chi square analysis of the variables was used to provide an alternative ranking of the importance of the variables. Since the inputs were continuous, a threshold was used for each input to generate the information needed for the contingency table. The Chi-square value varies depending on the setting of the threshold. The threshold value used in the ranking of variables was chosen to maximize the Chi square statistic.

The training runs made during the development of the automated procedure were chosen from these rankings. At the time that the training runs were made, an automated procedure had not been formulated. To save on overall processing time, only one partition of the training data was used. Combinations of variables that performed well in the first partition of the training and test data were then tried on remaining the partitions.

One method suggested in the literature for finding the best set of inputs has been to use a genetic algorithm to determine the highest performing set of inputs. Genetic algorithms typically require thousands of iterations to converge to a good solution. In working with the Western Blot data, this would represent a large amount of computer time, even

with the small training example size. For 10 variables, an enumeration of all combinations would require 1024 training runs. An alternative to the genetic algorithm was attempted. In this alternative method, a neural network was trained to predict the test set RMS Error based on the set of inputs chosen. The training examples used for this experiment were the results of training runs on the first partition of the Western Blot data. The predictor network was then tested with all combinations to determine the predicted minimum combination. The input combination was then used to train a network on the Western Blot data. The main drawback of this method and the genetic algorithm approach is that the sensitivity analysis information, found to be very effective is ignored in the process.

Results

The basic rankings for the 10 variables (bins) in the Western Blot Data are based on a consensus of 8 networks trained on the full database of 200 examples. The results are as follows:

7 : 1.182073
9 : 1.055611
3 : 1.053245
8 : 1.039028
6 : 1.027239
10 : 1.023135
4 : 0.978769
5 : 0.952821
2 : 0.899936
1 : 0.788143

The rankings of the 10 variables based on the Chi-square analysis are as follows:

3 : 4.380517
9 : 3.751625

7 : 3.372731
2 : 3.058437
6 : 3.022164
5 : 2.787982
10 : 1.614931
4 : 1.225725
1 : 0.975502
8 : 0.711958

During the analysis of the Western Blot data, a number of networks were trained on the first partition(s) of the training data. The test results are ranked below, showing the variables the were included in the training run.

Variables

1	2	3	4	5	6	7	8	9	10	Test Error	
0	0	0	1	0	1	0	0	1	1	0.49291	
0	0	1	0	0	0	1	0	1	0	0.49374	
0	0	0	0	0	1	0	0	1	1	0.49831	
1	0	0	0	0	0	1	0	0	1	0.50036	(11)
0	0	1	0	0	0	0	0	0	0	0.50145	
0	0	0	0	0	0	1	0	0	1	0.50164	(13)
0	0	0	0	0	0	0	0	0	1	0.50174	(5)
0	0	0	0	0	0	1	0	0	0	0.50182	
0	0	0	0	1	0	0	0	0	0	0.50295	(8)
0	0	1	0	1	0	1	0	0	0	0.50285	
0	0	0	0	0	0	0	0	1	0	0.50323	
0	0	0	0	0	1	0	0	0	0	0.50360	(7)
0	0	1	0	0	0	1	0	0	1	0.50587	
0	1	1	0	0	0	0	1	0	0	0.50682	(3)
0	0	1	0	0	0	0	0	1	0	0.50707	

1	2	3	4	5	6	7	8	9	10	Test Error	
0	0	0	1	1	1	0	0	1	1	0.50823	
1	0	0	0	1	1	1	0	0	0	0.50853	(2)
0	0	0	0	1	0	1	0	0	0	0.50995	(1)
1	1	0	0	0	0	0	0	0	0	0.51158	(9)
0	0	0	0	0	0	1	0	1	0	0.51163	
0	0	0	0	1	1	0	0	0	0	0.51372	(10)
0	0	0	0	0	0	0	1	1	0	0.51909	
1	0	0	1	0	0	0	1	0	0	0.52084	(4)
1	1	0	0	1	0	1	0	0	1	0.52950	
0	1	0	0	0	0	1	0	1	0	0.52978	
0	0	0	1	0	1	0	1	0	0	0.53196	
0	0	1	1	0	0	0	1	1	0	0.54841	
0	0	1	1	0	0	1	0	1	0	0.54934	(14)
0	1	0	1	0	0	0	0	0	1	0.55178	(12)
1	1	0	0	0	1	1	0	0	0	0.55290	(6)
1	0	1	0	0	0	0	1	0	1	0.56664	
0	0	1	1	0	1	0	1	1	1	0.59937	

() indicates combination was generated by prediction network training process.

In looking at the above test runs, it is clear that the more important variables in the rankings contributed to lower test set errors and that the more variables included, the lower the test set results. This shows the importance of choosing the best subset of variables in developing a high performance neural network.

Several combinations of variables were used to train networks on all partitions of the training data. The results of these runs are shown below.

VARIABLES	TIME SET PERFORMANCE
3	57.5%
3,9	53.5%
3,7,9	53.0%
4,6,9,10	57.0%

Since both rankings of variables showed 3, 7, and 9 as important, it is likely that this combination would give higher than 57.5% provided there was enough training data. Training example performance for this combination was 63.9%, showing the level of overtraining that occurred. Several of the first partition networks shown above had combinations of variables chosen by a neural network trained to predict the test performance. Those networks are indicated by a number in the last column. This number indicates the sequence in which tests were run. Combinations without a number were chosen manually from the rankings. If this process were continued the predictor network would eventually find the best combination. Since there are many factors that can effect the test set performance, it is likely that there is a lot of "noise" in the test set results. For this method to work better, a consensus approach may be needed to generate the training values for the predicted test set error. This problem would also be seen when using the consensus approach.

Conclusion

The process of using the sensitivity and contingency table rankings of variables is an effective and efficient technique for picking a set of variables to maximize the neural network performance. The top 3 variables under both rankings were the same, indicating that these methods are performing well. This method appears to work with the Western Blot data but should work well on any form of data, making this a general purpose neural network technique that can also be applied to patient history data.

The above results indicate more data would improve the level of performance. The sensitivity analysis shows little variation in the relative values of variables. Most of the variables contribute to the solution. This should be expected since the bins were chosen based on an analysis of neural network weights trained on the full Western Blot images. By using all or most of the variables, however, the neural networks quickly get into an overtraining situation. This can be avoided by adding data to the training example.

Tests with the neural network guided selection of variables proved to be less effective than the ranking approach. While the ranking approach was clearly the most effective, the neural network guided approach should also be able to eventually discover the best set of variables. Because it is a more direct approach than a genetic algorithm, it is likely to perform better than the genetic algorithm on similar data. The major drawback of this method is that it does not use the sensitivity analysis information to aid in the search.

EXAMPLE 6

Combine Patient History and ELISA Data

Requirements

Using the processing developed in the above examples, train a set of networks on the combination of Patient History Data and ELISA Data. An index generated from an ELISA test, based on the use of the entire set of antigens will be used to determine the improvement in performance achieved by combining this information with the patient historical data.

Additional Requirements

In addition to the above requirements, a comparison between the data from several ELISAs, ELISA 100 and ELISA 200 data and ELISA 2 data, and an analysis of the interrelationships of variables were performed to help determine to what variables the original ELISA tests related.

Methodology Used

In order to determine the improvement in the diagnostic test performance achieved by the inclusion of ELISA test results, several trainings were made using the holdout methods described in EXAMPLE 2. The partitions of the data were made so that in each partition, 80% of the data was used for training and the remaining 20% was used for testing.

To minimize the effects of the random starting weights, several networks were trained in the full training runs. In these runs three networks were trained in each of the five partitions of data, each from a different random start. The outputs of the networks were averaged to form a consensus result that has a lower variance than could be obtained from a single network. Since the number of patients for which all forms of ELISA data was available was 325, new training runs with the original 14 variables were made to provide an accurate means of comparing the effects of the ELISA data on the diagnosis of the disease. Analysis of the ELISA 2 data showed a large range of values for the test. Plots showing the relation of ELISA 2 to the ELISA 100 data suggested that the log of the ELISA 2 data might be better than the raw value.

The comparison training runs were organized as follows:

Run 1:ELISA 100, ELISA 200, log (ELISA 2) and the original 14 variables.

Run 2:(ELISA 2) and the original 14 variables

Run 3: The original 14 variables

After making these comparison runs, a final set of networks was trained on the complete data set of 325 patients. In the final set of networks, a consensus of eight networks was made to produce the final statistics. The final run statistics are reported only on the training data and represent an upper bound on the true performance. The results from the last holdout run represent a possible lower bound on the performance.

From the training data each of the 65 variables, including not available for a diagnosis, were built into a training example of 325 training examples. The TrainDos training program was modified to automate the generation of networks to provide relationships between the variables. In each of 65 networks, one of the variables was predicted by remaining 64. A sensitivity analysis was performed for each network to indicate the importance of each variable in making the prediction.

Results

The consensus results for the three comparison runs are as follows:

Run 1:All ELISA variables (CRFE:1)	66.46%
Run 2:Log of ELISA 2 (CRFEL2)	66.77%
Run 3:No ELISA Variables (CRFEL0)	62.76%

Comparison of Run 1 and Run 2 show that the addition of ELISA 100 and ELISA 200 data to ELISA 2 data had no effect. Therefore, ELISA 100 and ELISA 200 variables could be eliminated.

Comparisons of Run 2 and 3 showed that an input based on the ELISA test improved the diagnosis of the disease.

Comparison of Run 3 to pat06 showed a 5.47% drop in test performance. This could only be due to the decrease in the number of patients available on training. This also suggest that an increase in training data above 500 is likely to have a significant effect on the performance of the neural network on test data.

Based on these results, final networks were trained. Eight networks were trained on the 325 patients. The performance on the training data was 72.31 %. While this is similar results to the pat07 runs, it is clear that the improvement due to ELISA 2 data are being offset by a decrease in the amount of available training data.

Sensitivity Analysis results showed that the ELISA 2 variable ranked 7 out of the 15 variables used.

Plots of the hidden processing element outputs were made from the log files of the eight trained networks. A means was found so that the desired output could be indicated on the plots. By comparing the eight networks, it is clear that each performed the task in a different way. Some clustering of data points is seen in a few plots. Since this did not occur consistently, no conclusions could be drawn.

Statistics were generated on the last training run based on the use of a cutoff in the network output values. If the network output was below the cutoff, the example was rejected from consideration. The following table is a summary of the results for the consensus of eight networks in CRFLE2.

Cutoff	0.00	0.02	0.05	0.075	0.10
Sensitivity	.7790	.7911	.9016	.9596	.9595
Specificity	.6549	.6552	.7200	.7755	.8529
PPV	.7421	.7576	.8397	.8962	.9342
NPV	.6992	.6972	.8181	.9048	.9063
Odds Ratio	2.63	2.75	4.96	8.86	12.5
% Rejected	0.62	15.69	39.38	54.46	66.76

In general, these results are better than the results for pat07.

A test program named adzcrf2.exe (see, Appendix II) was produced as a demo of this final training. This program permits the running of pat07 and CRFEL2 based on the value input in the ELISA field. A value of 0 in the field causes pat07 to be used.

The analysis of variable relationships was performed. Based on the analysis of the relationships, the variables which showed Endo Present as a contributing factor were compared to the variables used in predicting Endo. Results of training two networks (PATVARSA and PATVARS3) showed that in the case of Endo, relationships were not symmetric, as

they are when using correlation. CRFVARSA.XLS was built from the sensitivity analysis results to summarize the results. These results show the nonlinear nature of the relationships. The importance of a variable is affected by the other variables in the training run. This suggests that a means of eliminating unimportant variables in an automated fashion may be required to increase the usefulness of this analysis.

Analysis of the variables relationship (CRFVAR00 - CRFVAR64) showed that in most cases, the log of the ELISA 2 test had greater significance than the raw ELISA 2 value. In particular, the log value ranked higher for both predicting Endo Present and AFS Stage.

Conclusions

The ELISA 2 test adds to the predictive power of the neural network. The ELISA 2 test has eliminated the need for the original ELISA tests. Based on this result it is likely that results of the work with the Western Blot data will further improve the power of the neural network diagnostic test.

The effects of increased training data could be clearly seen in the comparison of Run 3 to pat06. The difference in performance suggests that the performance of the neural network will increase substantially with an increase in training data. It appears from the comparison that doubling the data could improve the performance by 10-15%. With 8-10 times the data the performance might increase to 75-80%.

EXAMPLE 7

Patient History Stage/AFS Score Training

Requirements

Using methods developed in the above Examples, identify relevant variables for either the stage of the disease or the AFS Score. The selection of the target output variable to be used will be determined by a comparison of test set performance from a training run using the phase 1

list of important patient history variables. Once the list of important variables are selected, a consensus of 8 neural networks will be trained on the 510 patient database.

Methodology Used

Training examples were built for the Stage desired output and the AFS score desired output. There were 7 patients missing Stage information and 28 patients missing Score information. For the stage variable, the average value of 2.09 was used where the data was missing. For score, the missing data was replaced with a value depending on the value of the stage variable. For stage 1, a score of 3 was used. From stage 2, 10.5 was used. For stage 3, 28 was used and for stage 4 the value 55 was used. Stage and score were reprocessed so that the desired output would fall in the range of 0.0 to 1.0. Stage was translated linearly. Two methods were used for score. The first was the square root of the score divided by 12.5. The second was the log of score + 1 divided by the log of 150.

The holdout method was used to train networks on stage, square root score and log of score. These networks were trained using 45 variables. The results were compared to determine which variable and processing would be used for the remainder of the Example. The log of score was chosen.

At this point the procedure for isolating the set of important variables was begun. Eight networks were trained on the full training example and the consensus sensitivity analysis was generated to produce the first ranking of the variables. Then the Chi Square contingency table was generated to produce the second ranking of the variables. The procedure for isolating important variables was started manually, but was found to be too time consuming. The procedure was implemented as a computer program and was run on a computer for about one week.

From the results of the variable selection, a set of eight networks were trained on the full training example. The consensus results were analyzed and compared to the Endo present results.

Results

The sensitivity analysis of all 45 variables gave the following ranking of variables:

Name	Input #
Past history of Endo	33
Hist of Pelvic Surgery	34
Dysmenorrhea	40
#Birth	11
age(preproc)	1
Medication History	35
#Preg	10
Ectopic Pregnancy	26
Pelvic Pain	37
Adnnexal msses/Thickening	43
#Abort	12
Current Exod. Hormones	36
Hormone induced	18
Other history	32
Infertility	42
Packs/Day	8
Uterine/Tubal Anomalies	24
Dyspaarunia	41
Anovulatory	15
Hist of Infert	13

Name	Input #
Unknown	16
Menstrual Abnormalities	39
Gyn Cancer	31
Abnormal PAP/Dysplasia	30
Other Symptoms	45
Herpes	19
Ovulatory	14
Genital Warts	20
Fibroids	25
Ovarian Cysts	28
Drug Use	9
Dysfunctional Uterine Bld.	27
PID	23
Hypertension	4
Vag Infections	22
Undetermined	44
Other STD	21
Preg HTM	5
Autoimmune Disease	6
Abdominal Pain	38
Preg Induced DM	3
Oligoovulatory	17
Diabetes	2
Polycystic Ovarian Snyd	29
Transplant	7

The Chi Square analysis gave the following ranking of variables:

Name	Input #
Past history of Endo	33
#Preg	10
#Birth	11
Packs/Day	8
Dysmenorrhea	40
Pelvic Pain	37
Preg HTM	5
Hist of Infert	13
Abnormal PAP/Dysplasia	30
Infertility	42
Diabetes	2
Herpes	19
Hist Of Pelvic Surgery	34
#Abort	12
Other Symptoms	45
Medication History	35
Undetermined	44
Dysfunctional Uterine Bld.	27
Gyn Cancer	31
Uterine/Tubal Anomalies	24
Polycystic Ovarian Synd	29
Dyspaarunia	41
Genital Warts	20
Adnnexal masses/Thickening	43
Oligoovulatory	17

Name	Input #
Autoimmune Disease	6
Abdominal Pain	38
Unknown	16
Ectopic Pregnancy	26
Ovulatory	14
Fibroids	25
Current Exod Hormones	36
Ovarian Cyst	28
Drug Use	9
Vag Infections	22
Preg Induced DM	3
PID	23
age(preproc)	1
Hormone induced	18
Anovulatory	15
Menstrual Abnormalities	39
Hypertension	4
Other STD	21
Transplant	7
Other history	32

The variables chosen in the variable selection procedure were as follows, showing the ranking from the final sensitivity analysis:

Name	Input #
Past history of Endo	33
#Preg	10
Hist of Pelvic Surgery	34
age(proproc)	1
Dysmenorrhea	40
Ectopic Preg.	26
Oligoovulatory	17

The comparison of the score network to the Endo present network can be performed by forcing a threshold on the desired score output to produce an Endo present comparison. The results for the score and the pat07 networks are shown below.

NETWORK	PAT07	SCR07
Sensitivity	.828179	.679525
Specificity	.598174	.647399
PPV	.732523	.789655
NPV	.723757	.509091
Odds Ratio	2.695	2.017751

Conclusions

The set of variables identified in this Example appear to be reasonable.

The automated variable selection methodology appears to function properly. The choice of variables is well predicted by the sensitivity analysis.

Now that there are two methods for predicting the disease, the Endo present network and the Score network could be combined to improve the reliability of the prediction.

