Jacob L. Jaremko*
Philippe Poncet
Janet Ronsky
James Harder

Dept. of Surgery
Faculty of Medicine,
University of Calgary,
3330 Hospital Dr. NW,
Calgary, AB, T2N 4N1

Jean Dansereau
Dept. of Mechanical Engineering,
Ecole Polytechnique,
C.P. 6079,
Succ. Centre–Ville,
Montreal, PQ H3C 3A7

Hubert Labelle
Ste. Justine Hospital,
3175 Cote Ste-Catherine,
Montreal PQ H3T 1C5

Ronald F. Zernicke
Dept. of Surgery
Faculty of Medicine,
University of Calgary,
3330 Hospital Dr. NW,
Calgary, AB, T2N 4N1

Genetic Algorithm–Neural Network Estimation of Cobb Angle from Torso Asymmetry in Scoliosis

Scoliosis severity, measured by the Cobb angle, was estimated by artificial neural network from indices of torso surface asymmetry using a genetic algorithm to select the optimal set of input torso indices. Estimates of the Cobb angle were accurate within 5° in two-thirds, and within 10° in six-sevenths, of a test set of 115 scans of 48 scoliosis patients, showing promise for future longitudinal studies to detect scoliosis progression without use of X-rays. [DOI: 10.1115/1.1503375]

Keywords: Scoliosis, Torso Imaging, Diagnosis, Genetic Algorithms, Neural Networks

Introduction

Adolescent idiopathic scoliosis (AIS) is a three-dimensional spinal deformity most visible as a lateral spinal curvature and associated with asymmetry of the trunk and rib cage. AIS patients are currently monitored by measuring the progression of spinal curvature (measured by the Cobb angle) in a series of full spinal X-rays. This radiation exposure has been estimated to lead to an increased risk of cancer of up to 2.4/1000 scoliosis patients, similar to the risk of developing thyroid cancer in the general population [1]. Particularly since less than 10% of diagnosed scoliotic curves progress to require treatment [2], it is desirable to reduce this risk, for example by noninvasive assessment of curve progression from changes in torso surface deformity. Torso deformity is often the first clue to diagnosis of AIS and remains the most obvious feature of AIS for the patient, although the relation between torso and spinal deformity is complex and difficult to describe analytically [3–5]. While many investigators have focused on indices describing the asymmetry of the back surface [4–9], we found that use of indices describing the deformity of 360° torso cross-sections improved the prediction of the Cobb angle [10]. In that study we used an artificial neural network (ANN) to make the complex link between surface and spinal deformity, an approach that also gave improved results compared to linear regression when relating spinal deformity to rib rotation [11].

ANN's are pattern-recognition tools that relate input and output data even when the form of the relation is non-linear or unknown [12,13]. As with the biological nervous system, a typical ANN consists of several layers of interlinking nodes (Fig. 1). ''Learning'' occurs during repeated presentation of input-output pairs through link weight adjustment to reduce the error between actual and desired output [14]. The ''memory'' of the patterns the trained ANN has experienced is distributed throughout the network links, is modifiable at any time by further training, and can be generalized to classify new patterns [15]. Here, we trained an ANN to use a collection of input indices defining torso surface asymmetry to estimate the Cobb angle of spinal deformity for a group of scoliosis patients. The ability of the trained ANN to estimate the Cobb angle was evaluated by supplying inputs from a ''test set'' of data not used during training and comparing the ANN-estimated Cobb angle with the actual Cobb angle.

Effective ANN training required a set of input indices that described the surface deformity as completely and as efficiently as possible. Even after integrating thousands of raw torso surface data points into several dozen asymmetry indices, we were still faced with the challenge of selecting the most appropriate of these indices to use as ANN inputs. The simplest approach was to use the indices selected by a stepwise linear regression model using the same data set. However, the inputs giving optimal ANN performance may not necessarily be those that best correlate linearly to the output [11]. Several alternatives to linear regression in index-selection exist, including piecewise regression splines developed from combinations of input indices [16], principal com-

*Corresponding author: Jacob Jaremko, c/o Dr. R. F. Zernicke, Dept. of Surgery, Faculty of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, AB, Canada T2N 4N1, phone: (403) 220-5607, fax: (403) 270-0448, e-mail: jljaremk@ucalgary.ca, zernicke@ucalgary.ca

Contributed by the Bioengineering Division for publication in the JOURNAL OF BIOMECHANICAL ENGINEERING. Manuscript received July 2001; revised manuscript received, June 2002. Associate Editor: C. L. Vaughan.

