Ultrasound in Med. & Biol., Vol. 28, No. 8, pp. 991–996, 2002

0301-5629/02/$–see front matter

PII: S0301-5629(02)00554-9

● *Original Contribution*

ULTRASOUND ESTIMATION OF FETAL WEIGHT WITH THE USE OF COMPUTERIZED ARTIFICIAL NEURAL NETWORK MODEL

LOUISE CHUANG,* JENG-YANG HWANG,† CHIUNG-HSIN CHANG,* CHEN-HSIANG YU* and FONG-MING CHANG*

*Department of Obstetrics and Gynecology, National Cheng Kung University Medical College and Hospital, Tainan, Taiwan; and †Department of Information Engineering, I-Shou University, Kaohsiung, Taiwan

(*Received* 5 *November* 2001; *in final form* 9 *May* 2002)

Abstract—The aim of this study was to test if the computerized artificial neural network (ANN) model could improve ultrasound (US) estimation of fetal weight over estimation with the other commonly used formulas generated from regression analysis. First, as the training group, we performed US examinations on 991 singleton fetuses within 3 days of delivery. Six input variables were used to construct the ANN model: biparietal diameter (BPD), occipitofrontal diameter (OFD), abdominal circumference (AC), femur length (FL), gestational age and fetal presentation. Second, a total of 362 fetuses were assessed subsequently as the validation group. In this training group, the ANN model was better than the other compared formulas in fetal weight estimation ($n = 991$, mean absolute error 183.83 g, mean absolute percent error 6.02%, all $p < 0.0001$). In addition, the validation group further proved the results ($n = 362$, mean absolute error 179.91 g, mean absolute percent error 6.15%, all $p < 0.005$). In conclusion, the computerized artificial neural network (ANN) model could provide better US estimation of fetal weight than estimations by means of commonly used formulas generated from regression analysis. (E-mail: fchang@mail.ncku.edu.tw) © **2002 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Ultrasound, Fetal weight estimation, Artificial neural network.**

INTRODUCTION

Estimation of fetal weight is very important in prenatal care, and may be relevant to decision making and management (Chien et al. 2000). To date, ultrasound (US) remains the major tool for estimating fetal weight. In the literature, most published formulas for fetal weight estimation were generated from 2-D US measurements, such as biparietal diameter (BPD), occipitofrontal diameter (OFD), abdominal circumference (AC) and femur length (FL), with regression analysis (Hadlock et al. 1985; Hsieh et al. 1987; Warsof et al. 1977). Recently, 3-D US formulas generated from fetal thigh or upper arm volumes with regression analysis were reported to achieve higher level of accuracy (Chang et al. 1997; Liang et al. 1997). But, the disadvantages of 3-D US formulas remain to be overcome, such as time-consuming, method-sophisticated and cost-expensive.

Artificial neural network (ANN), a computerized analog of a biologic neural system, has been widely used in many different professional fields (Baxt 1995; Baxt and Skora 1996; Bottaci et al. 1997; Dybowski and Gant 1995; Farmer et al. 1992; Kothari et al. 1996; Park et al. 1996; Walker et al. 1999). The application of ANN to estimate fetal weight was first reported by Farmer and coworkers. Their result showed that ANN was superior to traditional formulas in fetal weight estimation in the macrosomia group. However, their subjects were limited to a specific group and no subsequent studies further proved their conclusion. The constructed architecture of the ANN model would develop relationships between the input and output data when training proceeds. The way of training the ANN model simulates how biologic neural connections are established and rectified perpetually. After an appropriate training process, the nonlinear neural network can afford a best fit "guess" as a result. The architecture, principles, characteristics and applications have been discussed in the literature (Cross et al. 1995).