EXAMPLE 8

Patient History Adhesions Training

Requirements

Using methods as outlined in EXAMPLE 7, identify relevant variables for Adhesions target output variable. This target output variable will be run using the phase 1 list of important patient history variables. This will also permit a comparison of the new outputs to the Endo present target variable used in phase 1. Once the list of important variables are selected, a consensus of 8 neural networks will be trained on the 510 patient database.

Methodology Used

Training data for the adhesions variable was generated in the same manner as for EXAMPLE 7. The adhesions variable generated two output variables in a manner similar to that used for Endo present. At this point the procedure for isolating the set of important variables was begun. Eight networks were trained on the full training example and the consensus sensitivity analysis was generated to produce the first ranking of the variables. Then the Chi Square contingency table was generated to produce the second ranking of the variables. The procedure for isolating important variables was started manually, but was found to be too time consuming. The procedure was implemented as a computer program and was run on a computer for about one week before completing.

From the results of the variable selection, a set of eight networks were trained on the full training example. The consensus results were analyzed and compared to the Endo present results.

Results

The sensitivity analysis of all 45 variables gave the following ranking of variables:

Name	Input #
Hist of Infert	13
Medication History	35
Ectopic Pregnancy	26
Packs/Day	8
Hist of Pelvic Surgery	34
Infertility	42
#Birth	11
Dyspaarunia	41
Hormone Induced	18
Past history of Endo	33
Herpes	19
#Preg	10
Current Exod. Hormones	36
age(preproc)	1
Dysmenorrhea	40
Uterine/Tubal Anomalies	24
Anovulatory	15
Other history	32
Ovarian Cyst	28
Pelvic Pain	37
Gyn Cancer	31
Ovulatory	14
Menstrual Abnormalities	39
#Abort	12
Unknown	16
Abnormal PAP/Dysplasia	30

Name	Input #
Abdominal Pain	38
Adnnexal masses/Thickening	43
Fibroids	25
Pelivic inflammatory disease (PID)	23
Dysfunctional Uterine Bld	27
Vag Infections	22
Drug Use	9
Other Symptoms	45
Genital Warts	20
Autoimmune Disease	6
Hypertension	4
Other STD	21
Preg Induced DM	3
Preg HTM	5
Polycystic Ovarian Synd	29
Oligoovulatory	17
Diabetes	2
Undetermined	44
Transplant	7

The Chi Square analysis gave the following ranking of variables:

Name	Input #
Hist of Infert	13
Infertility	42
Medication History	35

Name	Input #
Herpes	19
Ovulatory	14
Dysmenorrhea	40
Dyspaarunia	41
Packs/Day	8
Ectopic Preg.	26
Current Exod. Hormones	36
Menstrual Abnormalities	39
Anovulatory	15
Past history of Endo	33
Fibroids	25
Hormone induced	18
#Preg	10
Gyn Cancer	31
Hist of Pelvic Surgery	34
PID	23
Uterine/Tubal Anomalies	24
#Abort	12
#Birth	11
Other STD	21
Abdominal Pain	38
Unknown	16
Vag Infections	22
Abnormal PAP/Dysplasia	30
Dysfunctional Uterine Bld	27
Oligoovulatory	17

Name	Input #
Polycystic Ovarian Synd	29
Autoimmune Disease	6
Genital Warts	20
Other Symptoms	45
Ovarian Cyst	28
Other history	32
Pelvic Pain	37
age(preproc)	1
Preg Induced DM	3
Preg HTM	5
Adnnexal masses/Thickening	43
Undetermined	44
Diabetes	2
Drug Use	9
Hypertension	4
Transplant	7

The variables chosen in the variable selection procedure were as follows, showing the ranking from the final sensitivity analysis:

Name	Input #
Hist of Infert	13
Dyspaarunia	41
Ectopic Preg	26
Packs/Day	8
Hist of Pelvic Surgery	34
Medication History	35

Name	Input #
Ovulatory	14
Menstrual Abnormalities	39
Oligoovulatory	17

The comparison of the Score network to the Endo present network can be performed by forcing a threshold on the desired score output to produce an Endo present comparison. The results for the score and the pat07 networks are shown below.

NETWORK	PAT07	ADH07
Sensitivity	.828179	.825083
Specificity	.598174	.473430
PPV	.732523	.696379
NPV	.723757	.649007
Odds Ratio	2.695	2.148148

Conclusions

The set of variables identified in this Example appear to be reasonable. The automated variable selection methodology appears to function properly. The choice of variables is well predicted by the sensitivity analysis.

EXAMPLE 9

This example shows the reprodicibility of the process provided herein.

Methodology Used

Software used for the selection of important variables for Adhesions and Score was modified to operate with the Endo present desired output. The software was further modified to allow it to be run in

the general case instead of needing to be recompiled for each specific test.

The run was made on the Endo Present variable in the same fashion as the runs for Adhesion and score. This included using a consensus of 4 networks during the variable selection process. The training data was partitioned into five partitions during the training process, generating a total of 20 networks for each evaluation of the current set of variables being tested.

The results of runs with different random number seeds indicated that the number of networks in the consensus might need to be increased.

Two additional variable selection runs were made with a consensus of 10 networks used during the process. In this case, a total of 50 networks were trained to evaluate a single combination of variables. Two separate runs were made in this fashion with only the random starting seed changing.

From these final two variable selection runs, a set of eight networks were trained for each of the sets of variables (pat08, pat09), to allow their performance to be evaluated on new data (not included in the original 510 record database). Statistics on the performance of these networks were generated so that they could be compared to the original pat07 consensus nets.

Results

In each case where different random number seeds were used, the variable selection process found a different set of important variables. When the number of networks in the consensus was increased to 10, the variables common in the different runs increased.

Many of the original 14 variables used for pat07 were confirmed as being important in the variable selection runs using 10 consensus nets.

The final runs made on the selected variables were named pat08 and pat09.

The variables used in pat08 and pat09 consensus networks are shown below along with their sensitivity analysis rankings:

INPUT	PAT08 Variable Name	Average
10:	Past HX of Endo	1.525819
05:	#Birth	1.489470
14:	Dysmenorrhea	1.302913
12:	Medication History	1.263276
11:	Hist of Pelvic Surgery	1.235421
13:	Pelvic Pain	1.192193
01:	Age(preproc)	1.178076
04:	Packs/Day	1.063709
06:	Hist of Infert	0.943830
07:	Hormone Induced	0.925489
16:	Other Symptoms	0.879384
15:	Infertility	0.873955
09:	Abnormal PAP/Dysplasia	0.7061778
03:	Preg HTN	0.576799
08:	Herpes	0.441410
02:	Diabetes	0.402080
06:	Past HX of Endo	1.420401
04:	#Birth	1.353489
01:	age(preproc)	1.187359
09:	Pelvic Pain	1.183929
10:	Dysmenorrhea	1.141531
07:	Hist of Pelvic Surgery	1.064250

INPUT	PAT08 Variable Name	Average
03:	Packs/Day	1.042488
11:	Other Symptoms	0.780530
05:	Herpes	0.699654
08:	Medication History	0.623505
02:	Preg HTN	0.502863

Conclusions

The variable selection process appears works well and has produced two alternative networks that work as well or better than the the pat07 nets. The reason for this conclusion is that the performance statistics generated only on the training data appear slightly better for the pat07 then pat08 and pat09. Since the variable selection process carefully picks variables based on test set performance, the associated networks are not likely to have been overtrained. As a network becomes overtrained the typical characteristic is that the training example performance increases and the test set performance decreases. Thus the higher performance of pat07 may be the result of slight overtraining.

While the variable selection process appears to have produced two alternative selections on the same training data, the performance of the two selections appears to be very similar. This is based on the test set performance of the final variable selections for the two runs. It has become clear that when two variables are closer in their relative performance, then random factors can influence their relative ranking. The random factors in the variable selection runs included the random starting points and the use of added noise on the inputs during training. The random noise has been shown to aid in producing better generalization (translation: test set performance). As the number of

networks in the consensus increases, the effects of the random influences are decreased.

The determination of a set of variables that produces a high quality network seems to be addressed by the variable selection process. As more combinations of variables that work successfully are enumerated, it is evident that certain variables or combinations of variables are essential to good performance.

EXAMPLE 10

Evaluation of the Elimination of Past History of Endometriosis and History of Pelvic Surgery on Diagnostic Performance.

The purpose of this Example was to determine the importance of 'Past history of endometriosis' and 'Past history of pelvic surgery' variables in evaluating a patient's risk of having endometriosis, and to provide an alternative means (different from sensitivity analysis) to measure the importance of any given variable in predicting the outcome.

Tasks:

1. Apply the variable selection process excluding the 'Past history of endometriosis.
2. Repeat task (1) using different random seed variables for the variable selection process.
3. For both sets of 'endometriosis relevant variables' identified in tasks (1) and (2) above, complete the consensus network training process.
4. Repeat the above tasks (1), (2) and (3) excluding the 'Past history of Pelvic surgery' variable from the Endometriosis database.
5. Repeat the above tasks (1), (2), and (3) excluding both the 'Past history of endometriosis' and the 'Past history of endometriosis' variables from the Endometriosis database.

Methodology Used

The variable selection software developed in Example 9 was used as the basis to generate results for each of Example 10. The software was modified so that the user could identify variables that would be excluded from consideration based on the requirements of Example 10. This software was also modified to allow the reporting of classification performance for each of the sets of variables tested so that the effect of an eliminated variable could be more easily understood.

For each variable selection run that was made, parameters for the variable selection process were set as follows:

Number of partitions:	5
Consensus networks:	10
Training example size:	510
Number of passes:	999

The ordering of database variables in the variable selection process was based on the sensitivity analysis and Chi square analysis. This ordering was the same as used in pat08 and pat09.

The networks trained for this Example are identified as follows (the two nets have different random seeds);

Past Hist. of Endo eliminated:	pat10, pat11.
Hist. of Pelvic Surgery eliminated:	pat12, pat13.
Both variables eliminated:	pat14, pat15.

Once the variable selection process was completed for each of the combinations of variables and random seeds, a set of eight networks were trained using the selected variables identified. Each of these networks was trained on the complete 510 record database. From these training runs, a consensus of the outputs was generated in an Excel spreadsheet so that the performance of each of the networks could be evaluated.

Results

The typical performance of a consensus of networks was estimated using the holdout method with a partition of 5. When all variables were available, as in pat08 and pat09, the classification performance was estimated to be 65.23%.

When the past history of endometriosis variable was eliminated from consideration (pat10 and pat11), the performance was estimated at 62.47%. This represents a drop of 2.76%.

When the history of pelvic surgery variable was eliminated from consideration (pat12 and pat13), the performance was estimated at 64.52%. This represents a drop of only 0.72%.

When both variables were eliminated from consideration (pat14 and pat15), the performance was estimated to be 62.43%. This represents a drop of 2.80%. This is only slightly worse than just eliminating past history of endometriosis and appears to be consistent with other results based on the assumption that the variables are independent (are not correlated).

Conclusions

While history of pelvic surgery was used by the neural networks when it was available, the effect of eliminating this variable was minimal. The neural networks appear to be able to compensate for the elimination of this variable by using other information.

The removal of Past history of Endometriosis was significant. This variable was always at the top of the list in any sensitivity analysis. Its elimination caused about a 2.76% drop in performance over the average when all variables were available for use. Given that the average performance is estimated at 65.23%, and 50% can be achieved by chance, this represents an effective drop in performance of 18.12%.

When both variables were eliminated, there does not appear to be any significant drop in performance, which indicates there was no interaction between these two variables. This process of eliminating a variable and running the variable selection process appears to be a good approach to determining the true value of a given variable. It should be noted that there are two variables that are important to the diagnosis, but are highly correlated, the elimination of only one would have little effect since the network would compensate by using the other. It is only when both were eliminated that their value becomes clear.

EXAMPLE 11

Evaluation of the Elimination of Pelvic Pain and dysmenorrhea on Diagnostic Performance

Requirements

Goals:

1. To determine the importance of 'Pelvic Pain' and "Dysmenorrhea' variables in evaluating a patient's risk of having endometriosis.
2. To provide a separate mechanism (different from sensitivity analysis) to measure the importance of any given variable in predicting the outcome.

Tasks:

1. Apply the variable selection process described herein.
2. Repeat task (1) using different random seed variables for the variable selection process.
3. For both sets of 'endometriosis relevant variables' identified in tasks (1) and (2) above, and complete the consensus network training process.
4. Repeat the above tasks (1), (2) and (3) excluding the 'Dysmenorrhea' variable from the Endometriosis database.

5. Repeat the above tasks (1), (2), and (3) excluding both the 'Pelvic Pain' and the 'Dysmenorrhea' variables from the Endometriosis database.

Methodology Used

The variable selection software developed in Example 9 was used as the basis to generate results for each of these tasks.

For each variable selection run that was made, parameters for the variable selection process were sat as follows;

Number of partitions:	5
Consensus networks:	10
Training example size:	510
Number of passes:	999

The ordering of database variables in the variable selection process was based on the sensitivity analysis and Chi square analysis. This ordering was the same as used in pat08 and pat09. The networks trained for this task are identified as follows (the two nets have different random seeds);

Pelvic Pain eliminated:	pat16, pat17, pat17A
Dysmenorrhea eliminated:	pat18, pat19.
Both variables eliminated:	pat20, pat21.
Four variables (EXs. 11 and 12):	pat22, pat23, and pat23A

Once the variable selection process was completed for each of the combinations of variables and random seeds, a set of eight networks were trained using the selected variables identified. Each of these networks was trained on the complete 510 record database. From these training runs, a consensus of the outputs was generated in an Excel spreadsheet so that the performance of each of the networks could be evaluated.

Results

The typical performance of a consensus of networks was estimated using the holdout method with a partition of 5. When all variables were available, as in pat08 and pat09, the Classification performance was estimated to be 65.23%.

When the pelvic pain variable was eliminated from consideration (pat16 and pat17), the performance was estimated at 61.03%. This represents a drop of 4.20%.

When the dysmenorrhea variable was eliminated from consideration (pat18 and pat19), the performance was estimated at 63.44%. This represents a drop of only 1.79%.

When both variables were eliminated from consideration (pat20 and pat21), the performance was estimated to be 61.22%. This represents a drop of 4.00%. This is better than when pelvic pain was eliminated only. This suggests that the performance drop for pelvic pain is overstated. The best performing network that did not include pelvic pain had a performance of 62.29%, which gives a drop of 2.94%. This would be more reasonable an estimate given the performance when both were eliminated.

Conclusions

With the four variables tested, the ranking of the variables in order of importance is as follows:

Pelvic pain	2.94% - 4.20% drop
Past hist. of endo	2.76% drop
Dysmenorrhea	1.79% drop
Hist. of pelvic surg	0.72% drop

This process of eliminating a variable and running the variable selection process is a good approach to determining the value of a given variable. It should be noted that if there were two variables that were

important to the diagnosis, but were highly correlated, the elimination of only one would have little effect since the network would compensate by using the other. It is only when both were eliminated that their true value would become clear.

EXAMPLE 12

Training of Neural Network to Differentiate Mild Versus Severe Endometriosis

Goals:

1. To train a consensus of networks which differentiates between minimal/mild versus moderate/severe endometriosis.

Task:

1. Train networks to AFS score as follows:

   positive = Endo Stage III or IV

   negative = No Endo, Endo Stage I or II

2. Apply the variable selection process described in Method for Developing Medical and Biochemical Tests Using Neural Networks of the Endometriosis database.
3. Repeat task (2) using different random seed variables for the variable selection process.
4. Compare variables selected in (2) and (3) above before proceeding. If selected set of variables differ widely, repeat task (2) using different random seed weights.
5. Train final consensus networks for variables selected in (2) and (3) above.
6. Repeat steps (2) through (5) using only the subset of the endometriosis database for which Endo was present in the patient.

Methodology Used

The variable selection software developed in EXAMPLE 10 and modified in EXAMPLE 11, was used as the basis to generate results for each of the tasks in this example.

For each variable selection run that was made, parameters for the variable selection process were set as follows;

Number of partitions:	5
Consensus networks:	20
Training example size:	510 (290 for step (6))
Number of passes:	999

The ordering of database variables in the variable selection process was based on the sensitivity analysis an chi square analysis run specifically for the new target output as described in Example 1. The networks trained for this Example are identified as follows (the two nets have different random seeds);

Nets trained on full database: AFS01 and AFS02

Nets trained on Endo present subset: AFSEP1 and AFSEP2.

Once the variable selection process was completed for each of the combinations of variables and random seeds, a set of eight networks were trained using the selected variables identified. Each of these networks for AFS01 and AFS02 variables were trained on the complete 510 record database. Each of the networks for AFSEP1 and AFSEP2 variables were trained on the 291 records for which the endo present variable was positive. From these training runs, a consensus of the outputs was generated in an Excel spreadsheet so that the performance of each of the networks could be evaluated.

Results

The count of variables found in the reduced subset run was smaller than for the runs on the full training example. The typical performance of a consensus of networks was estimated using the holdout method with a partition of 5. The typical classification performance for the AFS run using the full training example was 77.22549%. The typical classification performance on the endo present subset was 63.008621%. If all examples were classified as negative, the performance for the full training example would be 78.82% and 65.29% for the subset. By changing the cutoff values for positive and negative classification better performance than suggested by these numbers can be achieved.