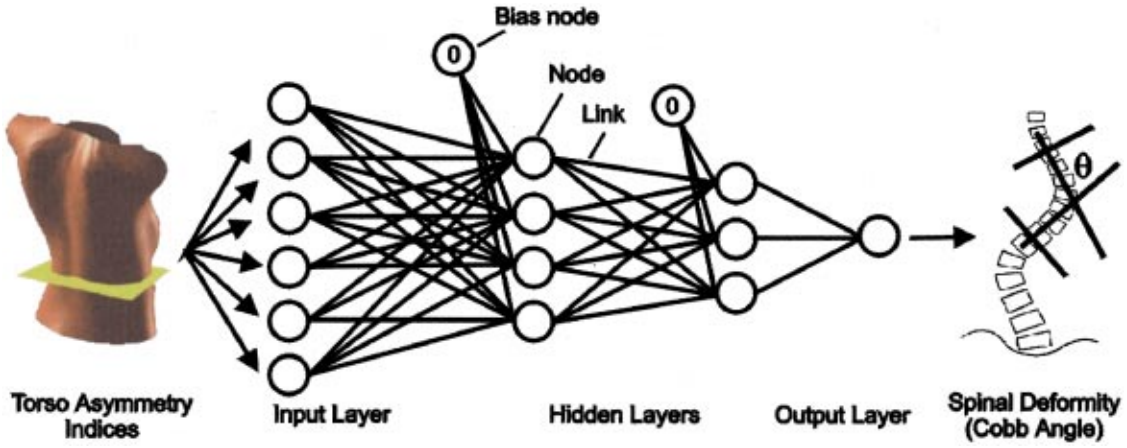

Fig. 1 Artificial neural network (ANN) architecture. The input layer accepted indices of torso asymmetry. In the hidden or processing layers, each node produced an output if the sum of inputs from connected links, multiplied by the link weights, was sufficiently large. Bias nodes functioned as constant terms in the ANN, like y-intercepts in linear regression models. Use of hidden layers enabled nonlinear calculations recognizing features of the input data. The ANN output in this study was an estimate of the Cobb angle.

ponent analysis [17], Kohonen feature mapping [18–20], and a well-studied, highly organized form of systematic trial called a genetic algorithm (GA) [21,22]. In a GA, a ''population'' of randomly selected sets of indices is generated and the ''fitness'' of each member is measured as the performance of an ANN using that set of inputs. By biological analogy, the ''fittest'' members of the population are ''selected'' for the next generation, with operators such as random mutation and cross-over enhancing the ''evolution'' of the population of indices toward increased fitness over successive generations. Ideally, this population converges toward an optimal set of indices for ANN use.

