Interpretation of nonstress tests by an artificial neural network

Shahar Kol, MD,[a] Israel Thaler, MD,[a, b] Noam Paz, PhD,[c] and Oded Shmueli, DSc[c]

Haifa, Israel

OBJECTIVE: Our purpose was to evaluate an artificial neural network in the interpretation of nonstress tests.

STUDY DESIGN: A nonlinear artificial neural network trained by backpropagation was taught to interpret records of nonstress tests by two learning sets. The first set contained nonstress tests that were similarly interpreted by three human experts; the second set contained a subset of nonstress tests that led to interobserver disagreement. Both "raw" fetal heart rate and uterine contraction data and 17 quantified variables obtained by automated computer analysis were introduced to the input layer. After training, the network was tested by presenting it with input patterns to which it had not been exposed. The performance of the system was examined in relation to the human expert.

RESULTS: After training the neural network with the first set, a sensitivity of 88.9% and a false-positive rate of 4.3% were obtained at testing. When the learning and test set contained records that led to interobserver disagreement, a sensitivity of 86.7% and a false-positive rate of 19.7% were obtained. Sixty percent of fetal heart rate records interpreted as abnormal by the neural network were interpreted likewise by the human experts.

CONCLUSIONS: The results obtained are encouraging in that the neural network could discriminate between normal and abnormal nonstress tests. Further evaluation of this new technique is mandatory to evaluate its efficacy and reliability in interpreting fetal heart rate records. (AM J OBSTET GYNECOL 1995;172:1372-9.)

Key words: Nonstress test, artificial neural network, backpropagation algorithm, automated analysis of fetal heart rate records

Since its introduction about two decades ago[1, 2] the nonstress test (NST) continues to be the most commonly used modality for the evaluation of fetal status during the antepartum period. Various definitions of reactivity have been used, but no standards were adopted, making it difficult to compare the different results. The American College of Obstetricians and Gynecologists provided some recommendations for defining a reactive NST,[3] but different interpretative algorithms are still used in many centers. In a critical review of the NST[4] it was suggested that the shortcomings of test interpretation based solely on the observation of reactive accelerations could be improved by the use of additional information that is already present in fetal heart rate (FHR) records. Such additional parameters include baseline FHR, beat-to-beat variation, and FHR decelerations.[5, 6] Another potential limitation of NST interpretation is the low interobserver and intraobserver agreement associated with visual assessment of FHR records.[7, 8]

Automated analysis of the NST was recently introduced in an attempt to achieve reproducible, objective analysis of FHR records.[9, 10] Such computer software was programmed to perform calculations and generate conclusions on the basis of the same or on a more extensive analytic array than the human expert is using.

However, although the computer measures events in the NST (e.g., accelerations, decelerations, FHR variation), the human interpreter relies primarily on pattern recognition. This was demonstrated in one study that showed that the age of the observer, years of experience with NST interpretation, and the volume of tests performed added minimal improvement to observer accuracy.[11]

The FHR is a "chaotic" signal and may contain more diagnostic information than is visually apparent. One approach that could detect such subtle but significant differences in the FHR during labor was recently described; this approach used approximate entropy, a mathematic formula quantifying regularity.[12]

Another technique that is capable of pattern recog-

From the Department of Obstetrics and Gynecology, Rambam Medical Center,[a] and the Faculties of Medicine[b] and Computer Sciences,[c] Technion, Israel Institute of Technology.

Received for publication January 31, 1994; revised March 31, 1994; accepted October 12, 1994.

Reprint requests: Israel Thaler, MD, Department of Obstetrics and Gynecology, Rambam Medical Center, Haifa 31096, Israel.

0002-9378/95 $3.00 + 0 6/1/61224

nition, reorganization of data and learning, is a biologic-simulated intelligence or artificial neural network. This novel approach is based on data reduction that resembles the human pattern of thinking. In this study we investigated the applicability of biologically simulated intelligence to the assessment of antenatal FHR records and tested this model against the results obtained by the human expert.