Conclusions

The results of the variable selection runs for the full training example and the subset of endo present examples suggests that the size of the training example is of importance in the determination of the important variables. It is clear that as the size of the training example increases, more variables will be considered important. This result can also be interpreted as an indication that more training data will improve the variable selection process and also the overall performance of the consensus networks used in building the diagnostic test.

EXAMPLE 13

Variable Selection and development of neural nets for predicting pregnancy related events and improvement of the performance of tests for fetal fibronectin

Data was collected from the over 700 test patients involved in a clinical trial of the assay described in U.S. Patent No. 5,468,619. Variable selection was performed without fetal fibronectin (fFN) test data. The final networks, designate EGA1-EGA4 were trained with the variables set forth in the table below.

EGA1 - EGA4 represent neural networks used for variable selection. For EGA1, the variable selection protocol was performed a network architecture with 8 inputs in the input layer, three processing elements in the hidden layer, and one output in the output layer. EGA2 is the same as EGA1, except that it is 9 inputs in the input layer. EGA3 has 7 inputs in the input layer, three processing elements in the hidden layer, and one output in the output layer. EGA4 is the same as EGA1, except that it is 8 inputs in the input layer.

The variables selected are as follows:

EGA1	EGA2	EGA3	EGA4
--	fFN	--	fFN
Ethnic Origin 1 (Caucasian)		Ethnic Origin 4 (Hispanic)	
EGA Sonogram		Marital Status 5 (living with partner)	
EGA Best (physician's determination of estimated gestational age)		Marital Status 6 (other)	
EGA Sampling		EGA Best	
Cervical dilation interpretation		Cervical dilation interpretation	
Vaginal bleeding (at time of sampling)		Vaginal bleeding (at time of sampling)	
1 complications (prev. preg w/o complications)		Other complications (prev. preg. w complications)	
Other complications (prev. preg. w complications)			

EGA = estimated gestational age

Final consensus network performance

Net	TP	TN	FP	FN	SN	SP	PPV	NVP	OR
EGA1	35	619	92	17	67.3	87.0	27.6	97.3	6.0
EGA2	37	640	71	15	71.2	90.0	34.3	97.7	7.9
EGA3	36	602	109	16	69.2	84.7	24.8	97.4	5.1
EGA4	32	654	57	20	61.5	92.0	36.0	87.0	8.9
fFN	31	592	119	21	59.6	83.3	20.7	96.6	7.3

EGA = estimated gestational age (less than 34 weeks); TP = true positives; TN = true negatives; FP = false positives; FN = false negative; SN = sensitivity; SP = specificity;, PPV = positive predictive value; NPV = negative predictive value; OR = odds ratio (total number correct/total number correct answers); and fFN = the results from the ELISA assay for fFN.

The results show that the network EGA4, the neural net that includes seven patient variables and includes the fFN ELISA assay and that predicts delivery at less than 34 weeks, has far fewer false positives than the fFN ELISA assay. In addition, the number of false positives was reduced by 50%. Incorporation of the fFN test into a neural net improved the performance of the fFN ELISA assay. All of the neural nets performed better than the fFN test alone. Thus, the methods herein, can be used to develop neural nets, as well as other decision-support systems, that can be used to predict pregnancy related events.

EXAMPLE 14

Training of Consensus Neural Networks on Specific subsets of Pat07 Variables

The examples shows the results of a task designed to quantitate the contribution of pat07 variables to pat07 performance, and to develop endometriosis networks using minimal numbers of pat07 variables.

Tasks:

1. Train final consensus networks using the following combination of Pat07 variables:
    a. All 14 minus Hx Endo (13 variables total)
    b. All 14 minus pelvic pain (13 variables total)
    c. All 14 minus dysmenorrhea (13 variables total)
    d. All 14 minus pelvic surgery (13 variables total)
2. Train final consensus networks using other combinations of Pat07 variables.
    a. Hx Endo, pelvic pain, and dysmenorrhea
    b. Hx Endo, pelvic pain, dysmenorrhea and Hx pelvic surgery
3. Train final consensus networks using other combinations of pat07 variables as indicated from above results.

Methodology Used

Using the original patient database, training examples were generated for each of the combinations of variables to be evaluated. These training examples contained only the variables required for the given consensus run. TrainDos™ was used in batch mode to train a set of eight neural networks for each of the combinations of variables to be evaluated. The networks were trained using the same parameters as the Pat07 training runs. The only difference was the setting of the random number seeds for each network. Each network was trained on the full 510 record database. From these training runs, a consensus of the outputs was generated in an Excel spreadsheet so that the performance of each of the networks could be evaluated.

Results

Since these runs were final training runs, the effects of eliminating variables could be seen but did not give as clear an indication as can be achieved by the holdout method.

Conclusions

The results of the variable selection runs on the full training example for the purposes of determining the contribution of a given set of variables is not as good a method as the evaluation method used in the variable selection process. The "holdout" method for evaluation with a partition of 5 and 20 net consensus gives a substantially better statistic for comparison of variables.

EXAMPLE 15

METHOD AND APPARATUS FOR AIDING IN THE DIAGNOSIS OF ENDOMETRIOSIS USING A PLURALITY OF PARAMETERS SUITED FOR ANALYSIS THROUGH A NEURAL NETWORK (PAT07)

Fig. 7 is a schematic diagram of an embodiment of one type of neural network 10 trained on clinical data of the form used for the consensus network (Fig. 10) of a plurality of neural networks. The structure is stored in digital form, along with weight values and data to be processed in a digital computer. This first type neural network 10 contains three layers, an input layer 12, a hidden layer 14 and an output layer 16. The input layer 12 has fourteen input preprocessors 17-30, each of which is provided with a normalizer (not shown) which generates a mean and standard deviation value to weight the clinical factors which are input into the input layer. The mean and standard deviation values are unique to the network training data. The input layer preprocessors 17-30 are each coupled to first and second processing elements 48, 50 of the hidden layer 14 via paths 51-64 and 65-78 so that each hidden layer processing element 48, 50 receives a value or signal from each

input preprocessor 17-30. Each path is provided with a unique weight based on the results of training on training data. The unique weights 80-93 and 95-108 are non-linearly related to the output and are unique for each network structure and initial values of the training data. The final value of the weights are based on the initialized values assigned for network training. The combination of the weights that result from training comprise a functional apparatus whose description as expressed in weights produces a desired solution, or more specifically a preliminary indicator of a diagnosis for endometriosis.

For the endometriosis test provided herein, the factors used to train the neural network and upon which the output is based are the past history of the disease, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts and diabetes. These fourteen factors have been determined to be a set of the most influential (greatest sensitivity) from the original set of over forty clinical factors. (Other sets of influential factors have been derived, see, EXAMPLES, above).

The hidden layer 14 is biased by bias weights 94, 119 provided via paths 164 and 179 to the processing elements 48 and 50. The output layer 16 contains two output processing elements 120, 122. The output layer 16 receives input from both hidden layer processing elements 48, 50 via paths 123, 124 and 125, 126. The output layer processing elements 120, 122 are weighted by weights 110, 112 and 114, 116. The output layer 16 is biased by bias weights 128, 130 provided via paths 129 and 131 to the processing elements 120 and 122.

The preliminary indication of the presence, absence or severity of endometriosis is the output pair of values A and B from the two processing elements 120, 122. The values are always positive between

zero and one. One of the indicators is indicative that endometriosis is present. The other one of the indicators is indicative that endometriosis is absent. While the output pair A, B provide generally valid indication of the disease, a consensus network of trained neural networks provides a higher confidence index.

Referring to Fig. 10, a final indicator pair C, D is based on an analysis of a consensus of preliminary indicator pairs from a plurality, specifically eight, trained neural networks 10A - 10H (Fig. 10). Each preliminary indicator pair A, B is provided to one of two consensus processors 150, 152. via paths 133-140 and 141-148. The first consensus processor 150 processes all positive indicators. The second consensus processor 152 processes all negative indicators. Each consensus processor 150, 152 is an averager, i.e., it merely forms a linear combination, such as an average, of the collection of like preliminary indicator pairs A, B. The resultant confidence indicator pair is the desired result, where the inputs are the set of clinical factors for the patient under test.

Fig. 9 illustrates a typical processor element 120. Similar processors 48 and 50 have more input elements, and processor element 122 is substantially identical. Typical processor element 120 comprises a plurality of weight multipliers 110, 114, 128 on respective input paths (numbering in total herein 15, 16 or 3 per element and shown herein as part of the processor element 120). The weighted values from the weight multipliers are coupled to a summer 156. The summer 156 output is coupled to an activation function 158, such as a sigmoid transfer function or an arctangent transfer function. The processor elements can be implemented as dedicated hardware or in a software function.

A sensitivity analysis can be performed to determine the relative importance of the clinical factors. The sensitivity analysis is performed on

a digital computer as follows: A trained neural network is run in the forward mode (no training) for each training example (input data group for which true output is known or suspected). The output of the network for each training example is then recorded. Thereafter, the network is rerun with each input variable being replaced by the average value of that input variable over the entire training example. The difference in values for each output is then squared and summed (accumulated) to obtain individual sums.

This sensitivity analysis process is performed for each training example. Each of the resultant sums is then normalized according to conventional processes so that if all variables contributed equally to the single resultant output, the normalized value would be 1.0. From this information, the normalized value can be ranked in order of importance.

In analysis of clinical data, it was determined that the order of sensitivity of factors for this neural network system are the past history of the disease, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts and diabetes.

A specific neural network system has been trained and has been found to be an effective diagnostic tool. The neural network system, as illustrated by Figs. 7 and 10, is described as follows:

Weights, in order of identification and not in order of sensitivity:

0. Bias
1. Age
2. Diabetes
3. Pregnancy hypertension
4. Smoking Packs/Day
5. #Pregnancies

6. #Births
7. #Abortions
8. Genital Warts
9. Abnormal PAP/Dysplasia
10. History of Endometriosis
11. History of Pelvic Surgery
12. Medication History
13. Pelvic Pain
14. Dysmenorrhea

are as follows for each of eight of the first type of neural networks 10:

First neural network A:

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	0.15	0.77	-0.12	0.12
1	-1.19	2.25	-0.44	0.44
2	-0.76	-2.30	0.66	-0.65
3	3.01	-1.48		
4	1.81	-0.85		
5	1.87	0.27		
6	3.56	-1.70		
7	-0.48	-0.47		
8	1.33	0.84		
9	-1.96	-6.19		
10	-4.45	0.50		
11	1.36	-0.95		
12	-1.61	0.40		
13	-1.97	2.38		
14	-0.91	1.86		

First Neural Network B:

Node/Weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	-0.16	-1.62	0.70	-0.70
1	-3.30	0.79	-0.69	0.69
2	0.85	0.45	-0.65	0.65
3	1.00	2.14		
4	1.00	3.82		
5	-0.81	3.93		
6	1.57	3.96		
7	-1.40	2.27		
8	0.46	-0.54		
9	1.16	1.51		
10	-0.80	-4.76		
11	-0.01	2.83		
12	-1.19	0.74		
13	-1.10	-0.43		
14	-2.29	-0.17		

First Neural Network C

Node/weight/	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	0.94	0.77	0.10	-0.10
1	1.43	3.31	-0.90	0.90
2	0.30	-1.48	0.87	-0.87
3	1.17	-0.83		
4	2.11	0.60		
5	-1.16	-2.09		
6	1.033	-1.39		
7	-0.68	-0.40		

Node/weight/	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
8	-0.88	-0.19		
9	0.31	-0.89		
10	-1.74	1.36		
11	1.62	0.59		
12	-1.49	-1.11		
13	-1.05	0.26		
14	-0.41	1.036		

First Neural Network D

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	1.08	-0.03	-1.43	1.30
1	1.27	-0.58	1.39	1.28
2	-0.89	-0.46	1.28	1.17
3	-1.00	-0.94		
4	-1.74	0.73		
5	-0.40	0.10		
6	-1.38	0.55		
7	1.26	-0.79		
8	1.06	-0.10		
9	0.66	-1.36		
10	0.71	1.01		
11	-0.57	0.00		
12	0.67	-0.38		
13	1.89	-0.49		
14	-0.90	1.57		

First Neural Network E

Node/Weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	0.14	-3.93	0.46	-0.46
1	-2.12	-1.07	-0.52	0.51
2	8.36	1.16	-0.80	0.82
3	1.02	1.39		
4	1.79	1.01		
5	0.31	-1.08		
6	2.87	2.33		
7	0.84	0.76		
8	-1.24	-0.51		
9	-1.75	-0.31		
10	-2.98	-1.92		
11	1.72	0.59		
12	-1.22	0.06		
13	-2.47	-0.76		
14	-1.14	-1.44		

First Neural Network F

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	-1.19	0.82	0.68	-0.68
1	-2.93	0.19	-0.67	0.67
2	1.19	0.72	-0.58	0.59
3	6.85	0.83		
4	1.08	0.59		
5	0.66	0.07		
6	1.65	1.06		
7	-0.28	0.51		

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
8	-1.63	1.04		
9	-1.15	1.47		
10	-0.80	-1.97		
11	0.43	0.97		
12	-0.13	-0.91		
13	-3.10	0.15		
14	-2.27	0.09		

First Neural Network G

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	-1.18	1.08	0.69	-0.69
1	-2.55	1.11	-0.70	0.70
2	0.48	0.52	-0.50	0.50
3	-1.40	1.41		
4	1.11	0.55		
5	-0.28	-0.48		
6	2.33	-0.23		
7	0.33	0.44		
8	-1.92	-1.23		
9	0.99	0.77		
10	-1.41	-2.96		
11	0.68	1.40		
12	-0.28	-0.28		
13	-1.66	-0.64		
14	-0.79	-2.38		

First Neural Network H

Node\weight	1st Hidden Layer	2nd Hidden Layer	1st Output Layer	2nd Output Layer
0	15.74	-2.48	0.017	-0.75
1	-0.76	-2.49	0.41	0.34
2	-0.91	0.99	-0.84	0.85
3	-1.13	1.97		
4	-0.75	2.41		
5	-0.66	1.51		
6	-0.83	1.01		
7	1.03	-0.26		
8	0.75	-0.76		
9	-0.48	2.00		
10	-0.48	-5.03		
11	2.01	1.77		
12	-0.015	-0.77		
13	0.25	-2.29		
14	1.11	-2.01		

Normalized Observation Value For First Type Neural Networks

Mean	Standard Deviation
-0.000000	1.000000
0.01	0.08
0.01	0.09
0.16	0.37
1.09	1.39
0.55	0.94
0.54	0.93

0.01	0.10
0.03	0.17
0.23	0.42
0.65	0.48
0.39	0.49
0.19	0.39
0.72	0.45

Further, as provided herein, the results of biochemical tests, such as tests according to the ELISA format test, may be used to produce trained augmented neural network systems to produce a relatively higher confidence level in terms of sensitivity and specificity. These second type neural networks are illustrated in Fig. 8. The numbering is identical to Fig. 7, except for the addition of a node 31 in the input layer 12 and a pair of weights 109 and 111. All weights, however, in the network change upon training with the additional biochemical result. The exact weight set is dependent on the specific biochemical test training example.

The training system provided herein may be used. Alternative training techniques may also be used (see, e.g., Baxt, "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction," Annals of Internal Medicine 115, p.843 (1 December 1991); "Improving the Accuracy of an Artificial Neural Network Using Multiple Differently Trained Networks," Neural Computation 4, p. 772 (1992)).

In evaluating the test results, it was noted that a high score correlated with presence of disease, a low score correlated with absence of the disease, and extreme scores raised confidence, while midrange scores reduced confidence. The presence of endometriosis was indicated by an output of 0.6 or above, and its absence by 0.4 or less. It was also noted that higher relative scores correlated with higher relative severity of

the disease. The methods herein, minimize the number of patients that require further, often surgical, procedures to establish the presence, absence or severity of the disease condition.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

WHAT IS CLAIMED:

1. A method for variable selection, comprising:

(a) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(b) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(c) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (b), wherein the best candidate variable does not improve performance, the process is completed.

2. The method of claim 1, wherein to step (a)the candidate variables are obtained from patients and include historical data and/or biochemical data.

3. A method of diagnosis, comprising:

selecting a set of important selected variables according to the method of claim 1; and

training a decision-support system using the selected final set of important selected variables to produce a test for diagnosis.

4. The method of claim 3, wherein the method of diagnosis assesses the likelihood that a medical condition or disorder is present, assesses the likelihood that a particular condition will develop or occur in the future, selects a course of treatment or determines the effectiveness of a treatment.

5. The method of claim 4, wherein the condition is a pregnancy-related condition or endometriosis.

6. The method of claim 3, wherein the method of diagnosis assesses the presence, absence or severity of a medical condition or determines the likely outcome resulting of a course of treatment.

7. A method of improving the effectiveness of a diagnostic biochemical test, comprising:

selecting a set of important selected variables according to the method of claim 1; and

training a decision-support system using the selected final set of important selected variables and the biochemical test data to produce a test that is more effective than the biochemical test alone.

8. A method of identifying a biochemical test that aids in diagnosis of a disorder or condition, comprising:

(a) selecting a set of important selected variables according to the method of claim 1;

(b) identifying a set of biochemical test data, and

training a decision-support system using the selected final set of important selected variables combined with each member of the set of biochemical test data, and assessing the performance of the resulting system;

(c) repeating the training and assessing with each member of the set of biochemical test data until all have been used in a training; and

(d) selecting the member of the set of biochemical data that results in a system that performs the best.