The genetic algorithm-neural network (GA-ANN) combination has been used successfully in several clinical studies. Dybowski et al. used a GA-ANN to predict the likelihood of death in 258 critically ill patients [23]. Although due to limited computing power they used the GA only on 18 indices chosen by decision tree from 157 available indices of clinical status, a 5-layer ANN using inputs selected from these 18 indices gave correct predictions 86% of the time compared to 75% for a logistic regression model. Jefferson et al. have used GA-ANN's in several studies, first predicting surgical outcome from indices such as age, sex, tumor volume and cell type for 620 lung cancer patients [24]. In this study the GA-ANN was up to twice as accurate as group statistics or logistic regression at predicting whether the patient would remain alive 6, 12, 18 or 24 months after surgery. This group also successfully estimated a quantitative output, using a 3-layer ANN with a 32-generation GA to estimate the volume of blood loss during hemorrhage in 33 rats from 3 of 11 available hemodynamic parameters [25]. The GA-ANN predicted blood loss with a significantly lower root-mean-square error than stepwise logistic regression. Jefferson et al. also used a GA-ANN to predict whether depression would occur after mania [26]. The GA was used both to select inputs from the 16 available in 100 patients and to ''prune'' the 3-layer ANN, selecting only the network links which contributed to the final solution. The combination of index-selection and pruning by GA resulted in an ANN that was at least as accurate as, and significantly more stable than either the ''ordinary'' ANN or regression models, though the GA-ANN was slow, requiring 100 generations for convergence. In all of these studies, the GA-ANN combination was more effective than linear regression or naïve neural-network techniques. Here, building on our earlier pilot study that used trial and error for index selection [10], we assessed whether genetic algorithm–neural network analysis of torso asymmetry indices could estimate the Cobb angle with clinically relevant accuracy. We hoped to improve upon the results of the pilot study by using a much more comprehensive set of torso indices (47 vs. 26, defining aspects of surface deformity not previously considered), a larger data set ($n=115$ vs. $n=48$), and the powerful genetic-algorithm analysis technique. The increased data set size allowed us to evaluate ANN accuracy overall and in subgroups based on bracing status, age, curve type and curve direction.

Methods

Data Acquisition Technique. We acquired for each patient 3-D surface coordinates of the full 360° torso (accuracy of location: 2.4 mm) and generated a 3-D model of the spine and rib cage from two X-rays (one postero-anterior (PA-0) and one at 20° to horizontal) using a four-scanner laser system and X-ray projector [27,28]. Contours were cut through the torso model at a 10 mm vertical separation and analyzed for asymmetry. Vertebral levels were estimated within ± 1 level from skin markers at the vertebra prominens and posterior superior iliac spines (PSIS) [29].

Clinical Indices. For each patient-visit, we recorded the following: the Cobb angle and location of the primary scoliotic curve (determined from the PA-0 X-ray film by the attending orthopaedic surgeon and/or a staff radiologist), age, sex, height, weight, bracing status (1 = patient has had brace treatment to date, 0 = patient remains unbraced), and scoliosis diagnosis (adolescent idiopathic scoliosis = 1, juvenile = 2, congenital or other = 3). The computer-Cobb angle generated from the 3-D X-ray reconstruction [28] was essentially interchangeable with the clinical Cobb angle ($r=0.99$).

Torso Indices. Indices of torso asymmetry were derived from the outline and shape of torso cross-sections (Fig. 2(*a*)). Each index had a numeric value at each vertical level (e.g., back surface rotation at T12 = 6.2°). The extrema, range and ''quasi-Cobb angle'' (Fig. 2(*b*)) were recorded for each index in the torso zone that gave optimal correlation to the Cobb angle (typically below the mid-thorax since above this the arms and breasts introduced non-spinal asymmetry). Torso indices included lateral deviation of the centroid line and of the spinous process line (estimated visually from cross-section profiles), orientation of the cross-sectional principal axes of inertia (PAX), cross-section eccentricity (i.e., degree of circularity), back surface rotation (BSR), difference between PAX and BSR, rib hump, left-right asymmetry of half-centroid locations, moments of inertia, and aspect ratios (Fig. 2(*c*)–(*d*).

Data Set. Consenting clinic patients with adolescent or juvenile ($n=42$) or congenital ($n=6$) idiopathic scoliosis who had