Material and methods

The study population consisted of patients seen at the Rambam Medical Center, Faculty of Medicine, Technion Israel Institute of Technology, either at the antenatal clinic or in the high-risk maternity unit. The gestational age ranged between 36 and 42 weeks at the time of NST recording. FHR and uterine contractions were recorded for 30 minutes with the patient in a semirecumbent position by means of an HP 8040A (Hewlett-Packard, Boeblingen, Germany) external monitor at a paper speed of 3 cm/min. The patients were asked to mark each perceived fetal movement with a handheld button. The recorded data (FHR, uterine contractions, and fetal movements) were sampled into a computer (Olivetti 380/M, Olivetti, Milan) with a digital serial interface. The sampling rate was set to 1 second so that each 30-minute NST generated 1800 values for FHR and 1800 values for uterine contractions. The sampled data of each NST session was stored in a separate file for subsequent analysis. A baseline was fitted with a first-order autoregressive digital filter, which was applied to the recording twice, first in the forward direction and then in the reverse direction, thus eliminating phase shift.[13] An algorithm to prevent excessive excursions caused by large deviations in FHR (e.g., accelerations and decelerations) was also used. A set of programs processed the data and quantified the following variables: baseline FHR, acceleration, decelerations, number of fetal movements, long- and short-term heart rate variability, mean acceleration area, mean deceleration area, the proportion of time FHR acceleration was present during the 30-minute recording (percent acceleration time), and the ratio between the number of FHR decelerations and the number of uterine contractions during the NST session. Accelerations (≥ 10 beats/min for ≥ 15 seconds) and decelerations (≥ 10 beats/min for ≥ 15 seconds) were classified according to amplitude and duration, as described in Table I. The overall heart rate variation was calculated by averaging 1-minute ranges in FHR values about the baseline (mean minute range). Short-term variation was calculated as the mean absolute beat-to-beat difference.[14] Although autocorrelated ultrasonographic pulses were used for this purpose rather than fetal electrocardiographic signals, an excellent correlation exists between the two measurements[15] and actual changes in either direction are detected by the former method (although the actual beat-to-beat variation is diminished). Fetal rest-activity cycles were also determined by the computer on the basis of the presence of high or low heart rate variation. The average signal loss in the records analyzed was 2.4% (43 of 1800 seconds). Whenever missing data were encountered (because of signal loss), a linear interpolation was introduced for correction.

Table I. Quantified NST variables obtained by automated analysis and presented to input layer

Decelerations (10-15 beats/min, 15-30 sec)
Decelerations (10-15 beats/min, >30 sec)
Decelerations (>15 beats/min, 15-30 sec)
Decelerations (>15 beats/min, >30 sec)
Total No. of fetal movements
Overall variation (mean Δ amplitude)
Mean Δ amplitude during high FHR variation
Mean Δ amplitude during low FHR variation
STV, overall
STV during high FHR variation
STV during low FHR variation
No. of FHR decelerations/No. of uterine contractions
Mean deceleration area
Mean acceleration area
Total acceleration duration/monitoring duration
Total No. of FHR accelerations
Baseline FHR

STV, Short-term variability.

Each NST trace was analyzed by visual inspection. Three examiners performed the visual analyses independently. All were practicing obstetricians experienced in interpreting NSTs on a daily basis. We chose to use multiple experts rather than a single expert, because this would better represent a "real life" situation and lead to more consistent results. Thus with this strategy "suspicious" traces (for example, a reactive NST with several short-lived decelerations) would either be considered normal or abnormal according to a "majority vote" and would therefore be less influenced by an individual reviewer's definition and tolerance of abnormal events (e.g., decelerations) in a given trace.

The reviewers were not given clinical information but were advised that all women were in the latter part of the third trimester. They had to state whether the NST tracings were normal (no further evaluation required) or abnormal (further evaluation or immediate intervention mandatory). No specific guidelines were provided, and interpretation was performed in the same manner as in daily practice. The examiners agreed unanimously on 816 NST traces (86%), whereas in 134 (14%) no complete interobserver consensus was reached.

In this study a nonlinear artificial neural network of the feed-foward type was used. This multilayered net-