9. A method for variable selection, comprising:

(a) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(b) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(c) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(d) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (c), if the variable does not improve performance in comparison to the performance of the selected important variables, evaluating is continued at step (e);

(e) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (d).

10. The method of claim 9, wherein the candidate variables include biochemical test data.

11. The method of claim 9, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

12. The method of claim 9, wherein ranking is based on process comprising a statistical analysis.

13. The method of claim 9, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

14. The method of claim 9, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

15. The method of claim 10, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

16. The method of claim 1, wherein the decision-support system includes a consensus of neural networks.

17. The method of claim 9, wherein the decision-support system includes a consensus of neural networks.

18. The method of claim 1 that is a computer-assisted method, wherein the set of n candidate variables and set of selected important variables are each stored in a computer.

19. The method of claim 9 that is a computer-assisted method, wherein the set of n candidate variables and set of selected important variables are each stored in a computer.

20. The method of claim 9, further comprising training a final decision-support system based on the completed set of selected important variables to produce a decision-support system based test for the condition.

21. The method of claim 3, further comprising training a final decision-support system based on the completed set of selected important variables to produce a decision-support system based test for the condition.

22. The method of claim 3, wherein the condition is a gynecological condition.

23. The method of claim 22, wherein the condition is selected from among infertility, a pregnancy related event, and pre-eclampsia.

24. In a computer system, a method for developing a decision-support system-based test to aid in diagnosing a medical condition, disease or disorder in a patient, comprising:

(a) collecting observations from a group of test patients in whom the medical condition is known;

(b) categorizing the observations into a set of candidate variables having observation values and storing the observation values as a observation data set in a computer;

(c) selecting a subset of selected important variables from the set of candidate variables by classifying the observation data set using a first decision-support system programmed into the computer system, whereby

the subset of selected important variables includes the candidate variables substantially indicative of the medical condition; and

(d) training a second decision-support system using the observation data corresponding to a subset of selected important variables, whereby the second decision-support system-based system constitutes a decision-support based diagnostic test for the condition, disease or disorder.

25. The method of claim 24, wherein the first decision-support system includes at least one neural network.

26. The method of claim 24, wherein the second decision-support system includes least one neural network.

27. The method of claim 24, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iii) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (ii), wherein the best candidate variable does not improve performance, the process is completed.

28. The method of claim 24, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(iii) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iv) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (iii), if the variable does not improve performance in comparison to the performance of the selected important variables, evaluating is continued at step (v);

(v) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (iv).

29. The method of claim 28, wherein the candidate variables include biochemical test data.

30. The method of claim 28, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

31. The method of claim 28, wherein ranking is based on process comprising a statistical analysis.

32. The method of claim 28, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

33. The method of claim 28, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

34. The method of claim 28, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

35. The method of claim 34, wherein the sensitivity analysis comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

36. The method of claim 35, further comprising:

(vi) ranking the variables according to their relative contribution to the determination of the decision-support system output.

37. The method of claim 24, wherein the step of training a second decision-support system includes a validating step wherein a previously-unused set of observation data is run through the second decision-support system after training to provide a performance estimate for indication of the medical condition, wherein the previously-unused set of observation data are collected from patients in whom the medical condition is known.

38. The method of claim 24, wherein the step of training a second decision-support system includes partitioning the observation data set into a plurality of partitions comprising at least one testing data partition and a plurality of training data partitions, wherein the second decision-support system is run using the plurality of training data partitions and the testing data partition is used to provide a final performance estimate for the second decision-support system after the training data partitions have been run.

39. The method of claim 38, wherein the second decision-support system comprises a plurality of neural networks, each neural network of the plurality having a unique set of starting weights and having a performance rating value.

40. The method of claim 39, wherein the final performance estimate is generated by averaging the performance rating values for the plurality of neural networks.

41. The method of claim 24, wherein the observation values are obtained from patient historical data results and/or biochemical test results.

42. The method of claim 24, wherein the condition is a pregnancy-related condition or endometriosis.

43. The method of claim 24, further comprising:

(e) collecting addition observations from patients and categorizing them into a set of candidate variables, which are then added to first set of candidate variables; and then

(f) repeating steps (c) and (d).

44. The method of claim 3, wherein the candidate variables include biochemical test data.

45. The method of claim 24, further comprising, after collecting observations from a group of test patients and before training the second decision-support based system,

collecting test results of a biochemical test from at least a portion of the test patients in whom the condition is known or suspected and categorizing them into a set of candidate variables, which are then added to first set of candidate variables; and then

repeating steps (c) and (d).

46. The method of claim 45, wherein the test assesses the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

47. The method of claim 45, wherein the test assess the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

48. The method of claim 45, wherein the decision-support system includes a neural network and the final set constitutes a consensus of neural networks.

49. The method of claim 45, wherein the first subset of relevant variables is identified using sensitivity analysis performed on the decision-support based system or consensus thereof.

50. The method of claim 45, wherein the first decision-support system includes at least one neural network.

51. The method of claim 45, wherein the second decision-support system includes at least one neural network.

52. The method of claim 45, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iii) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (ii), wherein the best candidate variable does not improve performance, the process is completed.

53. The method of claim 45, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(iii) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iv) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (iii), if the variable does not improve performance in

comparison to the performance of the selected important variables, evaluating is continued at step (v);

(v) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (iv).

54. The method of claim 53, wherein the candidate variables include biochemical test data.

55. The method of claim 53, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

56. The method of claim 53, wherein ranking is based on process comprising a statistical analysis.

57. The method of claim 53, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

58. The method of claim 54, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

59. The method of claim 54, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

60. The method of claim 59, wherein the sensitivity analysis comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

61. The method of claim 60, further comprising:

(vi) ranking the variables according to their relative contribution to the determination of the decision-support system output.

62. The method of claim 45, wherein the step of training a second decision-support system includes a validating step wherein a previously-unused set of observation data is run through the second decision-support system after training to provide a performance estimate for indication of the medical condition, wherein the previously-unused set of observation data are collected from patients in whom the medical condition is known.

63. The method of claim 45, wherein the step of training a second decision-support system includes partitioning the observation data set into a plurality of partitions comprising at least one testing data partition and a plurality of training data partitions, wherein the second decision-support system is run using the plurality of training data partitions and the testing data partition is used to provide a final performance estimate for the second decision-support system after the training data partitions have been run.

64. The method of claim 63, wherein the second decision-support system comprises a plurality of neural networks, each neural network of the plurality having a unique set of starting weights and having a performance rating value.

65. The method of claim 64, wherein the final performance estimate is generated by averaging the performance rating values for the plurality of neural networks.

66. The method of claim 45, wherein the observation values are obtained from patient historical data results and/or biochemical test results.

67. The method of claim 45, wherein the condition is a pregnancy-related condition or endometriosis.

68. The method of claim 45, further comprising:

(e) collecting addition observations from patients and categorizing them into a set of candidate variables, which are then added to first set of candidate variables; and then

(f) repeating steps (c) and (d).

69. The method of claim 24, wherein the test assesses the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

70. The method of claim 24, wherein the test assess the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

71. The method of claim 45, further comprising identifying any biochemical test data variable(s) that end up in the sinal subset of selected important variables, whereby the identified biochemical test data variable(s) serve as indicators of the disease, disorder or condition.

72. The method of claim 71, wherein the test assesses the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

73. The method of claim 71, wherein the test assess the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

74. The method of claim 71, wherein the decision-support system includes a neural network and the final set constitutes a consensus of neural networks.

75. The method of claim 71, wherein the first subset of relevant variables is identified using sensitivity analysis performed on the decision-support based system or consensus thereof.

76. The method of claim 71, wherein the first decision-support system includes at least one neural network.

77. The method of claim 71, wherein the second decision-support system includes at least one neural network.

78. The method of claim 71, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iii) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (ii), wherein the best candidate variable does not improve performance, the process is completed.

79. The method of claim 71, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(iii) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iv) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (iii), if the variable does not improve performance in comparison to the performance of the selected important variables, evaluating is continued at step (v);

(v) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (iv).

80. The method of claim 79, wherein the candidate variables include biochemical test data.

81. The method of claim 79, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

82. The method of claim 79, wherein ranking is based on process comprising a statistical analysis.

83. The method of claim 79, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

84. The method of claim 79, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

85. The method of claim 80, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

86. The method of claim 85, wherein the sensitivity analysis comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

87. The method of claim 86, further comprising:

(vi) ranking the variables according to their relative contribution to the determination of the decision-support system output.

88. The method of claim 71, wherein the step of training a second decision-support system includes a validating step wherein a previously-unused set of observation data is run through the second decision-support system after training to provide a performance estimate for indication of the medical condition, wherein the previously-unused set of observation data are collected from patients in whom the medical condition is known.

89. The method of claim 71, wherein the step of training a second decision-support system includes partitioning the observation data set into a plurality of partitions comprising at least one testing data partition and a plurality of training data partitions, wherein the second decision-support system is run using the plurality of training data partitions and the testing data partition is used to provide a final performance estimate for the second decision-support system after the training data partitions have been run.

90. The method of claim 89, wherein the second decision-support system comprises a plurality of neural networks, each neural network of the plurality having a unique set of starting weights and having a performance rating value.

91. The method of claim 90, wherein the final performance estimate is generated by averaging the performance rating values for the plurality of neural networks.

92. The method of claim 71, wherein the observation values are obtained from patient historical data results and/or biochemical test results.

93. The method of claim 71, wherein the condition is a pregnancy-related condition or endometriosis.

94. The method of claim 71, further comprising:

(e) collecting addition observations from patients and categorizing them into a set of candidate variables, which are then added to first set of candidate variables; and then

(f) repeating steps (c) and (d).

95. The method of claim 71, further comprising developing a diagnostic biochemical test for the identified biochemical test data variable(s).

96. A method for developing new biochemical tests or identifying new disease markers, comprising:

performing the method of claim 71, and

identifying biochemical data variables that are selected important variables;

and developing tests that detect the biochemical data or disease marker from which the variable is derived.

97. In a computer system, a method for analyzing effectiveness of a diagnostic test to aid in diagnosing the presence, absence or severity of a medical condition or to assess a course of treatment or the effectiveness of a particular treatment in a patient comprising:

(a) collecting observations from a group of test patients in whom the medical condition is known;

(b) categorizing the observations into a set of candidate variables having observation values and storing the observation values as a observation data set in a computer;

(c) selecting a subset of selected important variables from the set of candidate variables by classifying the observation data set using a first decision-support system programmed into the computer system; and

(d) training a second decision-support system using the observation data corresponding to the subset of selected important variables;

(e) collecting results of the diagnostic test under analysis or collecting observations after or during treatment from same the group of test patients;

(f) categorizing the observations into a second set of candidate variables having observation values, combining them with the observations from step (b), and storing the observation values as a observation data set in a computer;

(g) selecting a second subset of selected important variables by classifying the observation data set using a first decision-support system programmed into the computer system;

(h) training a third decision-support system using the observation data corresponding to a subset of selected important variables from step (g);

(i) comparing the performance of the second and third systems, and thereby identifying assessing the effectiveness of a diagnostic test to aid in diagnosing the presence, absence or severity of a medical condition or to assessing the effectiveness of a course of treatment or the effectiveness of a particular treatment in treating a disease, disorder or condition.

98. The method of claim 97, wherein the method assesses the effectiveness of a diagnostic test in aiding in a diagnosis.

99. The method of claim 97, wherein the method assesses the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

100. The method of claim 97, wherein the decision-support system includes a neural network and the final set constitutes a consensus of neural networks.

101. The method of claim 97, wherein the first subset of relevant variables is identified using sensitivity analysis performed on the decision-support based system or consensus thereof.

102. The method of claim 97, wherein the first decision-support system includes at least one neural network.

103. The method of claim 97, wherein the second decision-support system includes at least one neural network.

104. The method of claim 97, wherein the third decision-support system includes at least one neural network.

105. The method of claim 97, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iii) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (ii), wherein the best candidate variable does not improve performance, the process is completed.

106. The method of claim 97, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(iii) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iv) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (iii), if the variable does not improve performance in comparison to the performance of the selected important variables, evaluating is continued at step (v);

(v) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (iv).

107. The method of claim 106, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

108. The method of claim 106, wherein ranking is based on process comprising a statistical analysis.

109. The method of claim 106, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

110. The method of claim 107, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

111. The method of claim 107, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

112. The method of claim 111, wherein the sensitivity analysis comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

113. The method of claim 112, further comprising:

(vi) ranking the variables according to their relative contribution to the determination of the decision-support system output.

114. The method of claim 97, wherein the step of training a second and/or third decision-support system includes a validating step wherein a previously-unused set of observation data is run through the second decision-support system after training to provide a performance estimate for indication of the medical condition, wherein the previously-unused set of observation data are collected from patients in whom the medical condition is known.

115. The method of claim 97, wherein the step of training a second and/or third decision-support system includes partitioning the observation data set into a plurality of partitions comprising at least one testing data partition and a plurality of training data partitions, wherein the second decision-support system is run using the plurality of training data partitions and the testing data partition is used to provide a final performance estimate for the second decision-support system after the training data partitions have been run.

116. The method of claim 115, wherein the each decision-support system comprises a plurality of neural networks, each neural network of the plurality having a unique set of starting weights and having a performance rating value.

117. The method of claim 116, wherein the final performance estimate is generated by averaging the performance rating values for the plurality of neural networks.

118. The method of claim 97, wherein the observation values are obtained from patient historical data results and/or biochemical test results.

119. The method of claim 97, wherein the condition is a pregnancy-related condition or endometriosis.

120. The method of claim 97, further comprising:

collecting addition observations from patients and categorizing them into a set of candidate variables, which are then added to first set of candidate variables at step (b); and then

(j) repeating steps (c) - (i).

121. In a computer system, a method for developing a condition-specific biochemical test to aid in diagnosing the presence, absence, or severity of a medical condition in a patient comprising:

(a) collecting test results of a biochemical test from a group of test patients in whom the condition is known or suspected;

(b) categorizing the observations into a set of candidate variables having observation values and storing the observation values as a observation data set in a computer;

(c) selecting a subset of selected important variables from a set of variables comprising the candidate variables by classifying the observation data set using a first decision-support system programmed into the computer system, whereby the subset of selected important variables includes the candidate variables substantially indicative of the medical condition; and

(d) identifying those variable(s) in the selected important variable set that correspond to biochemical data; and

(e) designing or selecting a biochemical test that assesses the data that corresponds to the identified variable(s).

122. The method of claim 121, wherein the decision-support system includes neural network or consensus thereof.

123. The method of claim 121, wherein the test assesses the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

124. The method of claim 121, wherein the test assess the presence, absence, severity or course of treatment of a disease, disorder or other medical condition or aids in determining the outcome resulting from a selected treatment.

125. The method of claim 121, wherein the decision-support system includes a neural network and the final set constitutes a consensus of neural networks.

126. The method of claim 121, wherein the first subset of relevant variables is identified using sensitivity analysis performed on the decision-support based system or consensus thereof.

127. The method of claim 121, wherein the first decision-support system includes at least one neural network.

128. The method of claim 121, wherein the second decision-support system includes at least one neural network.

129. The method of claim 121, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) taking candidate variables one at a time and evaluating each by training a decision-support system on that variable combined with the current set of selected important variables;

(iii) selecting the best of the candidate variables, wherein the best variable is any one that gives the highest performance of the decision-support system, and if the best candidate variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (ii), wherein the best candidate variable does not improve performance, the process is completed.

130. The method of claim 121, wherein the step of selecting a subset of selected important variables includes:

(i) providing a first set of n candidate variables and a second set of selected important variable, wherein the second set is initially empty;

(ii) ranking all candidate variables, wherein the ranking is either arbitrary or ordered;

(iii) taking the highest m ranked variables one at a time, wherein m is from 1 up to n, and evaluating each by training a decision-support

system on that variable combined with the current set of selected important variables;

(iv) selecting the best of the m variables, wherein the best variable is the one that gives the highest performance of the decision-support system, and if the best variable improves performance compared to the performance of the selected important variables, adding it to the selected important variable set, removing it from the candidate set and continuing evaluating at step (iii), if the variable does not improve performance in comparison to the performance of the selected important variables, evaluating is continued at step (v);

(v) determining if all variables on the candidate set have been evaluated, wherein if they have been evaluated, the process is complete and the set of selected important variables is a completed set, otherwise continuing by taking the next highest m ranked variables one at a time, and evaluating each by training a decision-support-system on that variable combined with the current set of important selected variables and performing step (iv).

131. The method of claim 130, wherein the candidate variables include biochemical test data.

132. The method of claim 130, wherein ranking is based on an analysis comprising a sensitivity analysis or other decision-support system-based analysis.

133. The method of claim 130, wherein ranking is based on process comprising a statistical analysis.

134. The method of claim 130, wherein ranking is based on a process comprising chi square, regression analysis or discriminant analysis.

135. The method of claim 131, wherein ranking is determined by a process that uses evaluation by an expert, a rule based system, a sensitivity analysis or combinations thereof.

136. The method of claim 131, wherein the sensitivity analysis, comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

137. The method of claim 136, wherein the sensitivity analysis comprises:

(i) determining an average observation value for each of the variables in the observation data set;

(ii) selecting a training example, and running the example through a decision-support system to produce an output value, designated and stored as the normal output;

(iii) selecting a first variable in the selected training example, replacing the observation value with the average observation value of the first variable, running the modified example in the decision-support system in the forward mode and recording the output as the modified output;

(iv) squaring the difference between the normal output and the modified output and accumulating it as a total, wherein the total for each variable designed the selected variable total for each variable;

(v) repeating steps (iii) and (iv) for each variable in the example;

(vi) repeating steps (ii)-(v) for each example in the data set, wherein each total for the selected variable represents the relative contribution of each variable to the determination of the decision-support system output.