Although the 2-D US measurement methods were strictly followed (Callen 1983; Campbell and Wilkin 1975; Ko et al. 1983), there were errors between estimated fetal weight and actual fetal weight. Instead of the

Address correspondence to: Fong-Ming Chang, M.D., Ph.D., Department of Obstetrics and Gynecology, National Cheng Kung University Hospital, 138 Victory Road, Tainan, Taiwan. E-mail: fchang@mail.ncku.edu.tw

fixed formulas using regression analysis, we hypothesized that the ANN model that develops nonlinear relationships between input variables and output outcomes could reduce the errors. In this report, we incorporated 2-D US parameters, fetal presentation and gestational age together, to construct the ANN model in estimating fetal weight. By using the ANN model, we attempted to establish an accurate and convenient way for estimating fetal weight, which could preserve the advantages of 2-D US formulas and avoid the disadvantages of 3-D US. The performance of our ANN model and the traditional 2-D US formulas were compared. To the best of our knowledge, our report is the first well-established ANN model in predicting fetal weight covering the entire range of fetal weight.

MATERIALS AND METHODS

Subjects

From November 1997 to June 1999, we consecutively performed fetal biometry by US on every singleton fetus admitted to the delivery room at the National Cheng Kung University Hospital. The inclusion criteria were singletons delivered within 3 days of US examination. The exclusion criteria were anomalous fetuses, multiple gestations and fetuses not delivered within 3 days of US examination. A total of 991 fetuses that met the above criteria were included as the training group of the ANN model.

To further validate the established ANN model in fetal weight estimation, 362 fetuses that were delivered within the subsequent 6 months (from July 1999 to December 1999) and met the criteria described above were used as the validation group.

ANN model

The architecture of the trained ANN (Fig. 1a) model in our report was composed of three layers: 1. one input layer with six inputs; 2. one hidden layer with 12 neurons and 72 connections; 3. one output layer with one outcome and 12 connections. In this study, the back propagation network algorithm is used as the learning algorithm to train the artificial neural network (Cross et al. 1995). The ANN model development flow chart in our report is shown in Fig. 1b. Six input variables were BPD, OFD, AC, FL, gestational age (GA) and fetal presentation. The AC was calculated from the abdominal anteroposterior diameter (APD) and abdominal transverse diameter (ATD) with the equation: $AC = \pi/2 \times (APD + ATD)$.

All the US measurements were conducted by trained resident doctors, according to the methods previously described (Callen 1983; Campbell and Wilkin 1975; Ko et al. 1983). We used the commercially available 2-D US scanners (Aloka SSD-650 or SSD-680, Tokyo, Japan) with a 3.5-MHz transabdominal probe. The gestational ages (input unit: weeks) were determined by the last menstrual period and were confirmed by US at early pregnancy. Fetal presentations were divided into vertex and malpresentation groups according to US examination. The fetuses with breech or transverse presentations were represented in the group of malpresentations. Number 1 and 0 were used to indicate malpresentation and vertex presentation, respectively, as the sixth input variable. All ANN programs were programmed by