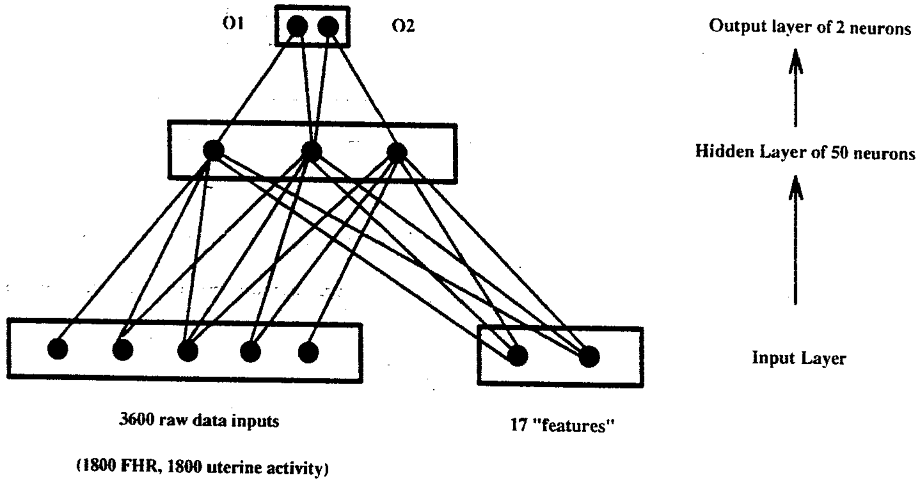

Fig. 1. Schematic representation of neural network architecture. Each of 50 hidden units is connected to 72 raw data inputs and to all 17 computed variables ("features"). Output layer consists of two neurons: *01*, normal result; *02*, abnormal result.

work was trained with backpropagation, which is the emost widely used learning procedure.[16] A major problem was the large number of inputs required at the input layer. This was solved with the aid of URSUS, a neural network simulator, developed in the faculty of computer sciences at the Technion, Israel Institute of Technology. URSUS provides a development and working environment for neural networks, featuring a neural network generator, a neural network simulator, a graphic interface, and a tool kit of useful functions for preprocessing and postprocessing of data. URSUS is user friendly and enables the operator to define and run very large neural networks. It is capable of completing a partial network description provided by the user and offers various extensions for the basic network configuration. It may then train several configurations in parallel and choose the optimal one providing the best results. This considerably reduces the time required for training. In addition, each training cycle is very fast, because the URSUS simulator itself is faster than other simulators we have tried.

The layout of the artificial neural network used in this study is shown in Fig. 1. The input layer is presented with 3600 raw data (1800 FHR and 1800 uterine contraction values) and 17 quantified NST variables generated by the computer (Table I), as previously described. The inclusion of these 17 variables was meant to provide the network with artificially derived "hints," which greatly facilitate learning.

A hidden layer of 50 neurons is connected to the input layer. Each hidden neuron receives input from 72 raw-data neurons and from all 17 computed variables. The hidden layer is connected to an output layer consisting of two neurons. One neuron represents a normal result and the other an abnormal result for a particular NST. A naive postprocessing step was used to interpret the network's output. This step determines (separately for each output unit and repeatedly for every reasonable number of diagnostic misses) where to place the threshold if no more than "*n*" misses of abnormal cases (false negatives) is desirable. For example, if we won't tolerate any false negatives, a high threshold should be considered for the normal output neuron, albeit at the cost of a parallel increase in the false-positive rate. If one miss is to be allowed, the threshold for the normal output neuron should be lowered.

The network was trained by dividing the available data into a training set and a test set. Training took place by choosing input patterns (NSTs) from the training set and allowing activation to flow from the inputs through the hidden units to the output units (feed-forward). The value of the output unit was then compared with the documented diagnosis for each pattern (as provided by the human experts). The difference (error) between the actual activation of the output unit and the correct value was then used by the backpropagation algorithm[16] to modify all weights of the network so that future outputs approximate the correct diagnosis. This step repeats itself until the criteria for convergence are met (i.e., the error ceases to decrease). Fig. 2 shows a typical learning curve. A control group was used to define the cutoff level, the point where optimal learning has been obtained. Testing the network was accomplished by using the weights derived in the training set and presenting the network with input patterns (NSTs) to which it had not been exposed. The performance of the neural net was exam-