138. The method of claim 137, further comprising:

(vi) ranking the variables according to their relative contribution to the determination of the decision-support system output.

139. The method of claim 121, wherein the step of training a second decision-support system includes a validating step wherein a previously-unused set of observation data is run through the second decision-support system after training to provide a performance estimate for indication of the medical condition, wherein the previously-unused set of observation data are collected from patients in whom the medical condition is known.

140. The method of claim 121, wherein the step of training a second decision-support system includes partitioning the observation data

set into a plurality of partitions comprising at least one testing data partition and a plurality of training data partitions, wherein the second decision-support system is run using the plurality of training data partitions and the testing data partition is used to provide a final performance estimate for the second decision-support system after the training data partitions have been run.

141. The method of claim 140, wherein the second decision-support system comprises a plurality of neural networks, each neural network of the plurality having a unique set of starting weights and having a performance rating value.

142. The method of claim 141, wherein the final performance estimate is generated by averaging the performance rating values for the plurality of neural networks.

143. The method of claim 121, wherein the observation values are obtained from patient historical data results and/or biochemical test results.

144. The method of claim 121, wherein the condition is a pregnancy-related condition or endometriosis.

145. The method of claim 121, further comprising:

(e) collecting addition observations from patients and categorizing them into a set of candidate variables, which are then added to first set of candidate variables; and then

(f) repeating steps (c) and (d).

146. A method for diagnosing endometriosis, comprising assessing a subset containing at least three of the following variables:

01. 10. Past history of Endo

02. 6. number of births

03. 14. dysmenorrhea

04. 1. age (preproc)

05. 13. pelvic pain
06. 11. history of pelvic surgery
07. 4. smoking (packs/day)
08. 12. medication History
09. 5. number of pregnancies
10. 7. number of abortions
11. 9. Abnormal PAP smear/dysplasia
12. 3. Pregnancy hyperplasia
13. 8. Genital Warts
14. 2. Diabetes

using a decision-support system that has been trained to diagnose endometriosis.

147. The method of claim 146, wherein the selected subset of these variables contains one or more of the following combinations of three variables set forth in sets (a)-(m):

a) number of births, history of endometriosis, history of pelvic surgery;

b) diabetes, pregnancy hypertension, smoking;

c) pregnancy hypertension, abnormal pap smear/dysplasia, history of endometriosis;

d) age, smoking, history of endometriosis;

e) smoking, history of endometriosis, dysmenorrhea;

f) age, diabetes, history of endometriosis;

g) pregnancy hypertension, number of births, history of endometriosis;

h) Smoking, number of births, history of endometriosis;

i) pregnancy hypertension, history endometriosis, history of pelvic surgery;

j) number of pregnancies, history of endometriosis, history of pelvic surgery;

k) number of births, abnormal PAP smear/dysplasia, history of endometriosis;

l) number of births, abnormal PAP smear/dysplasia, dysmenorrhea;

m) history of endometriosis, history of pelvic surgery, dysmenorrhea; and

n) number of pregnancies, history of endometriosis, dysmenorrhea.

148. The method of claim 147, wherein the decision support system is a neural network.

(c) selecting a subset of selected important variables from the set of candidate variables by classifying the observation data set using a first decision-support system programmed into the computer system, whereby the subset of selected important variables includes the candidate variables substantially indicative of the medical condition; and

(d) training a second decision-support system using the observation data corresponding to a subset of selected important variables, whereby the second decision-support system-based system constitutes a decision-support based diagnostic test for the condition, disease or disorder.

149. The method of claim 24, wherein the disorder is endometriosis and the candidate variables comprise at least four of the variables selected from:

(i) past history of endometriosis, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of

pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts, and diabetes, or

(ii) age, parity, gravidity, number of abortions, smoking quantity per day, past history of endometriosis, dysmenorrhea, pelvic pain, abnormal PAP, history of pelvic surgery, medication history, pregnancy hypertension, genital warts and diabetes.

150. The method of claim 149, wherein the decision-support system comprises a neural network or a consensus of neural networks.

151. The method of claim 149, wherein at least five variables are selected.

152. In a computer system, a method to aid in diagnosis of the presence, absence or severity of endometriosis in a patient comprising:

(a) collecting observation values reflecting presence and absence of specified clinical data factors and storing the observed clinical data factors in storage means of the computer system, the specified clinical data factors comprising at least four of the factors selected from:

(i) past history of endometriosis, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts, and diabetes, or

(ii) age, parity, gravidity, number of abortions, smoking quantity per day, past history of endometriosis, dysmenorrhea, pelvic pain, abnormal PAP, history of pelvic surgery, medication history, pregnancy hypertension, genital warts and diabetes;

(b) applying the observation values from the memory means to a first decision-support system trained on samples of the specified factors; and thereupon

(c) extracting from the first decision-support system an output value, wherein the output value is a quantitative objective aid to enhance decision processes for a diagnosis of endometriosis.

153. The method of claim 152, wherein the decision-support system comprises a neural network.

154. The method of claim 152, wherein at least five factors are selected.

155. The method of claim 152, further comprising:

b1) applying said observation values from said memory means to a plurality of the first decision-support system, wherein each one of the first decision-support systems is trained on the samples of the specified factors with different starting weights for each training;

c1) extracting from the first decision-support system, output value pairs for each one of said first neural networks; and

d) forming a linear combination of said first ones of said output value pairs and forming a linear combination of said second ones of said output value pairs, to obtain a confidence index pair, said confidence index pair being said quantitative objective aid.

156. The method of claim 155, wherein the first decision support system is a neural network that comprises a three-layer network containing an input layer, a hidden layer and an output layer, the input layer having fourteen input nodes, first and second hidden layer nodes, a hidden layer bias for each the hidden layer node, first and second output layer nodes in the output layer, and an output layer bias for each the output layer node.

157. The method of claim 155, wherein the first decision support system is a neural network and each of the plurality of first trained neural networks comprises a three-layer network comprising an input layer, a hidden layer and an output layer.

158. The method of claim 155, wherein the first decision support system is a neural networks and each comprises a three-layer network comprising an input layer, a hidden layer and an output layer, the input layer having fourteen input nodes, first and second hidden layer nodes and a hidden layer bias for each the hidden layer node and first and second output layer nodes in the output layer and an output layer bias for each the output layer node,

wherein weights, in order of identification as follows:

0. Bias
1. Age
2. Diabetes
3. Pregnancy hypertension
4. Smoking Packs/Day
5. Number of Pregnancies
6. Number ofBirths
7. Number ofAbortions
8. Genital Warts
9. Abnormal PAP/Dysplasia
10. History of Endometriosis
11. History of Pelvic Surgery
12. Medication History
13. Pelvic Pain
14. Dysmenorrhea

are as follows for each of eight the first neural networks:

First neural network A

to processing element at the first hidden layer node:

0.15 -1.19 -0.76 3.01 1.81 1.87

3.56 -0.48 1.33 -1.96 -4.45 1.36

-1.61 -1.97 -0.91

to processing element at the second hidden layer node:

0.77 2.25 -2.30 -1.48 -0.85 0.27

-1.70 -0.47 0.84 -6.19 0.50 -0.95

0.40 2.38 1.86

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

-0.12 -0.44 0.66

to processing element at the second output layer node:

0.12 0.44 -0.65

First neural network B

to processing element at the first hidden layer node:

-0.16 -3.30 0.85 1.00 0.99 -0.81

1.57 -1.40 0.46 1.16 -0.80 -0.01 -1.19 -1.10 -2.29

to processing element at the second hidden layer node:

-1.62 0.79 0.45 2.14 3.82 3.93

3.96 2.27 -0.54 1.51 -4.76 2.83

0.74 -0.43 -0.17

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.70 -0.69 -0.65

to processing element at the second output layer node:

-0.70 0.69 0.65

First neural network C

to processing element at the first hidden layer node:

0.94 1.43 0.29 1.17 2.11 -1.16

1.033 -0.68 -0.88 0.31 -1.74 1.62

-1.49 -1.05 -0.41

to processing element at the second hidden layer node:

0.77 3.31 -1.48 -0.83 0.60 -2.09

-1.39 -0.40 -0.19 -0.89 1.36 0.59

-1.11 0.26 1.04

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.10 -0.90 0.87

to processing element at the second output layer node:

-0.10 0.90 -0.87

First neural network D

to processing element at the first hidden layer node:

1.08 1.27 -0.89 -1.00 -1.74 -0.40

-1.38 1.26 1.06 0.66 0.71 -0.57

0.67 1.89 -0.90

to processing element at the second hidden layer node:

-0.03 -0.58 -0.46 -0.94 0.73 0.10

0.55 -0.79 -0.098 -1.36 1.01 0.00

-0.38 -0.49 1.57

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

-1.43 1.39 1.28

to processing element at the second output layer node:

1.30 -1.28 -1.17

First neural network E

to processing element at the first hidden layer node:

0.14 -2.12 8.36 1.02 1.79 0.31

2.87 0.84 -1.24 -1.75 -2.98 1.72

-1.22 -2.47 -1.14

to processing element at the second hidden layer node:

-3.93 -1.07 1.16 1.39 1.01 -1.08

2.33 0.76 -0.51 -0.31 -1.92 0.59

0.06 -0.76 -1.44

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.46 -0.52 -0.80

to processing element at the second output layer node:

-0.46 0.51 0.82

First neural network F

to processing element at the first hidden layer node:

-1.19 -2.93 1.19 6.85 1.08 0.66

1.65 -0.28 -1.63 -1.15 -0.79 0.43

-0.13 -3.10 -2.27

to processing element at the second hidden layer node:

0.82 0.19 0.72 0.83 0.59 0.07

1.06 0.51 1.04 1.47 -1.97 0.97

-0.91 -0.15 0.09

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.68 -0.67 -0.58

to processing element at the second output layer node:

-0.68 0.67 0.58

First neural network G

to processing element at the first hidden layer node:

-1.18 -2.55 0.48 -1.40 1.11 -0.28

2.33 0.33 -1.92 0.99 -1.41 0.68

-0.28 -1.65 -0.79

to processing element at the second hidden layer node:

1.07 1.11 0.52 1.41 0.55 -0.48

-0.23 0.44 -1.23 0.77 -2.96 1.39

-0.28 -0.64 -2.38

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.69 -0.70 -0.50

to processing element at the second output layer node:

-0.69 0.70 0.50

First neural network H

to processing element at the first hidden layer node:

15.74 -0.76 -0.91 -1.13 -0.75 -0.66

-0.83 1.03 0.75 -0.48 -0.47 2.01

-0.02 0.25 1.11

to processing element at the second hidden layer node:

-2.48 -2.49 0.99 1.97 2.41 1.51

1.01 -0.26 -0.76 2.00 -5.03 1.77

-0.77 -2.29 -2.01

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.02 0.41 -0.84

to processing element at the second output layer node:

-0.75 0.34 0.85.

159. The method of claim 158, wherein normalized observation values for each one of the first neural networks have the following mean and standard deviations, in order of the identification:

-0.00 1.00
0.01 0.08
0.01 0.09
0.16 0.37
1.09 1.39
0.55 0.94
0.54 0.93
0.01 0.10
0.03 0.17
0.23 0.42
0.65 0.48
0.39 0.49
0.19 0.39
0.72 0.45.

160. In a computer system, a method to aid in diagnosis of the presence, absence or severity of endometriosis in a patient comprising the steps of:

(a) collecting observation values reflecting presence and absence of specified factors and storing the observation factors in storage means of the computer system, the specified factors comprising: past history of the disease, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts, and diabetes;

(b) obtaining results from the patient of a biochemical test relevant to endometriosis and storing in the memory mean;

(c) applying the observation values and the relevant biochemical test results from the memory means to a second neural network trained on samples of the specified factors and the test results; and thereupon

(d) extracting from the second trained neural network an output value pair, the output value pair being a preliminary indicator for the diagnosis of endometriosis.

161. The method of claim 160, further including the steps of:

(c1) applying the observation values and the relevant biochemical test results from the memory means to a plurality of the second neural networks, each one of the first neural networks being trained on the samples of the specified factors with starting weights for each training being randomly initialized;

(d1) extracting from each one of the first trained neural networks, output value pairs for each one of the first neural networks; and

(e) forming a linear combination of the first ones of the output value pairs and forming a linear combination of the second ones of the output value pairs, to obtain a confidence index pair, the confidence index pair being a final indicator for the diagnosis of endometriosis.

162. The method of claim 160, wherein the first trained neural network comprises a three-layer network containing an input layer, a hidden layer and an output layer, the input layer having fifteen input nodes, first and second hidden layer nodes, a hidden layer bias for each the hidden layer node, first and second output layer nodes in the output layer, and an output layer bias for each the output layer node.

163. The method of claim 161, wherein the plurality of the second trained neural networks each comprises a three-layer network containing an input layer, a hidden layer and an output layer, the input layer having fifteen input nodes, first and second hidden layer nodes and a hidden layer bias for each the hidden layer node and first and second output layer nodes in the output layer and an output layer bias for each the output layer node,

wherein weights, in order of identification as follows:

0. Bias
1. Age
2. Diabetes
3. Pregnancy hypertension
4. Smoking Packs/Day
5. Number ofPregnancies
6. Number ofBirths
7. Number ofAbortions
8. Genital Warts
9. Abnormal PAP/Dysplasia
10. History of Endometriosis
11. History of Pelvic Surgery
12. Medication History
13. Pelvic Pain
14. Dysmenorrhea
15. Biochemical test results.

164. In a computer system, a neural network system to aid in diagnosis of the presence, absence or severity of endometriosis in a patient, the neural network system comprising:

a plurality of first trained neural networks each comprising a three-layer network containing an input layer, a hidden layer and an

output layer, the input layer having fourteen input nodes, first and second hidden layer nodes, a hidden layer bias for each the hidden layer node, first and second output layer nodes in the output layer, and an output layer bias for each the output layer node, each trained neural network for generating a preliminary indicator for the diagnosis of endometriosis;

input means for observed values of clinical data factors;

storage means of the computer system for the observed values of the clinical data factors, the clinical data factors comprising: past history of the disease, number of births, dysmenorrhea, age, pelvic pain, history of pelvic surgery, smoking quantity per day, medication history, number of pregnancies, number of abortions, abnormal PAP/dysplasia, pregnancy hypertension, genital warts, and diabetes; and

means for building a consensus from the output layer nodes, the consensus being a quantitative objective aid to enhance the decision process for the diagnosis of endometriosis.

165. The neural network system of claim 164, further including:

an input normalizer for normalizing the observation values from the memory means to a plurality of the first neural networks, each one of the first neural networks being trained on the samples of the specified factors with starting weights for each training being randomly initialized.

166. The neural network system of claim 164, wherein the consensus builder comprises a linear combiner of first ones of output value pairs and of second ones of output value pairs, to obtain a confidence index pair, the confidence index pair being the consensus and final indicator for the diagnosis of endometriosis.