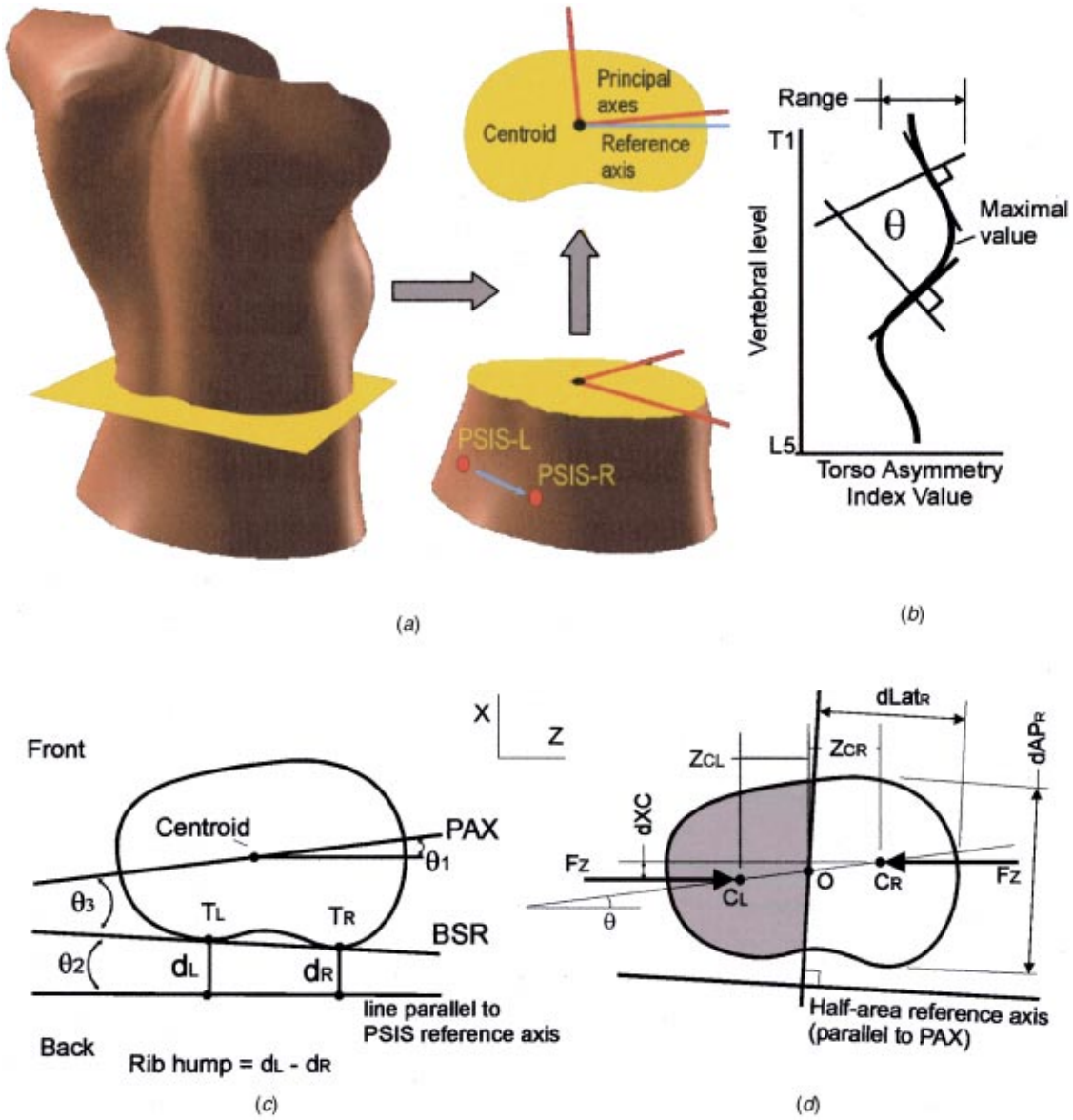

Fig. 2 Calculation of selected torso asymmetry indices. (*a*) Cross-sections were cut through the torso surface model. PSIS axis=line joining skin markers on posterior superior iliac spines. (*b*) The quasi-Cobb angle (θ) was computed for each index of asymmetry between the most appropriate pair of points of inflection on the index-curve. (*c*) Angles of principal axis (PAX) rotation (θ_1), back surface rotation (BSR, θ_2), and the difference between BSR and PAX rotation (θ_3) were recorded relative to the PSIS axis along with the "rib hump" ($d_L - d_R$). (*d*) Half-areas were cut relative to the PAX reference axis. Asymmetry of half-centroids C_L and C_R was measured in the antero-posterior (dXC) and lateral (Z_{CL} vs. Z_{CR}) directions. The angle of rotation of the line joining the half-centroids (θ), and the difference between aspect ratios (dAP/dLat) between left and right half-areas were also computed.