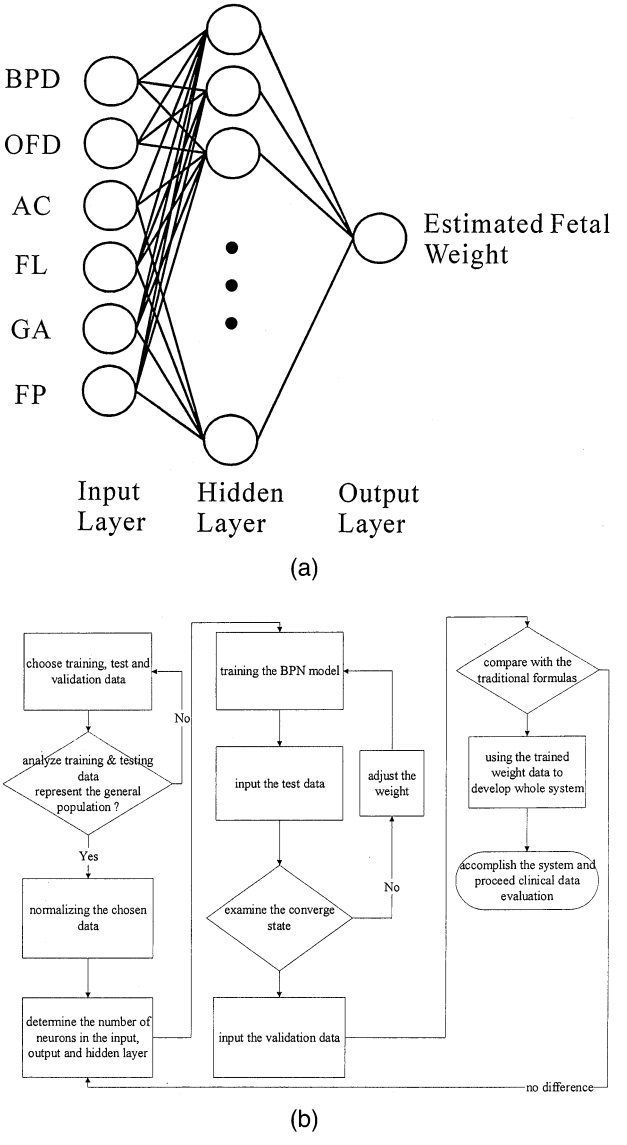

Fig. 1. (a) The architecture of ANN model. BPD = biparietal distance, OFD = occipital frontal distance, AC = abdominal circumference, FL = femur length, GA = gestational age, FP = fetal presentation. (b) The neural network development flow chart.

Table 1. Clinical widely-used US formulas for estimation of fetal weight by regression analysis

Authors	Ultrasound formulas by regression analysis
Hsieh et al. (1987) formula 1B	$\text{Log}_{10}BW = 5.6541 \cdot 0.001 \cdot AC \cdot BPD - 1.5515 \cdot 0.0001 \cdot AC^2 \cdot BPD + 1.9782 \cdot 0.00001 \cdot AC^3 + 5.2594 \cdot 0.01 \cdot BPD + 2.13153$
Hsieh et al. (1987) formula 2B	$\text{Log}_{10}BW = 9.4962 \cdot 0.001 \cdot AC \cdot BPD - 0.1432 \cdot FL - 7.6742 \cdot 0.0001 \cdot AC \cdot BPD^2 + 1.7450 \cdot 0.001 \cdot BPD^2 \cdot FL + 2.7193$
Warsof et al. (1977)	$\text{Log}_{10}BW = -1.599 + 0.144 \cdot BPD + 0.032 \cdot AC - 0.000111 \cdot (BPD^2 \cdot AC)$
Hadlock et al. (1985)	$\text{Log}_{10}BW = 1.304 + 0.05281 \cdot AC + 0.1938 \cdot FL - 0.004 \cdot AC \cdot FL$

BW = birth weight; *AC* = abdominal circumference; *BPD* = biparietal diameter; *FL* = femur length.

C^{++} language and compiled with Inspire Borland C^{++} Builder version 4.0 (Scotts Valley, CA). The training method is similar to that in the report of Cross et al. (1995).

Statistics

Several traditional formulas used in Taiwan (Hsieh et al. 1987) and in the western countries (Hadlock et al. 1985; Warsof et al. 1977) are listed in Table 1. These formulas were all generated by regression analysis. To examine the performance of the ANN model, we compared the differences of the mean absolute error and mean absolute percent error between the ANN model and the traditional formulas by paired *t*-tests.

RESULTS

The distributions of actual fetal body weight of the training group and the validation group are shown in Fig. 2. The average fetal body weights were 3108.2 ± 532.5 g (200 to 4400 g) and 3014.6 ± 669.7 g (275 to 4340 g) in the training group and the validation group, respectively. The estimated fetal weight by ANN model showed no significant difference from the actual fetal weight either in the training group ($n = 991$, $t = 1.88$, $p = 0.06$, not significant) or in the validation group ($n = 362$, $t = 0.55$, $p = 0.58$, not significant). The highly significant correlations between the actual fetal weight and the estimated fetal weight by the ANN model of the training group and the validation group are shown in Fig. 3a ($r = 0.9027$, $n = 991$, $p < 0.0001$) and Fig. 3b ($r = 0.9379$, $n = 362$, $p < 0.0001$).