167. The neural network system of claim 164, wherein weights, in order of identification as follows:

0. Bias
1. Age
2. Diabetes
3. Pregnancy hypertension
4. Smoking Packs/Day
5. Number ofPregnancies
6. Number ofBirths
7. Number ofAbortions
8. Genital Warts
9. Abnormal PAP/Dysplasia
10. History of Endometriosis
11. History of Pelvic Surgery
12. Medication History
13. Pelvic Pain
14. Dysmenorrhea

are as follows for each of eight the first neural networks:

First neural network A

to processing element at the first hidden layer node:

0.15 -1.19 -0.76 3.01 1.81 1.87
3.56 -0.48 1.33 -1.96 -4.45 1.36
-1.61 -1.97 -0.91

to processing element at the second hidden layer node:

0.77 2.25 -2.30 -1.48 -0.85 0.27
-1.70 -0.47 0.84 -6.19 0.50 -0.95
0.40 2.38 1.86

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

-0.12 -0.44 0.66

to processing element at the second output layer node:

0.12 0.44 -0.65

First neural network B

to processing element at the first hidden layer node:

-0.16 -3.29 0.85 1.00 0.99 -0.81

1.57 -1.40 0.46 1.16 -0.80 -0.01

-1.19 -1.10 -2.29

to processing element at the second hidden layer node:

-1.62 0.79 0.45 2.14 3.82 3.93

3.96 2.27 -0.54 1.51 -4.76 2.83

0.74 -0.43 -0.17

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.70 -0.69 -0.65

to processing element at the second output layer node:

-0.70 0.69 0.65

First neural network C

to processing element at the first hidden layer node:

0.94 1.43 0.29 1.17 2.11 -1.16

1.03 -0.68 -0.88 0.31 -1.74 1.62

-1.49 -1.05 -0.41

to processing element at the second hidden layer node:

0.77 3.31 -1.48 -0.83 0.60 -2.09

-1.39 -0.40 -0.19 -0.89 1.36 0.59

-1.11 0.26 1.04

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.10 -0.90 0.87

to processing element at the second output layer node:

-0.10 0.90 -0.87

First neural network D

to processing element at the first hidden layer node:

1.08 1.27 -0.89 -1.00 -1.74 -0.40

-1.38 1.26 1.06 0.66 0.71 -0.57

0.67 1.89 -0.90

to processing element at the second hidden layer node:

-0.03 -0.58 -0.46 -0.94 0.73 0.10

0.55 -0.79 -0.10 -1.36 1.01 0.00

-0.38 -0.49 1.57

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

-1.43 1.39 1.28

to processing element at the second output layer node:

1.30 -1.28 -1.17

First neural network E

to processing element at the first hidden layer node:

0.14 -2.12 8.36 1.02 1.79 0.31

2.87 0.84 -1.24 -1.75 -2.98 1.72

-1.22 -2.47 -1.14

to processing element at the second hidden layer node:

-3.93 -1.07 1.16 1.39 1.01 -1.08

2.33 0.76 -0.51 -0.31 -1.92 0.59

0.06 -0.76 -1.44

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.46 -0.52 -0.80

to processing element at the second output layer node:

-0.46 0.51 0.82

First neural network F

to processing element at the first hidden layer node:

-1.19 -2.93 1.19 6.85 1.08 0.66

1.65 -0.28 -1.63 -1.15 -0.79 0.43

-0.13 -3.10 -2.27

to processing element at the second hidden layer node:

0.82 0.19 0.72 0.83 0.59 0.07 1.06

0.51 1.04 1.47 -1.97 0.97

-0.91 -0.15 0.09

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.68 -0.67 -0.58

to processing element at the second output layer node:

-0.68 0.67 0.58

First neural network G

to processing element at the first hidden layer node:

-1.18 -2.55 0.48 -1.40 1.11 -0.28

2.33 0.33 -1.92 0.99 -1.41 0.68

-0.28 -1.65 -0.79

to processing element at the second hidden layer node:

1.08 1.11 0.52 1.41 0.55 -0.48

-0.23 0.44 -1.23 0.77 -2.96 1.39

-0.28 -0.64 -2.38

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.69 -0.70 -0.50

to processing element at the second output layer node:

-0.69 0.70 0.50

First neural network H

to processing element at the first hidden layer node:

15.74 -0.76 -0.91 -1.13 -0.75 -0.66

-0.83 1.03 0.75 -0.48 -0.47 2.01

-0.02 0.25 1.11

to processing element at the second hidden layer node:

-2.48 -2.49 0.99 1.97 2.41 1.51

1.01 -0.26 -0.76 2.00 -5.03 1.77

-0.77 -2.29 -2.01

output layer weights (0 bias, first hidden layer weight, second hidden layer weight)

to processing element at the first output layer node:

0.017 0.41 -0.84

to processing element at the second output layer node:

-0.75 0.34 0.85.

168. The system of claim 167, wherein normalized observation values for each one of the first neural networks have the following mean and standard deviations, in order of the identification:

-0.00 1.00

0.01 0.08

0.01 0.09

0.16 0.37
1.09 1.39
0.55 0.94
0.54 0.93
0.01 0.10
0.03 0.17
0.23 0.42
0.65 0.48
0.39 0.49
0.19 0.39
0.72 0.45.

169. The system of claim 164, further comprising storage means for biochemical results, and wherein the plurality of networks have been trained to include biochemical test results.

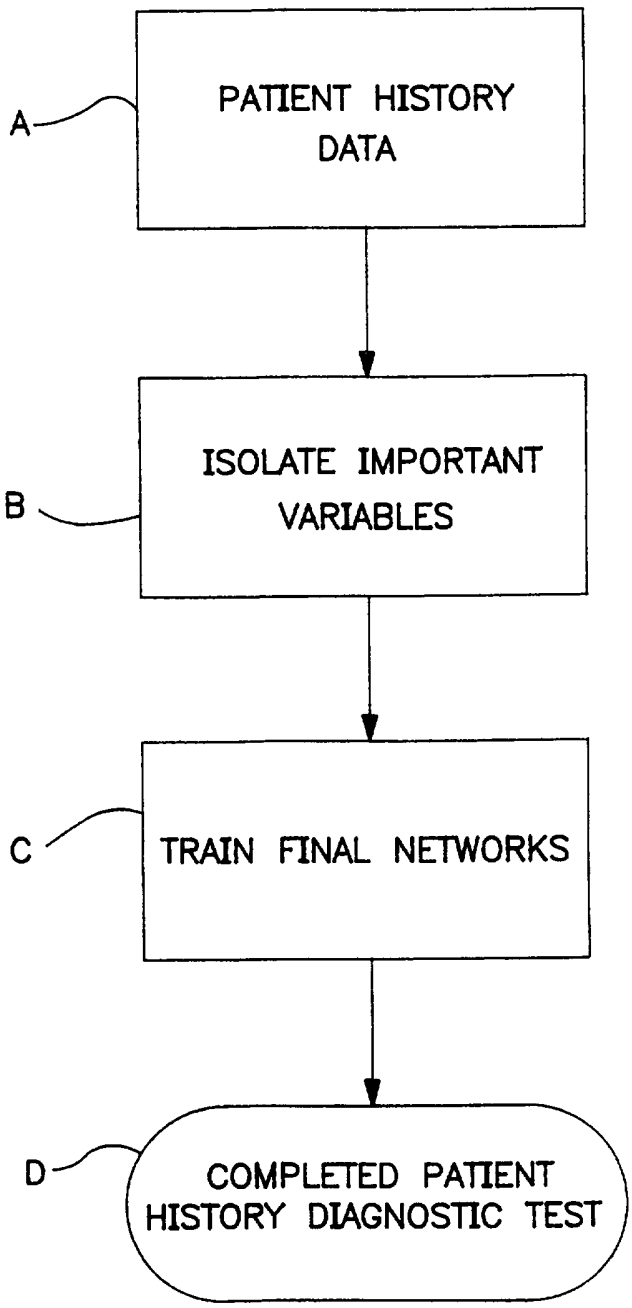
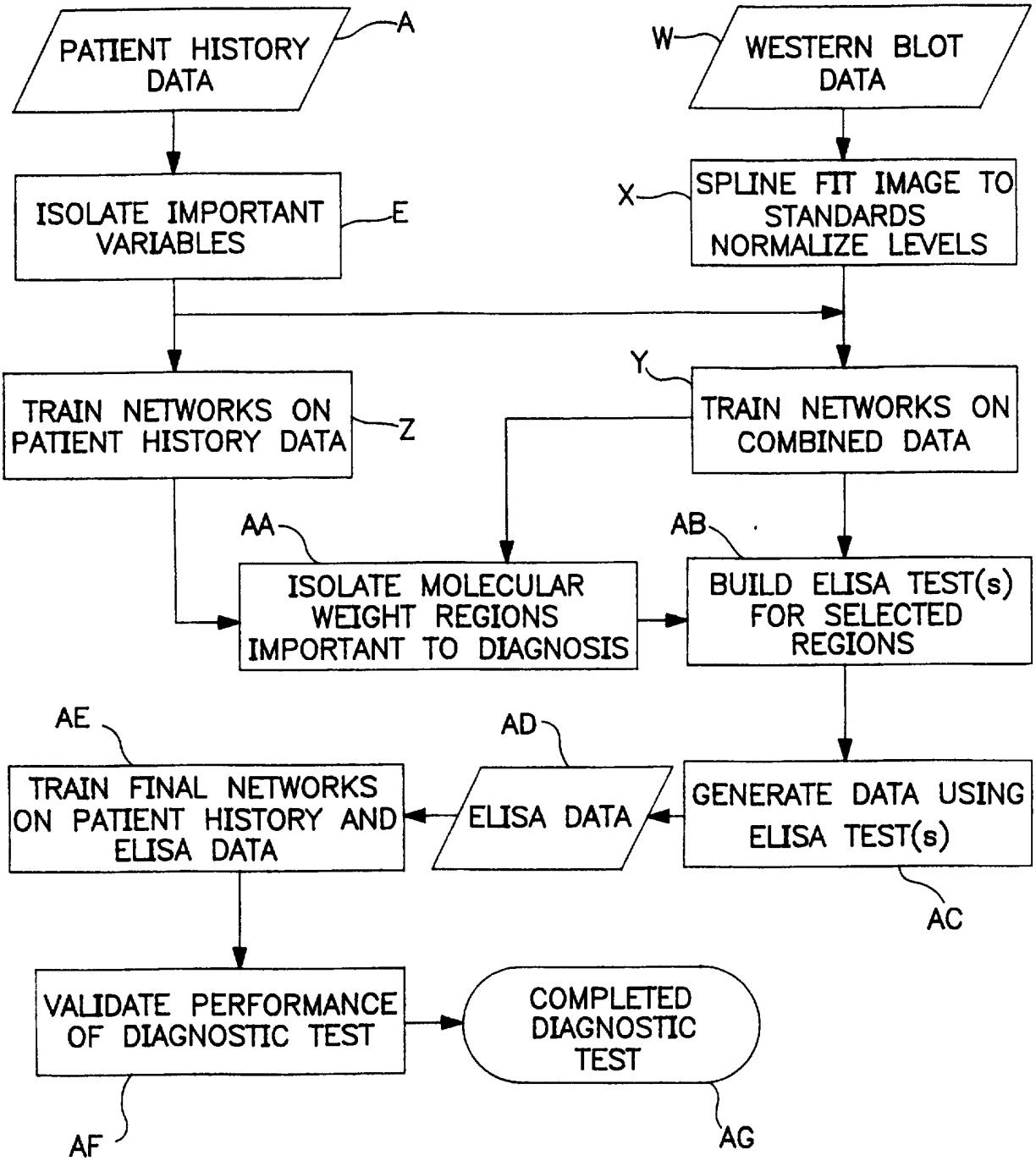

A
PATIENT HISTORY DATA
B
ISOLATE IMPORTANT VARIABLES
C
TRAIN FINAL NETWORKS
D
COMPLETED PATIENT HISTORY DIAGNOSTIC TEST

FIG. 1

FIG. 2


PATIENT HISTORY DATA
A
W
WESTERN BLOT DATA
ISOLATE IMPORTANT VARIABLES
E
X
SPLINE FIT IMAGE TO STANDARDS NORMALIZE LEVELS
Y
TRAIN NETWORKS ON PATIENT HISTORY DATA
Z
TRAIN NETWORKS ON COMBINED DATA
AA
AB
ISOLATE MOLECULAR WEIGHT REGIONS IMPORTANT TO DIAGNOSIS
BUILD ELISA TEST(s) FOR SELECTED REGIONS
AE
AD
TRAIN FINAL NETWORKS ON PATIENT HISTORY AND ELISA DATA
ELISA DATA
GENERATE DATA USING ELISA TEST(s)
AC
VALIDATE PERFORMANCE OF DIAGNOSTIC TEST
COMPLETED DIAGNOSTIC TEST
AF
AG

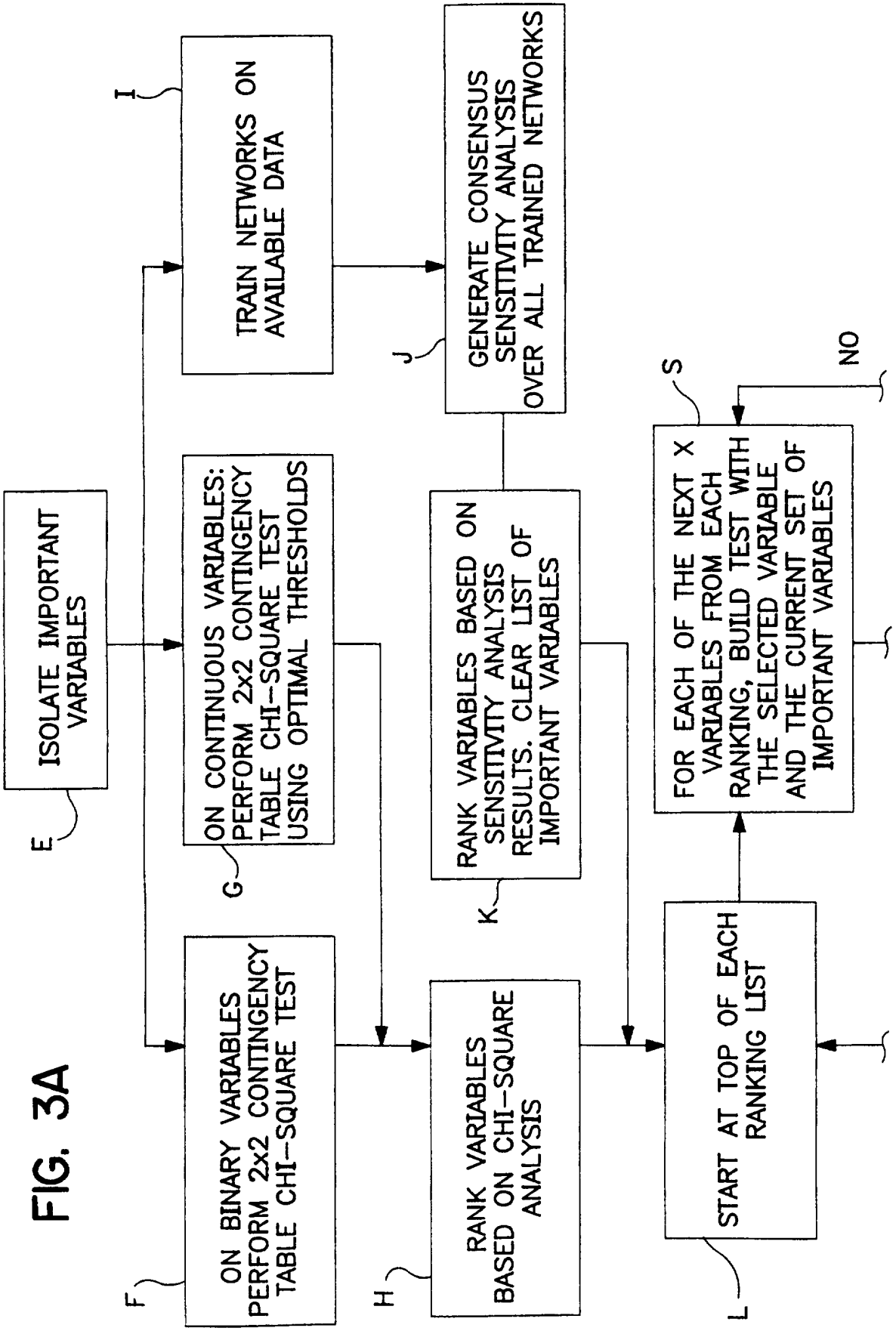

ISOLATE IMPORTANT VARIABLES
E
ON CONTINUOUS VARIABLES: PERFORM 2x2 CONTINGENCY TABLE CHI-SQUARE TEST USING OPTIMAL THRESHOLDS
G
ON BINARY VARIABLES PERFORM 2x2 CONTINGENCY TABLE CHI-SQUARE TEST
F
TRAIN NETWORKS ON AVAILABLE DATA
I
GENERATE CONSENSUS SENSITIVITY ANALYSIS OVER ALL TRAINED NETWORKS
J
RANK VARIABLES BASED ON SENSITIVITY ANALYSIS RESULTS. CLEAR LIST OF IMPORTANT VARIABLES
K
RANK VARIABLES BASED ON CHI-SQUARE ANALYSIS
H
START AT TOP OF EACH RANKING LIST
L
FOR EACH OF THE NEXT X VARIABLES FROM EACH RANKING, BUILD TEST WITH THE SELECTED VARIABLE AND THE CURRENT SET OF IMPORTANT VARIABLES
S
NO