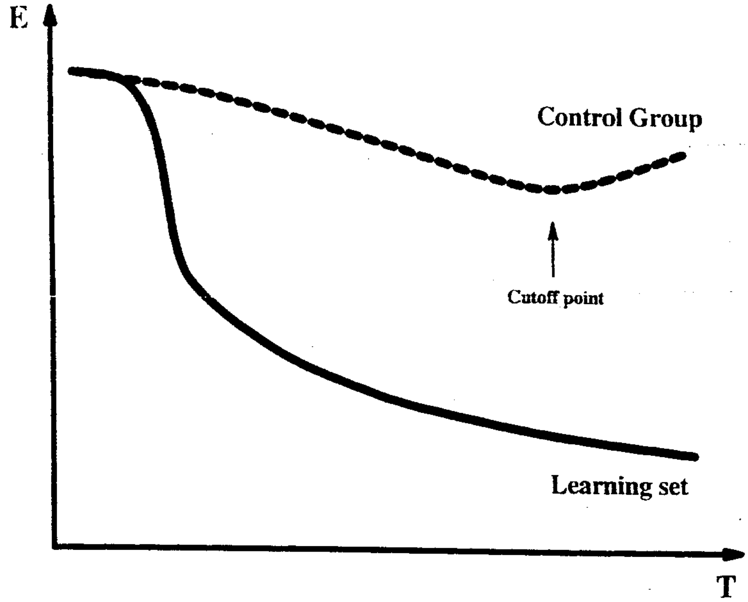

Fig. 2. Plot demonstrating training stage of neural network. *Horizontal axis,* Time *(T)* (or number of iterations); *vertical axis,* error *(E)* (i.e., difference between desired and actual output). *Solid line,* Change in error as function of time. Initially error rapidly declines (i.e., network "learns" fast), and then rate of change is decreased. Separate set of input patterns (control group) is used to define point of minimal error *(broken line).* This is optimal point of convergence, because only marginal improvement is achieved by increasing number of iterations. Network learns by backpropagation rule.

ined in relation to the human expert (the "gold standard"). Detection rate (sensitivity) was defined as the number of NSTs in the test set correctly diagnosed as being abnormal divided by the total number of abnormal NSTs in the test set. False-positive rate (specificity) was defined as the number of NSTs in a test set correctly diagnosed as being reactive divided by the total number of reactive NSTs.

The neural network program was run on a Sun 4/260 SPARCstation (Sun Microsystems, Inc., Mountain View, Calif.).

Validation. To validate the performance of the neural network, we performed the following examinations.

Internal consistency. The neural network was trained with a learning set of 530 NSTs. Each 30-minute NST was divided into six segments of 25 minutes each (i.e., minutes 1 to 25, 2 to 26, 3 to 27, 4 to 28, 5 to 29, 6 to 30). In this manner 3180 segments were obtained.

Another set of 264 NSTs was divided into 25-minute segments as described above. One segment (minutes 2 to 26) served as the control group and another segment (minutes 3 to 27) as the test group. After 2700 cycles the neural network interpreted both the control group and the test group as being normal in 175 traces. Another 53 traces were interpreted as being abnormal in both groups. The neural network judged 18 traces in the control group to be abnormal, whereas the corresponding traces in the test group were judged to be normal. In another 18 traces this disagreement was reversed: 18 traces in the control group that were interpreted as being normal in the control group were interpreted as being abnormal in the test group. The results demonstrate an 86.4% internal consistency. By improving the learning process (e.g., by dividing the traces into longer or shorter segments and by increasing the number of NSTs in the learning and test sets) this figure should be expected to increase.

External consistency. The performance of the neural network was evaluated against the human expert, once for the control group (minutes 2 to 26) and once for the test group (minutes 3 to 27), with the same 264 traces. The results obtained after 2700 cycles were nearly identical, with the neural network giving conflicting results in just two of 264 traces, demonstrating 99.2% consistency. This remained unchanged with advancing exposure to the same input data (tested up to 10,200 cycles).

Results

In the first stage the network was trained by means of a randomly chosen subset of 545 NSTs. All NSTs in this set were similarly interpreted by the three experts (complete interobserver consensus). A total of 510 (93.6%) were interpreted as normal and 35 (6.4%) as abnormal. Learning was accomplished after about 15,000 cycles (i.e., the number of times the network

Table II. Results obtained by testing network with NSTs that were similarly interpreted by all experts ($n = 271$)

Cycle	Threshold	Test set					
		NN	*Sensitivity*	*NA*	*AN*	*AA*	*Specificity*
2,003	0.9080	95	37.55	158	0	18	100.00
2,903	0.9282	104	41.11	149	1	17	94.44
10,708	0.7503	242	95.65	11	2	16	88.89
12,208	0.6593	249	98.42	4	3	15	83.33
12,808	0.5597	250	98.81	3	4	14	77.78
12,808	0.5523	250	98.81	3	5	13	72:22
13,708	0.5250	250	98.81	3	6	12	66.67

NN, NSTs interpreted as normal by both expert and network; *NA,* NSTs interpreted as normal by experts and abnormal by network; *AN,* NSTs interpreted as abnormal by experts and normal by network; *AA,* NSTs interpreted as abnormal by both experts and network.