As shown in Table 2, we compared the mean values of the absolute error (AE) and absolute percent error (APE) by the ANN model with those by the traditional formulas in the training group. The ANN model had the lowest values of mean absolute error (183.83 g, $n = 991$, all $p < 0.0001$) and mean absolute percent error (6.02%, $n = 991$, all $p < 0.0001$). In the validation group (Table 3), the lowest mean absolute error (179.91 g, $n = 362$, all $p < 0.005$) and mean absolute percent error (6.15%, $n =$

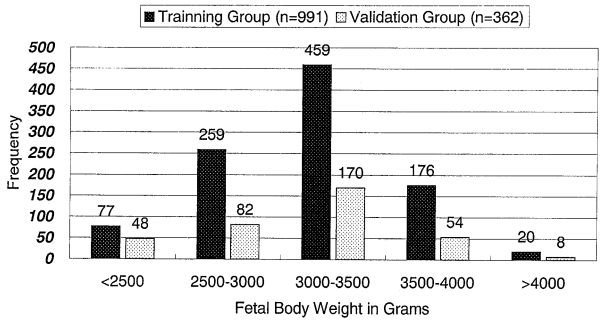

Fig. 2. The actual fetal body weight distribution of the training group and the validation group.

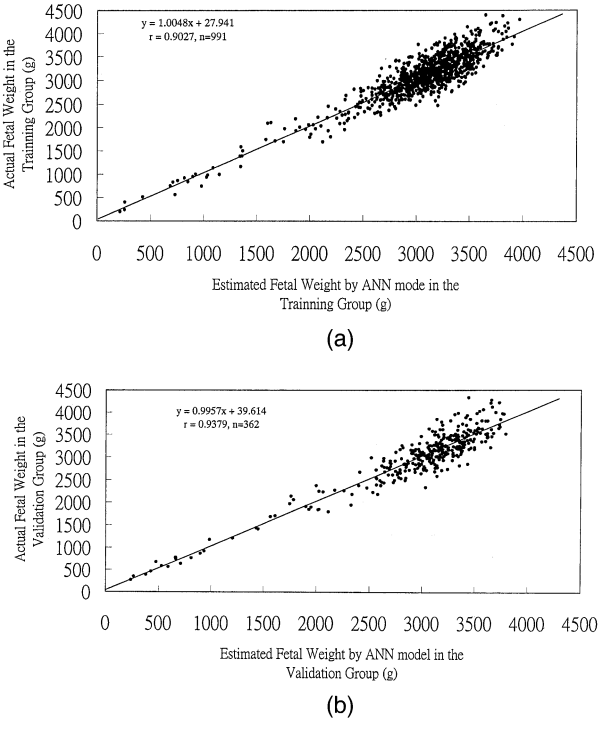

Fig. 3. (a) Scattergram of the actual fetal body weight vs. the estimated fetal weight by the ANN model in the training group. (b) Scattergram of the actual fetal body weight vs. the estimated fetal weight by the ANN model in the validation group.

Table 2. Comparison of accuracy of different formulas in the training group ($n = 991$)

Formula	*APE* (%, mean ± SD)	Significance* (p)	*AE* (g, mean ± SD)	Significance* (p)
Hsieh et al. (1987) formula 1B	7.3% ± 6.62%	<0.0001	216.92 ± 162.76	<0.0001
Hsieh et al. (1987) formula 2B	7.65% ± 10.09%	<0.0001	216.81 ± 160.47	<0.0001
Warsof et al. (1977)	7.98% ± 5.86%	<0.0001	246.60 ± 183.23	<0.0001
Hadlock et al. (1985)	7.74% ± 6.08%	<0.0001	238.34 ± 190.06	<0.0001
ANN model	6.02% ± 4.72%	–	183.83 ± 143.09	–

* Paired t-test, ANN model used as a comparison; *APE* = absolute percent error; *AE* = absolute error

362, all $p < 0.005$) of the ANN model further proved the statistical superiority. The result with the ANN model with fetuses at extremes of weight range (< 2500 g and > 4000 g) was less accurate both in the training group (< 2500 g APE: 10.25%, > 4000 g APE: 10.71%), and in the validation group (< 2500 g APE: 9.63%, > 4000 g, APE: 13.03%).