FIG. 3A

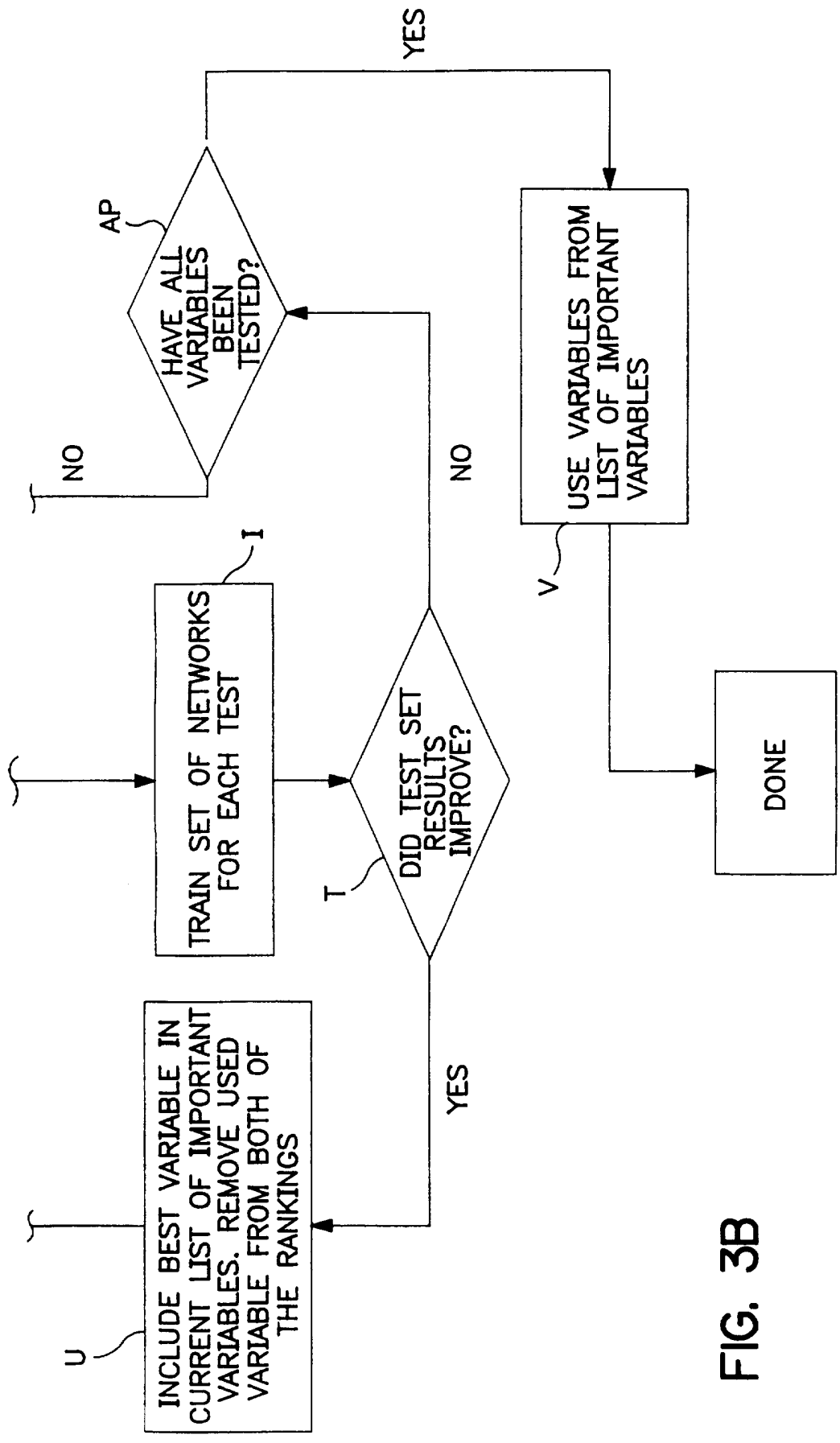

YES
AP
HAVE ALL VARIABLES BEEN TESTED?
NO
USE VARIABLES FROM LIST OF IMPORTANT VARIABLES
V
NO
I
TRAIN SET OF NETWORKS FOR EACH TEST
DID TEST SET RESULTS IMPROVE?
T
DONE
YES
INCLUDE BEST VARIABLE IN CURRENT LIST OF IMPORTANT VARIABLES. REMOVE USED VARIABLE FROM BOTH OF THE RANKINGS
U

FIG. 3B

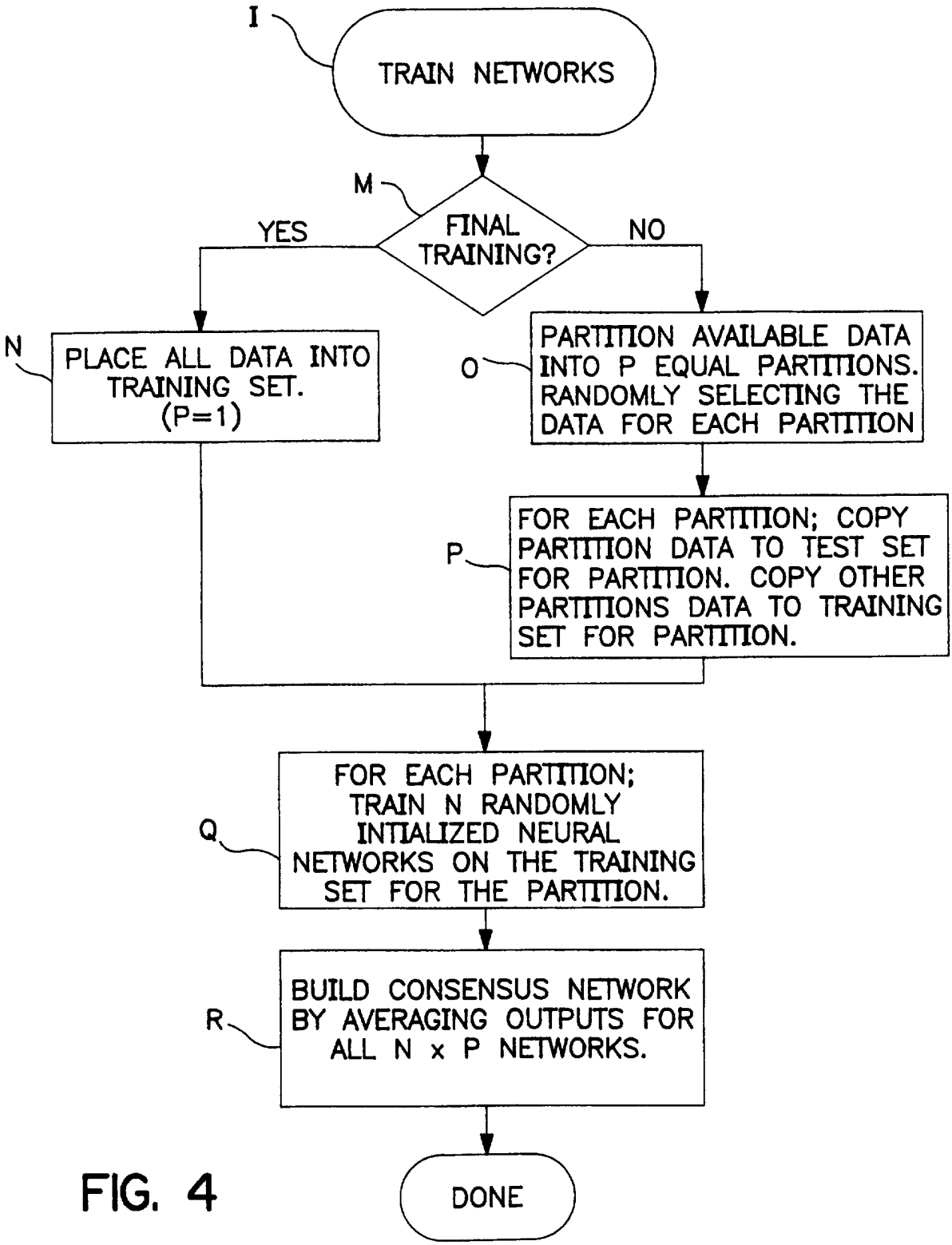

I
TRAIN NETWORKS
M
FINAL TRAINING?
YES
NO
N
PLACE ALL DATA INTO TRAINING SET. (P=1)
O
PARTITION AVAILABLE DATA INTO P EQUAL PARTITIONS. RANDOMLY SELECTING THE DATA FOR EACH PARTITION
P
FOR EACH PARTITION; COPY PARTITION DATA TO TEST SET FOR PARTITION. COPY OTHER PARTITIONS DATA TO TRAINING SET FOR PARTITION.
Q
FOR EACH PARTITION; TRAIN N RANDOMLY INTIALIZED NEURAL NETWORKS ON THE TRAINING SET FOR THE PARTITION.
R
BUILD CONSENSUS NETWORK BY AVERAGING OUTPUTS FOR ALL N x P NETWORKS.
DONE

FIG. 4

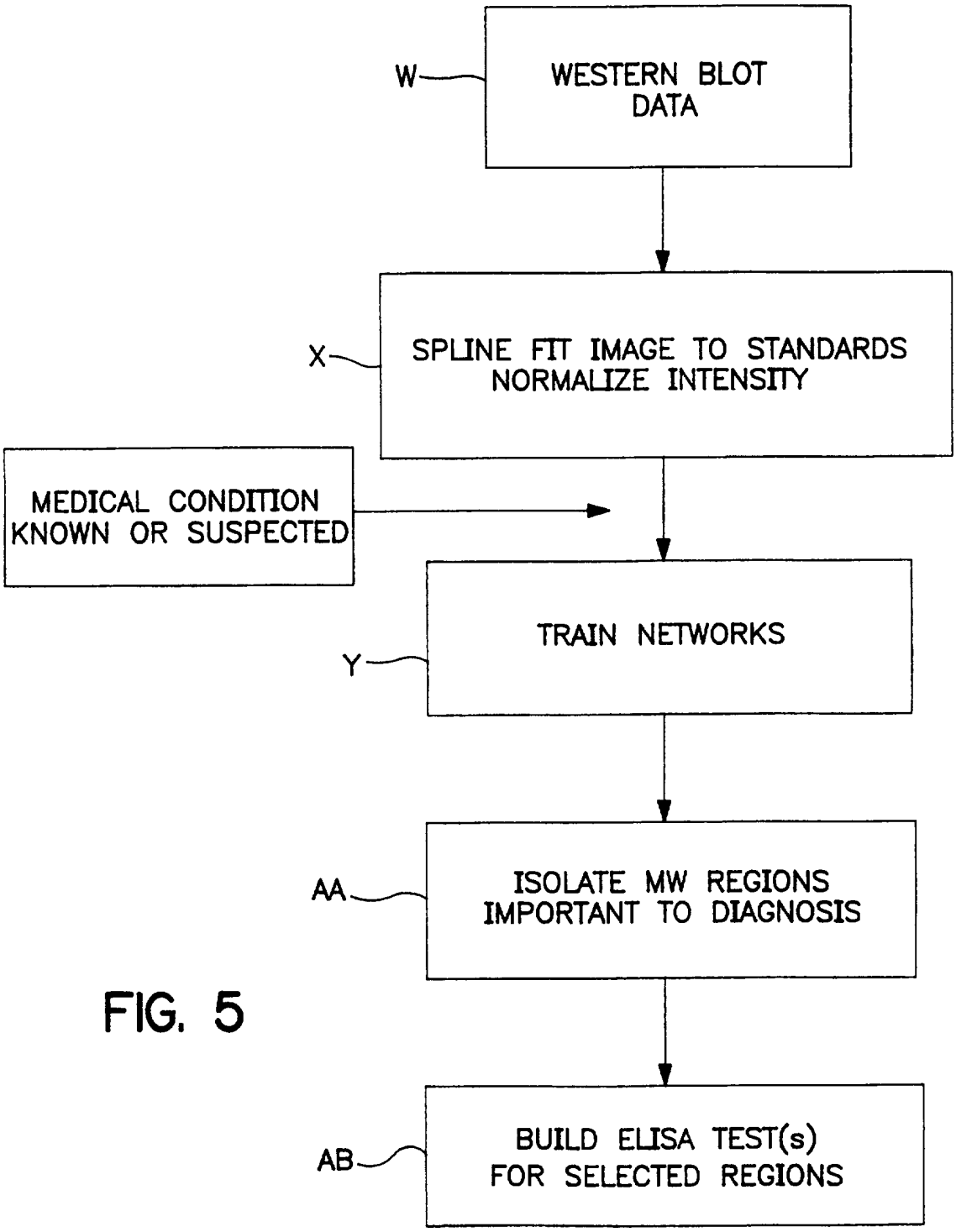

W
WESTERN BLOT DATA
X
SPLINE FIT IMAGE TO STANDARDS NORMALIZE INTENSITY
MEDICAL CONDITION KNOWN OR SUSPECTED
Y
TRAIN NETWORKS
AA
ISOLATE MW REGIONS IMPORTANT TO DIAGNOSIS
AB
BUILD ELISA TEST(s) FOR SELECTED REGIONS

FIG. 5

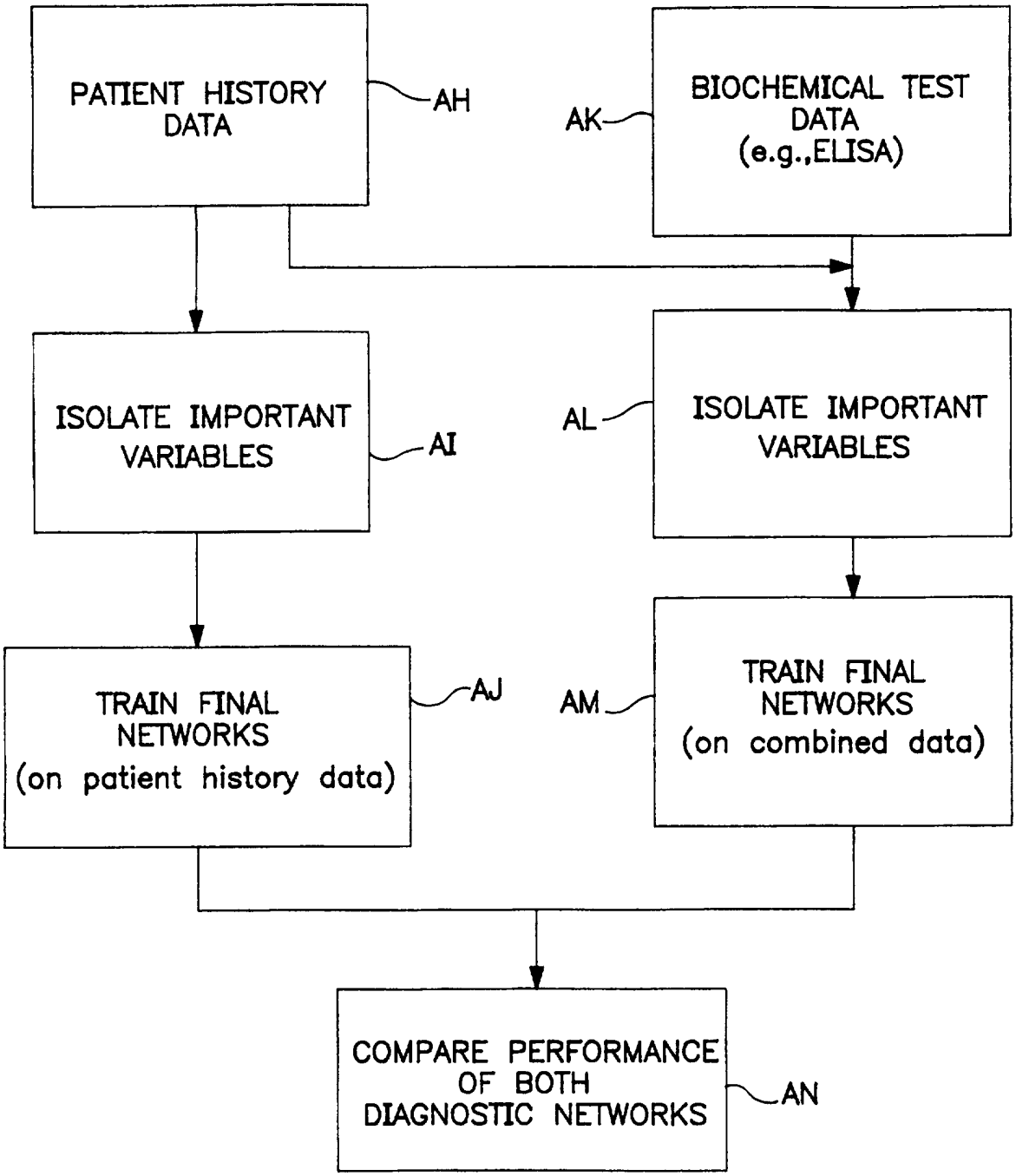

PATIENT HISTORY DATA
AH
AK
BIOCHEMICAL TEST DATA (e.g.,ELISA)
ISOLATE IMPORTANT VARIABLES
AI
AL
ISOLATE IMPORTANT VARIABLES
TRAIN FINAL NETWORKS (on patient history data)
AJ
AM
TRAIN FINAL NETWORKS (on combined data)
COMPARE PERFORMANCE OF BOTH DIAGNOSTIC NETWORKS
AN