not undergone spinal fusion surgery were concurrently scanned and X-rayed at one or more of five data collections between May, 1998 and May, 2000. The study protocol was approved by the hospital ethics committee. Since patient age, curve severity and posture varied at each semi-annual visit, each scan-X-ray pair was considered a unique data set, giving 115 data sets in 48 patients (38/48 female; 26/48 undergoing brace treatment). The average patient age at time of scan (mean $\pm$ standard deviation, S.D.) was 13.1 $\pm$ 2.8 yr (range: 8.2–18.6 yr), while Cobb angles averaged 34.6 $\pm$ 14.0° (range: 10–71°). To simulate a "real-world" ANN application, data from the first four collection periods (May, 1998 to November, 1999, n=89) were used for GA and ANN training and data from the final collection period (May, 2000, n=26) were used to test the trained ANN. For subgroup analysis, patients were subdivided based on age, curve severity, direction of primary curve (right, n=91; left, n=24), and whether they had ever undergone brace treatment at the time of scan (n=63) or not (n=52).

Genetic Algorithm Design. To determine which combination of the 47 available input indices resulted in the most effective ANN estimation of the Cobb angle, we used a customized genetic algorithm (GA) based on a freeware package [30] running on Matlab (The Mathworks Inc., Natick, MA, v. 6.0, 2000). For each GA trial, an initial "population" of "chromosomes" each made up of 47 "genes" was generated (Fig. 3.1). Each "gene" represented one of the available input indices, and was randomly given a value of zero ("do not use") or one ("okay to use"). The "fitness" of