DISCUSSION

The ANN model has been widely used in many different professional fields, such as pattern recognition, classification problems, image analysis and prediction model construction (Baxt 1995; Baxt and Skora 1996; Bottaci et al. 1997; Dybowski and Gant 1995; Farmer et al. 1992; Kothari et al. 1996; Park et al. 1996; Walker et al. 1999), etc. The back propagation network (BPN) algorithm is one of the most useful learning algorithms in ANN models (Cross et al. 1995). It is a supervised training algorithm that performs fixed gradient descent in weights, to reduce the mean squared error between the target outputs and network actual outputs. After training (Cross et al. 1995), best fit results can be achieved by the ANN. Unlike a linear analysis, neural networks are nonlinear and no calculation about the sample size needed to achieve a given level of confidence is yet possible (Cross et al. 1995). In this study, we used the BPN algorithm to develop the ANN and prove our hypothesis that the ANN model could reduce errors between estimated fetal weight and actual fetal weight.

Farmer et al. (1992) reported the first study using ANN to estimate fetal weight by US. However, their study was focused mainly to the macrosomia group. No subsequent study further proved the conclusion and no similar study covered the entire weight range of fetal weight. The subjects in our series were collected from our delivery room, which included actual fetal weight from 200 g to 4400 g. It is our routine that all fetuses accepted US examination when admitted to the delivery room. Our report showed that the ANN model could reach a statistically higher degree of accuracy than the traditional formulas using regression analysis in fetal weight estimation, even when applied to the entire weight range.

In our series, the ANN performance was less accurate at the extremes of the weight range, especially when actual fetal body weight was < 2500 g or > 4000 g. A reason for the lower accuracy might be fewer cases of extreme fetal weight. Collecting enough cases of extreme fetal weight and developing ANN models for the weight groups of < 2500 g and > 4000 g, respectively, may

Table 3. Comparison of accuracy of different formulas in the validation group ($n = 362$)

Formula	*APE* (%, mean ± SD)	Significance* (p)	*AE* (g, mean ± SD)	Significance* (p)
Hsieh et al. (1987) formula 1B	7.85% ± 6.52%	<0.0001	218.87 ± 158.68	<0.001
Hsieh et al. (1987) formula 2B	8.53% ± 10.82%	<0.0002	217.01 ± 154.76	<0.002
Warsof et al. (1977)	7.86% ± 5.79%	<0.0001	235.07 ± 174.57	<0.0001
Hadlock et al. (1985)	7.54% ± 6.44%	<0.002	217.87 ± 183.93	<0.003
ANN model	6.15% ± 4.99%	–	179.91 ± 148.99	–

* Paired t-test, ANN model used as a comparison; *APE* = absolute percent error; *AE* = absolute error.

solve this problem, and is being undertaken at our department. In addition, our ANN model could not be applied to fetuses for which 2-D US data were unavailable. When fetal heads were deeply engaged in the pelvis or when fetal bodies were severely compressed in oligohydramnios, 2-D US fetal-weight estimation formulas might be the only effective way (Chang et al. 1997) to estimate fetal weight correctly.

Some critics may wonder because it is likely that, over the 18 months of the initial study, many more than 991 deliveries would have taken place in a major obstetric hospital like ours. Although we are a tertiary medical center in Southern Taiwan, we are not the major obstetric hospital compared with the other community hospitals nearby, in which 2500 to 3000 babies were delivered per year. In our hospital, we delivered around 1400 to 1500 babies per year. In the phase one study, from November 1997 to June 1999, we had 2501 babies delivered in our delivery room. And our US unit only operated in the daytime (from 9:00 AM to 5:00 PM). Those admitted to our hospital in the evening and in the night were missed for US examination. Therefore, it is reasonable to have 40% (991 of 2501) babies examined by prenatal US during the daytime.