Table III. Results obtained by testing network with NSTs, some of which did not lead to complete interobserver agreement ($n = 319$)

Cycle	Threshold	Test set					
		NN	*Sensitivity*	*NA*	*AN*	*AA*	*Specificity*
2,000	0.9478	55	19.03	234	0	30	100.00
2,600	0.9248	90	31.14	199	1	29	96.67
2,600	0.8889	143	49.48	146	2	28	93.33
2,900	0.8797	165	57.00	124	3	27	90.00
2,900	0.7990	232	80.28	57	4	26	86.67
2,900	0.7925	234	80.97	55	5	25	83.33
3,802	0.7721	247	85.47	42	6	24	80.00
7,404	0.6704	273	94.46	16	7	23	76.67
8,906	0.6201	276	95.50	13	8	22	73.33
10,406	0.5939	276	95.00	13	9	21	70.00

For abbreviations see footnote to Table II.

evaluated the input data) when the performance of the system stabilized (i.e., the criteria for convergence were met). The closed recognition rate was almost 100%. The network was then tested on 271 NSTs, 253 (93.4%) interpreted as normal and 18 (6.6%) as abnormal by the human experts. Here too there was complete interobserver agreement. The results are shown in Table II. The most important determinant of the system's performance is the value assigned to the subset *AN* (abnormal diagnosed as normal) in the test set, which defines how many false-negative cases are acceptable. This value depends on the threshold for the normal-output neuron (the likelihood of normal NST). If a high threshold was used (e.g., 0.908), the sensitivity was 100%, but there was an overhead of a high false-positive rate (62.5%). By decreasing the threshold to 0.75, the sensitivity decreased to 88.9% (two false-negative results were introduced), whereas the false-positive rate decreased to 4.3% (the system correctly diagnosed 242 of 253 normal NSTs). Lowering the threshold further resulted in improved specificity, but the number of false negatives became unacceptably high. The abnormal neuron gave similar results.

In the second stage the network was trained and tested on 950 NSTs, 816 from the first stage and an additional 134 for which the interpretation did not lead to complete interobserver consensus. Each of these NSTs was assigned an interpretation according to the majority decision (i.e., two of three). Training was performed on a subset of 631 NSTs, 571 (90.5%) interpreted as normal and 60 (9.5%) as abnormal by the human experts. The test set included 319 NSTs, 289 (90.6%) interpreted as normal and 30 (9.4%) as abnormal by the human experts. Here too learning was completed after about 15,000 cycles, when the closed recognition rate was almost 100%. Table III shows the results of the test. On the basis of the normal neuron (01), a threshold of 0.8 (cycle 3200) yielded the best result, specificity 80.3% and sensitivity 86.7%. There were four false-negative cases.

Comment

Artificial neural networks are computer models inspired by the structure and behavior of real neurons. Like the brain, they can recognize patterns, reorganized data, and, most interesting, learn. Artificial neural networks are typically composed of interconnected units that serve as model neurons. The units are connected by links that act as the axons and dendrites. The function of the synapse is modeled by modifiable weight, which is associated with each connection, representing the connection strength at the synapse. The link passes the weighted output value to another unit, which sums up the values passed to it by all other incoming links. If the total input value exceeds some threshold value, the unit fires. Modifications in the firing pattern constitute the learning. This is achieved by changes in the weighting factors on the links and is analogous to changing the effectiveness of the synapses in the brain so that the influence of one neuron on another changes.[17]

Artificial neural networks are made up of three types of units. The input units take in information from the outside. The output units send out signals that are visible to the external world. The hidden units act as a go-between from the input to the output unit. They neither receive input directly from the outside nor produce a visible output. To develop a predictive model, the network analyzes multiple parameters presented at the input as the training data set, with the number of variables corresponding to the number of input neurons. The input neurons are connected to the neurons in the hidden layer. Similarly, the neurons in the hidden layer are connected to an output layer, the results to be modeled. The entire data set is run through the network numerous times, with the network attempting to vary the interconnections between the input-hidden layers and the hidden-output layers until all inputs match all outputs (i.e., a specified outcome is obtained for each input). The model is then ready for use as a predictive tool.