FIG. 6

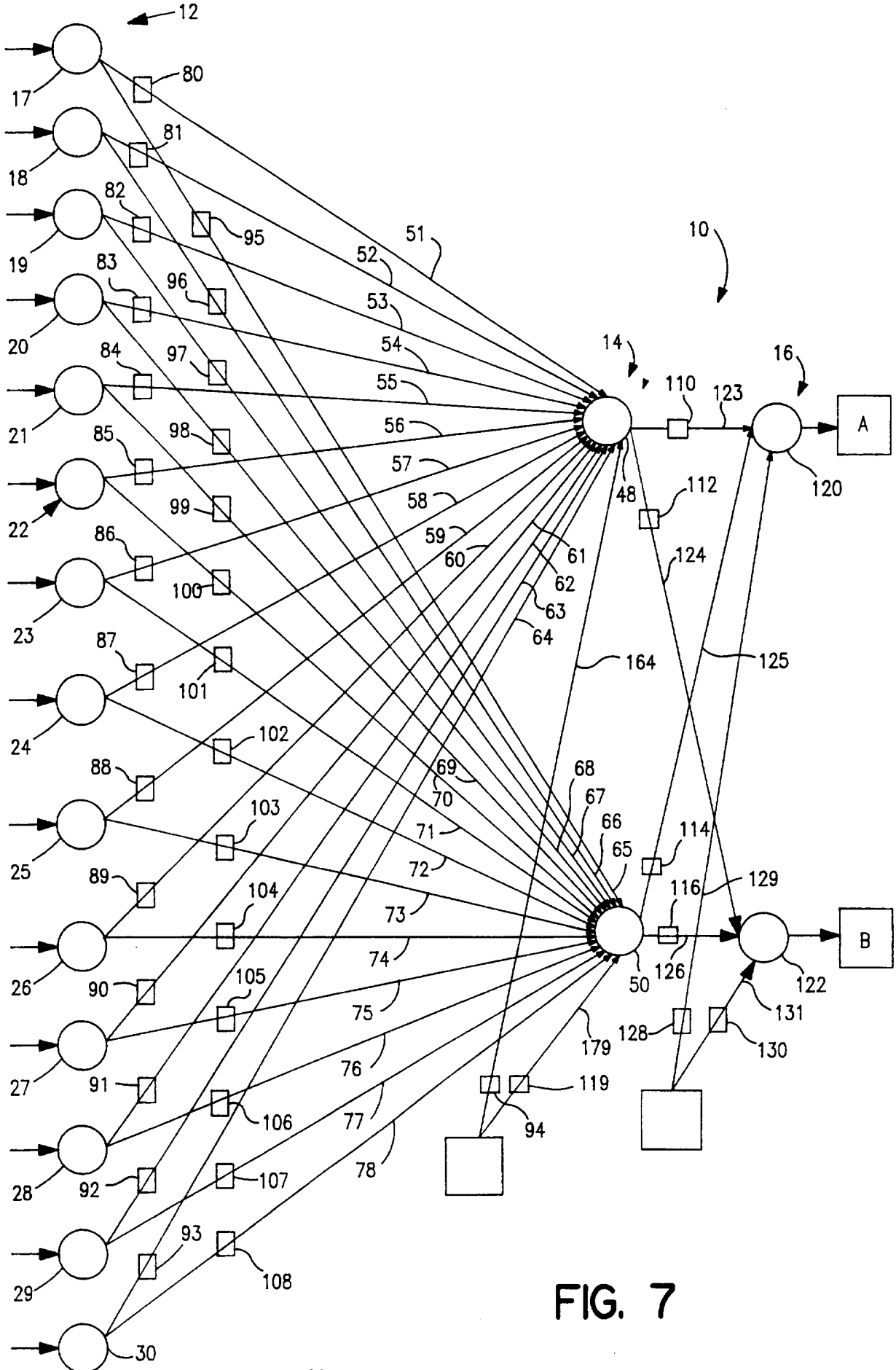

12
17
18
19
20
21
22
23
24
25
26
27
28
29
30
80
81
82
83
84
85
86
87
88
89
90
91
92
93
95
96
97
98
99
100
101
102
103
104
105
106
107
108
51
52
53
54
55
56
57
58
59
60
61
62
63
64
10
14
16
110
123
A
120
48
112
124
164
125
69
70
68
67
66
65
71
72
73
74
75
76
77
78
114
116
129
B
122
50
126
131
128
130
179
119
94

FIG. 7

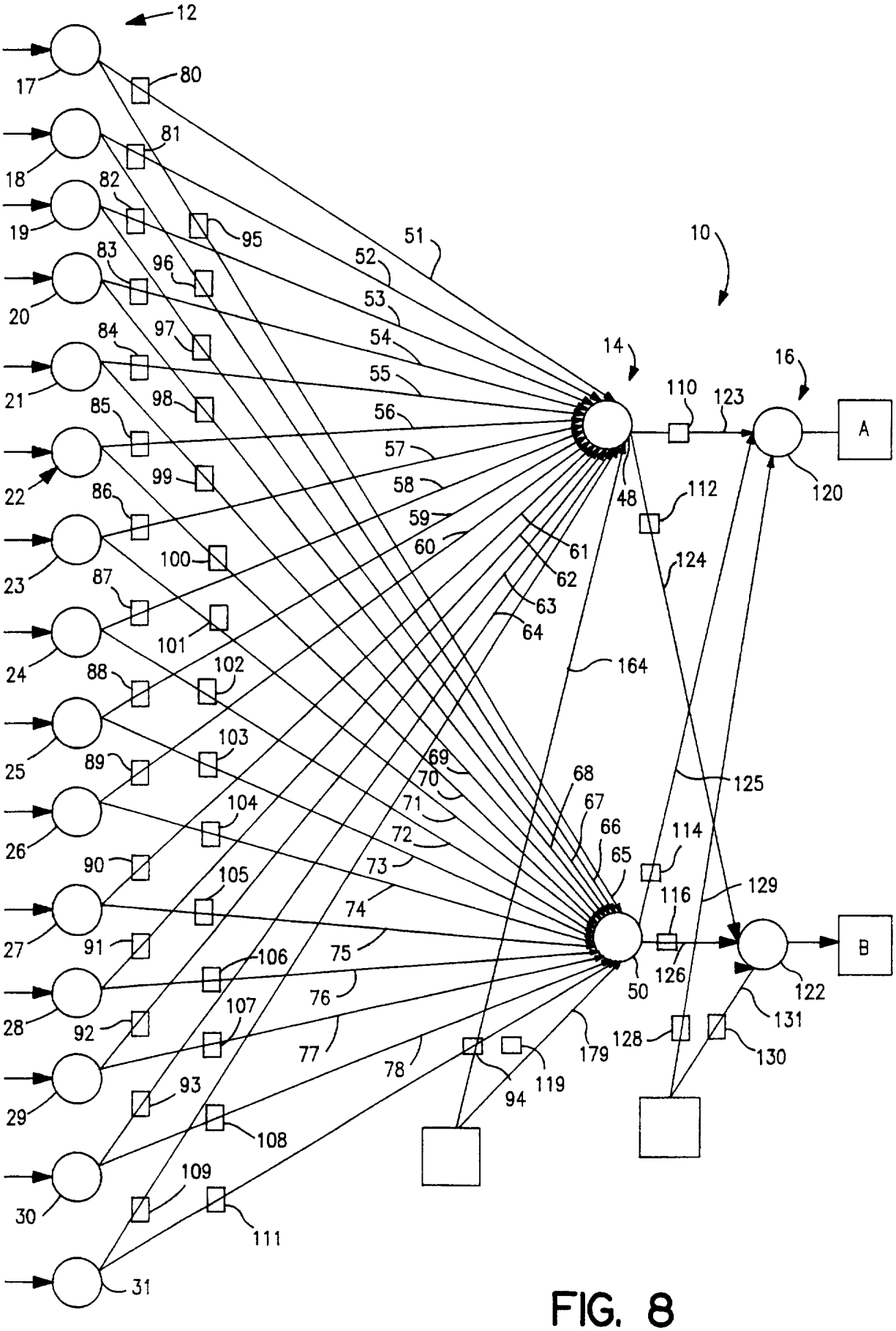

12
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
80
81
82
83
84
85
86
87
88
89
90
91
92
93
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
111
51
52
53
54
55
56
57
58
59
60
61
62
63
64
69
70
71
72
73
74
75
76
77
78
94
119
10
14
16
110
123
A
120
48
112
124
164
68
67
66
65
114
116
129
125
B
50
126
122
131
128
130
179

FIG. 8

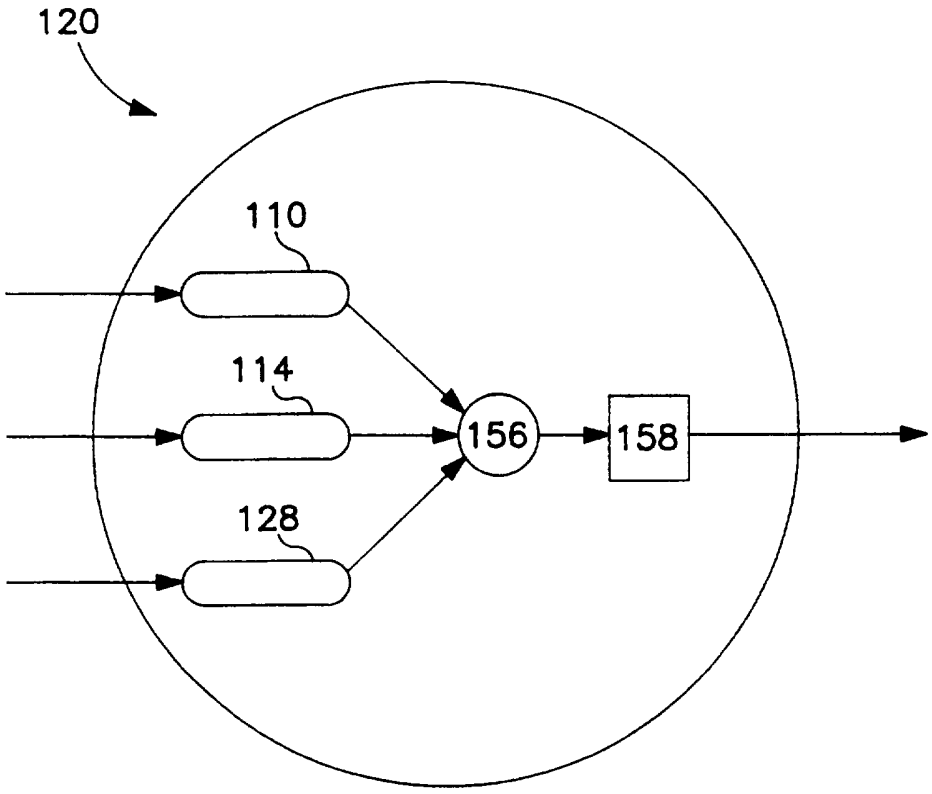


FIG. 9

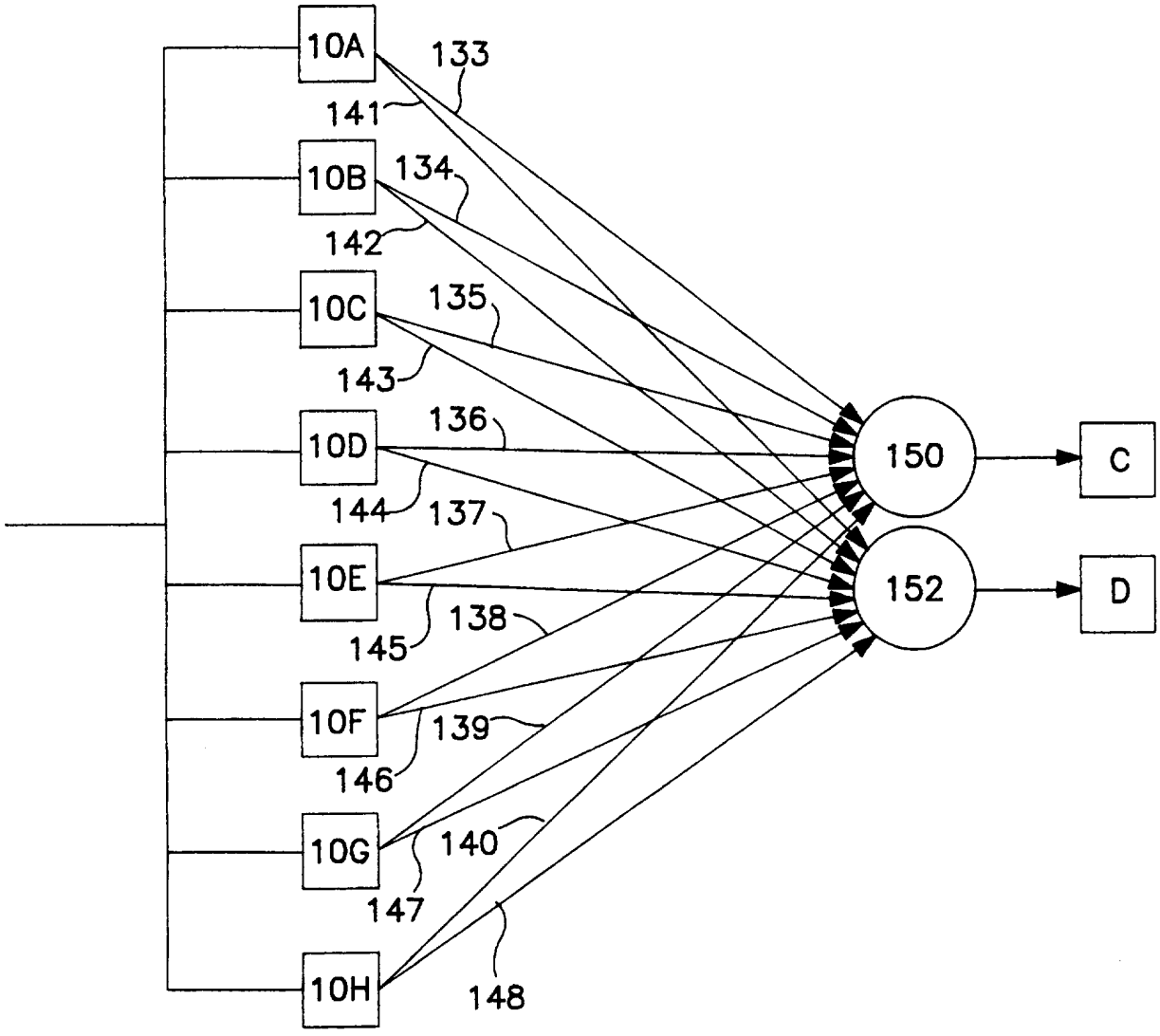

10A
10B
10C
10D
10E
10F
10G
10H
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
150
152
C
D

FIG. 10

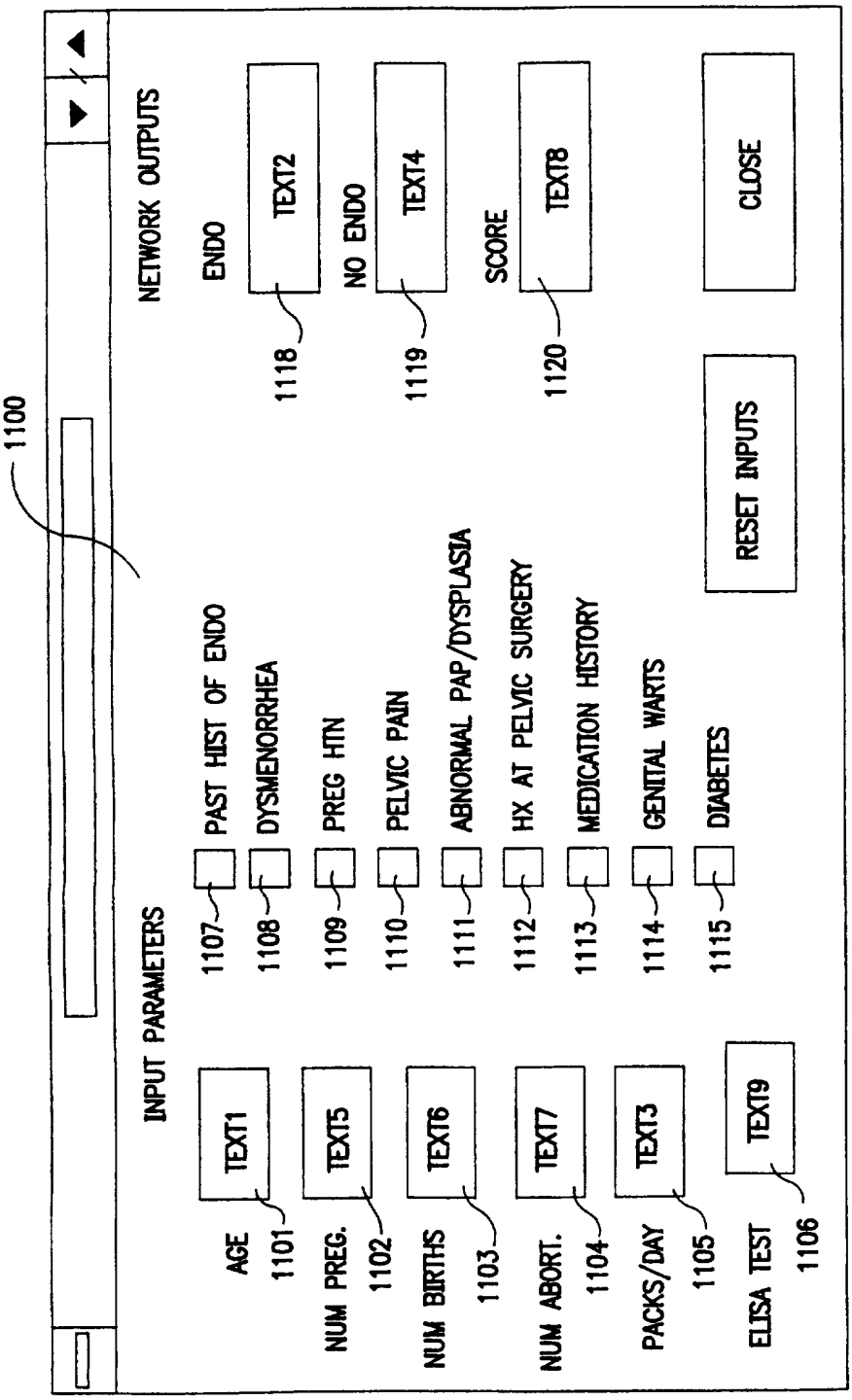

1100
INPUT PARAMETERS
AGE
TEXT1
1101
NUM PREG.
TEXT5
1102
NUM BIRTHS
TEXT6
1103
NUM ABORT.
TEXT7
1104
PACKS/DAY
TEXT3
1105
ELISA TEST
TEXT9
1106
1107 PAST HIST OF ENDO
1108 DYSMENORRHEA
1109 PREG HTN
1110 PELVIC PAIN
1111 ABNORMAL PAP/DYSPLASIA
1112 HX AT PELVIC SURGERY
1113 MEDICATION HISTORY
1114 GENITAL WARTS
1115 DIABETES
NETWORK OUTPUTS
ENDO
TEXT2
1118
NO ENDO
TEXT4
1119
SCORE
TEXT8
1120
RESET INPUTS
CLOSE

FIG. 11