In our study, the definition of an anomaly was for any fetus with a major structural anomaly that could be diagnosed prenatally, such as holoprosencephaly, omphalocele, cystic hygroma, etc. These were excluded from the study. The rationale was that the traditional 2-D US measurements are not appropriate for the assessment of fetal biometry in fetuses with structural anomalies. In our study, we might include some fetuses with rare and nonstructural anomalies that could only be diagnosed postnatally by genetic screening or metabolic methods, in which the prenatal ultrasonic examination cannot demonstrate any structural abnormality. However, we believe that this point makes only little impact on the study because these nonstructural anomalies are too rare.

Although there is no universally acknowledged rule, it would be useful to know why 991 (approximately three quarters) cases were selected for the training group and the remaining 362 cases as the validating group, rather than some other division. In our phase one study (November 1997 to June 1999), it was planned initially to have around 1000 babies as the training group before the study was undertaken. Therefore, 991 babies were included for the training group. In the phase two study (July 1999 to December 1999), our statistical consultants advised that we already had enough samples to reach a solid conclusion for our initial study goal. In other words, in our phase two study, they considered 362 cases already sufficient to have a statistical power that could prove our hypothesis that the ANN model is superior to other conventional 2-D regression methods. It seems redundant to have more cases in 2000 to reach our conclusion.

Some may wonder at our choice of the six input parameters, thinking that they are not well justified. The four dimensional variables are reasonable because of the previous literature (Hadlock et al. 1985; Hsieh et al. 1987; Warsof et al. 1977), but gestational age (GA) and presentation are less obvious. First, according to Sabbagha et al. (1989), GA is very important to the prediction of fetal weight because the density and/or the morphological features of the fetuses may change with different GA. Therefore, it is not our hypothesis only. It is already proved by their study that GA is an important factor in predicting fetal weight. Moreover, they had included GA in their formula of fetal weight prediction (Sabbagha et al. 1989), as we did in this series.

Second, why should presentation influence the accuracy of the estimation? It is well documented that, in the breech or the transverse position, the fetal head is subjected to greater side-to side pressure than in the vertex presentation (Hadlock et al. 1985; Hsieh et al. 1987). Thus, the shape of the fetal head could be affected by malpresentation so as to appear in an abnormal shape, such as dolicocephaly or brachycephaly. In the literature, there are formulas to correct BPD or to use head circumference for solving this situation (Hadlock et al. 1985; Sabbagha et al.1989). In this series, we did not use them to modify our data, because we believe that our ANN model could learn by making presentation as one input variable and then turn out to be an appropriate output variable. And our results proved that fetal presentation is a useful parameter in predicting fetal weight. However, further studies are warranted to show the exact impact of fetal presentation on fetal weight prediction.

Some may question the reproducibility of our study using the previous formulas by Warsof et al. (1977) or Hadlock et al. (1985), and that it will create a bias in the control group to show the superiority of the ANN model. Indeed, it is very important that we should be able to reproduce the same level of accuracy for the Hadlcok or the Warsof formulas when using them as the control group. The accuracy reported by Warsof and colleagues was 8.0% in the mean absolute percent error. The accuracy reported by Hadlock was 7.5%. In our study, as shown in Tables 2 and 3, we reproduced the APE as 7.98% in the training group and 7.86% in the validation group for the Warsof formula, respectively; both of them are very close to the data in the Warsof report (8.0%). Furthermore, we reproduced the APE as 7.74% in the training group and 7.54% in the validation group for the Hadlock formula, respectively; both of them are also very close to that in the Hadlock report (7.5%). Therefore, we reproduced the same level of accuracy using

their formulas in our study and, thus, can avoid bias when using them as the control group.