In this study the neural network was efficiently trained by both training sets, with a diagnostic closure of nearly 100%. If the expert's interpretations are taken as the "gold standard," the neural network produced good resuls for both test sets. In the "clear-cut" experiment the detection rate was 89% and the false-alarm rate was 4.35% when the threshold was set at 0.75. When the training set and the test set contained NSTs that were interpreted by one expert different from the other two human analysts, the detection rate was 87% and the false-alarm rate was 19.7%, with a threshold of 0.8 and 77% and 5.5%, respectively, with a threshold of 0.67. These results demonstrate better agreement between the neural network and the human analyst than was shown for automated FHR analysis. In a comparison between computerized FHR analysis and visual NST interpretation, only 44.4% of nonreactive tests did not meet the required computer criteria for a normal test, whereas 34% of tests that did not meet those criteria were nonreactive. The false-alarm rate was 6.9%.[18] Only cases with a priori interobserver agreement were compared. In the current study 60% of NSTs interpreted as abnormal by the neural network were similarly interpreted by the experts in the two data sets tested. It should be pointed out that in the former study the diagnostic performance of the computer was similar to the performance of the human observer in relation to fetal outcome, and there was a highly significant decrease in the intervention rate when computerized interpretation was used as opposed to visual assessment.[18] In this study no attempt was made to examine the performance of the neural network in relation to fetal outcome. The latter may be influenced by many factors, which are not necessarily reflected by an NST performed at various intervals from delivery. Also the test does not simulate the risk conditions of labor. These and other factors account for the relatively low sensitivity and positive predictive value of the NST.[19]

Poor agreement between visual interpretation and automated NST analysis was demonstrated in two other studies.[20, 21]

Unlike expert systems to which knowledge is explicitly provided, neural networks define their own rules (or decision criteria) in an implicit manner by modifying the weights of each connection so that the network produces a better approximation of the desired output. Because knowledge is distributed over the entire network, the latter is tolerant to both subtle distortions of weights and to incomplete or noisy inputs. These characteristics not only provide the network with its credibility, but also with its most important feature: generalization, the power to generate a reasonable output to an input that was not present in the training series. This fundamental characteristic that enables the network to function more accurately in the clinical environment is lacking when other statistical strategies are used.

Previous clinical experience with neural networks demonstrated encouraging results. Fifteen input variables were used to predict the weight of fetuses suspected of having macrosomia,[22] and 41 variables were used in an attempt to diagnose myocardial infarction in patients arriving at the emergency department with chest pain.[23] The networks performed substantially better than reported for physicians or for any other analytic approach (e.g., logistic models).

Unlike the two previous studies,[22, 23] where outcome was indisputable (newborn's weight or clinical diagnosis of myocardial infarction), the end point in the current study is less clearly defined. Visual interpretation of the NST is subjected to substantial interobserver and intraobserver variation.[7, 8, 21]

Reviewers of NST records are often asked to rate the NST explicitly (Is the trace reactive [normal], suspicious [questionable], nonreactive [pathologic], or implicit [Is further evaluation or intervention mandatory?]). We felt that by asking the experts to make management decisions we simulated a familiar situation to which they were exposed in daily practice. To eliminate interobserver variation, we included in the "clear-cut" learning and test sets only traces that were interpreted similarly by all reviewers. When NSTs were included that did not have complete interobserver consensus for their interpretation, there was less agreement between the network and the human raters. Whether this was because of a more accurate interpretation of NSTs by the network, which identified and used new information, remains to be established. It should be pointed out that the NST features are a function of gestational age. As the fetus matures, its sleep cycles show less variability and the awake phase produces more dramatic patterns than in the younger fetus. Other behavior patterns (e.g., sucking) will likely produce differing interpretations by the experts, which may confuse the network.

This study represents the first reported attempt to use a very large number of input variables in the clinical setting. In a preliminary experiment we have presented the input layer with either raw FHR data or with 17 quantified FHR variables (which were generated by a computer). The network did not demonstrate optimal convergence, and the diagnostic closure was not as satisfactory as when both inputs were used in parallel. If the neural network could receive the output as part of the input, the learning procedure would have been an easy task. This, however, is not feasible, so by introducing at the input information that is derived from the basic raw data ("hints") we can use a smaller network and substantially shorten the learning procedure. A large enough network can approximate any reasonable nonpolynomial function.[24, 25] Theoretically then learning could also have been accomplished by means of the raw data alone. However, this would require a very large and fast network and an extremely fast computer, making it virtually impractical.

The ability of the network to discover relationships in the data that may not be immediately apparent to physicians, its resistance to substantial input perturbations (as is often the case with "noisy" FHR records), and its ability to improve by learning new input data are the primary advantages of the biologically simulated intelligence. In this study we evaluated the ability of the neural network to discriminate between normal and abnormal NSTs. The results obtained are encouraging and would justify further investigations to evaluate the efficacy and reliability of this new technique in interpreting FHR records.

REFERENCES

1. Hammacher K. The clinical significance of cardiotocography. In: Huntingford PJ, Hunter KA, Saling E, eds. Perinatal medicine. New York: Academic Press, 1969:80-9.
2. Rochard F, Schifrin BS, Goupil F, et al. Nonstressed fetal heart rate monitoring in the antepartum period. Am J Obstet Gynecol 1976;126:699-706.
3. American College of Obstetricians and Gynecologists. Antepartum fetal surveillance. Washington, DC: American College of Obstetricians and Gynecologists, 1987:1-4. (American College of Obstetricians and Gynecologists technical bulletin no 107.)
4. Devoe LD, Castillo RA, Sherline DM. The nonstress test as a diagnostic test. A critical reappraisal. Am J Obstet Gynecol 1985;152:1047-53.
5. Sill H, Wilson EW. Antepartum fetal heart rate monitoring. Aust N Z J Obstet Gynaecol 1975;15:132-7.
6. Devoe LD, Yanowitch G, Azor H. The application of multiple-parameter scoring to antepartum fetal heart rate testing. J Reprod Med 1981;26:250-4.
7. Lotgering FK, Wallenburg HCS, Schouten HJA. Interobserver and intraobserver variation in the assessment of antepartum cardiotocograms. Am J Obstet Gynecol 1982;144:701-5.
8. Trimbos JR, Keirse MJ. Observer variability in assessment of antepartum cardiotocograms. Br J Obstet Gynaecol 1978;85:900-6.
9. Dawes GS, Moulden M, Redman CWG. The advantages of computerized fetal heart rate analysis. J Perinat Med 1991;19:39-45.
10. Searle JR, Devoe LD, Phillips MC, Searle NS. Computerized analysis of resting fetal heart rate tracings. Obstet Gynecol 1988;71:407-12.
11. Hage ML. Interpretation of nonstress tests. Am J Obstet Gynecol 1985;153:490-5.
12. Pincus SM, Viscarello R. Approximate entropy: a regularity measure of fetal heart rate analysis. Obstet Gynecol 1992;79:249-55.
13. Orr WC, Hoffman HJ. A 90-min cardiac biorythms: methodology and data analysis using modified periodograms and complex demodulation. IEEE Trans Biomed Eng 1974;21:130-42.
14. Dalton KJ, Dawes GS, Patrick JE. Diurnal, respiratory, and other rhythms of the fetal heart rate in lambs. Am J Obstet Gynecol 1977;127:414-24.
15. Lawson GW, Belcher R, Dawes GS, Redman CWG. A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems. Am J Obstet Gynecol 1983;147:721-2.
16. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by backpropagation errors. Nature 1986;323:533-6.
17. Hinton GE. Connectionist learning procedures. Artif Intell 1989;40:185-234.
18. Hiett AK, Devoe LD, Youssef A, Gardner RN, Black BS. A comparison of visual and automated methods of analyzing fetal heart rate tests. Am J Obstet Gynecol 1993;168:1517-21.
19. Thacker SB, Berkelman RL. Assessing the diagnostic accuracy and efficacy of selected antepartum fetal surveillance techniques. Obstet Gynecol Surv 1986;41:121-41.
20. Gagnon R, Campbell K, Hunse C. A comparison between visual and computer analysis of antepartum fetal heart rate tracings. Am J Obstet Gynecol 1993;168:842-7.
21. Schneider E, Schulman H, Farmakides G, Paksima S. Comparison of the interpretation of antepartum fetal heart rate tracings between a computer program and experts. J Matern Fetal Invest 1991;1:205-8.
22. Farmer RM, Medearis A, Hirata GI, Platt LD. The use of a neural network for the ultrasonographic estimation of fetal weight in the macrosomic fetus. Am J Obstet Gynecol 1992;166:1467-72.
23. Baxt WG. Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion. Neural Computation 1990;2:480-9.

24. Cybenko G. Approximation by superpositions of sigmoidal function. Math Control Signals Systems 1989;2:303-14.
25. Kolmogorov AN. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Dokl Akad Nauk 1957;114:953-6.