Finally, in Figs. 3a and b, it seems that the data regarding the large and small babies do not lie along a straight line and, therefore, the errors at both ends might be systemic. We consider the small case numbers in both the extreme fetal weight groups might be the main cause of the error in predicting fetal weight, and studies of larger sample size are needed to reduce this error in the future. On the other hand, other possible causes for the errors at both ends might be due to the density difference, or the variation of fetal morphologic feature, or the variety in body configurations, as suggested by Sabbagha et al. (1989). However, further studies are warranted to clarify these issues.

In conclusion, our study demonstrates that the ANN model can be used to estimate fetal weight covering the entire weight range, with statistical superiority to the traditional formulas. However, the accuracy of fetal weight estimation at the extremes of the weight range is not satisfactory. Further studies are needed to test the ANN model in fetal weight estimation of extreme weight ranges.

Acknowledgements—This study was supported in part by grants from National Science Council, Taipei, Taiwan, to F. M. Chang. The authors are grateful to all the resident doctors for the ultrasound measurements; Ms. Wen-Chu Chen, Ms. Yeh-Chin Cheng and Ms. I-Chan Wang for their assistance; and T. K. Truong, Fellow IEEE, at the Department of Information Engineering, I-Shou University, for equipment supply.

REFERENCES

Baxt WG. Application of artificial neural networks to clinical medicine. Lancet 1995;346:1135–1138.

Baxt WG, Skora J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. Lancet 1996;347: 12–15.

Bottaci L, Drew PJ, Hartley JE, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet 1997;350:469–472.

Callen PW. Ultrasonography in obstetrics and gynecology. Philadelphia: Saunders, 1983:25–33.

Campbell S, Wilkin D. Ultrasonic measurement of the fetal abdominal circumference in estimation of fetal weight. Br J Obstet Gynaecol 1975;82:689–697.

Chang FM, Liang RI, Ko HC, Yao BL, Chang CH, Yu CH. Three-dimensional ultrasound-assessed fetal thigh volumetry in predicting birth weight. Obstet Gynecol 1997;90:331–339.

Chien PFW, Owen P, Khan KS. Validity of ultrasound estimation of fetal weight. Obstet Gynecol 2000;95:856–860.

Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet 1995;346:1075–1079.

Dybowski R, Gant V. Artificial neural networks in pathology and medical laboratories. Lancet 1995;346:1203–1207.

Farmer RM, Medearis AL, Hirata GI, Platt LD. The use of a neural network for the ultrasound estimation of fetal weight in the macrosomic fetus. Am J Obstet Gynecol 1992;166:1467–1472.

Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. Am J Obstet Gynecol 1985;151:333–337.

Hsieh FJ, Chang FM, Huang HC, et al. Computer-assisted analysis for prediction of fetal weight by ultrasound—comparison of biparietal diameter, abdominal circumference and femur length. J Formos Med Assoc 1987;86:957–964.

Ko TM, Hsieh FJ, Chen CA, Cheng YT, Chen HY. Ultrasonic measurement of fetal femur length. Obstet Gynecol ROC 1983;22:255–260.

Kothari R, Cualing H, Balachander T. Neural network analysis of flow cytometry immunophenotype data. IEEE Trans Biomed Eng 1996; 43:803–810.

Liang RI, Chang FM, Yao BL, et al. Predicting birth weight by fetal upper-arm volume with use of three-dimensional ultrasound. Am J Obstet Gynecol 1997;177:632–638.

Park YR, Murray TJ, Chen C. Predicting sun spots using a layered perceptron neural network. IEEE Trans Neural Networks 1996;7: 501–505.

Sabbagha RE, Minogue J, Tamura RK, Hungerford SA. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, and small-for-gestational-age fetuses. Am J Obstet Gynecol 1989;160:854–862.

Walker AJ, Cross SS, Harrison RF. Visualization of biomedical datasets by use of growing cell structure networks: A novel diagnostic classification technique. Lancet 1999;354:1518–1521.

Warsof SL, Gohan P, Berkowitz RL, Hobbins JC. The estimation of fetal weight by computer-assisted analysis. Am J Obstet Gynecol 1977;128:881–892.