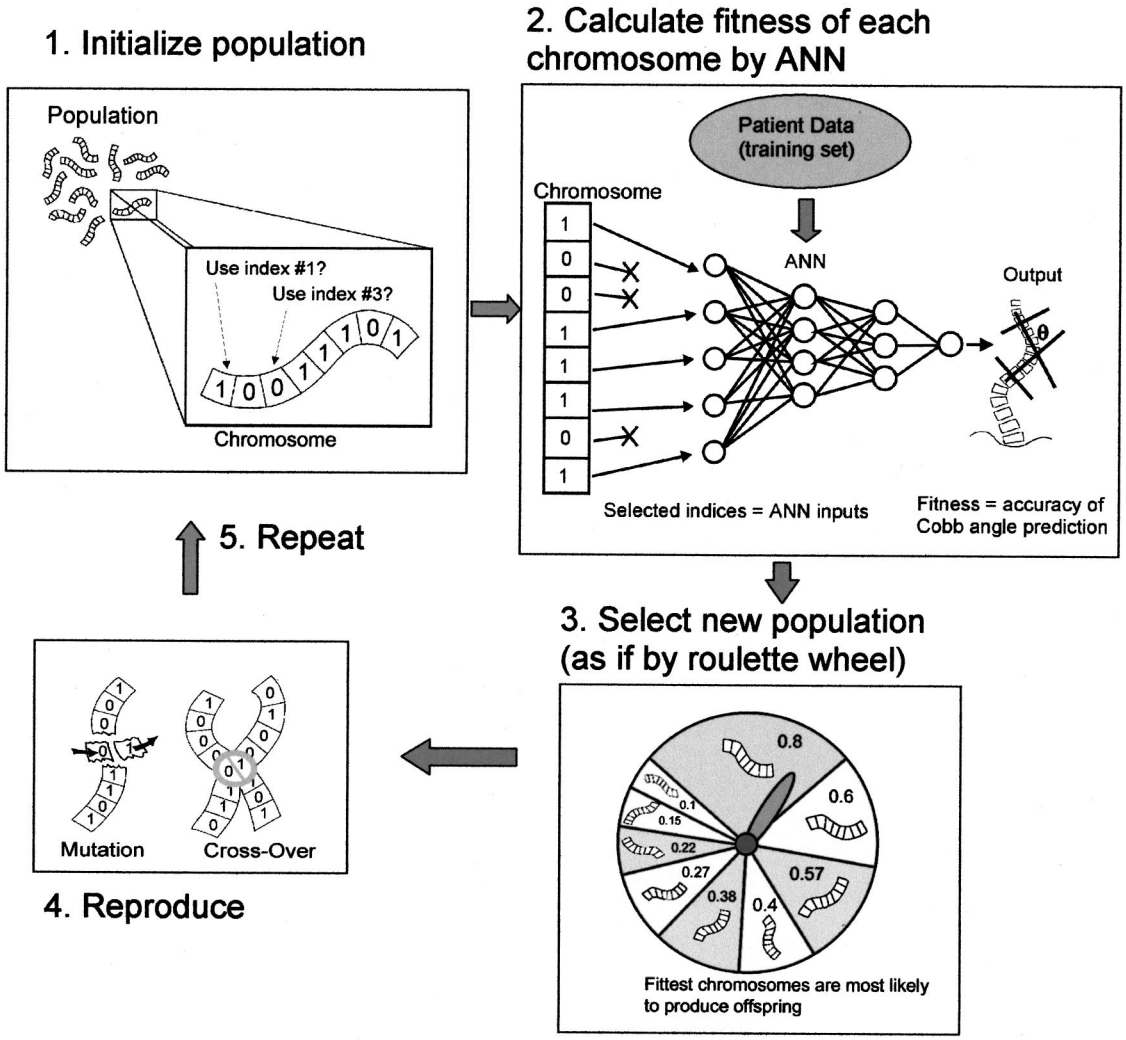

Fig. 3 Genetic algorithm for selection of neural-network (ANN) input indices. (1) The population was initialized by creating a set of "chromosomes" containing a "gene" for each available input index, randomly set to zero ("do not use") or one ("ok to use"). (2) The fitness of each chromosome was the accuracy of Cobb-angle predictions from an ANN using only the indices whose genes have non-zero values for that chromosome. (3) A new population of the same size as the original one was created by selecting chromosomes as if spinning a roulette wheel, with the chance of a chromosome being selected for the new population being proportional to that chromosome's fitness. (4) These selected chromosomes then "reproduced" via mutation and cross-over. (5) The algorithm repeated, starting a new "generation" with the updated population. Over time, the population "evolved" towards optimal fitness.

each chromosome was calculated based on the accuracy of Cobb angle estimation by an ANN using only those input indices whose genes had a value of one in that chromosome (Fig. 3.2). This method of fitness calculation was the most direct way to solve the ANN-optimization problem, but it necessitated substantial computing power because it required a complete ANN trial using the full training set for each chromosome in each generation. A new population of the same size as the previous one was generated by "selection" based partly on the "fitness" of each chromosome and partly on random chance (to maintain variation in the "gene pool"; Fig. 3.3), followed by random "mutation" and "cross-over" of genes (Fig. 3.4). This process of fitness-evaluation, selection, and reproduction was repeated for many "generations" until the population as a whole converged towards the optimal "chromosome" whose pattern of zeros and ones defined the best set of ANN input indices.

In pilot testing we found that performance was best with an initial population of 20 or more chromosomes. Convergence was faster (10–20 generations vs. 30–40 generations) when the population was "seeded" with non-random chromosomes representing the sets of indices that contributed significantly ($p<0.05$) to a linear regression model predicting the Cobb angle or that correlated moderately ($r>0.5$) to the Cobb angle. However, if we "seeded" more than 10% of the population, the GA converged prematurely (e.g., in just 3-5 generations when 50% of the population was seeded), presumably because it then lacked variability in its members. Premature convergence typically resulted in an ANN that was no more accurate than the regression model its inputs were derived from. In pilot testing we found that the rate of binary mutation (conversion of ones to zeros and vice versa) for optimal performance was 5%, that different cross-over rates had little effect on performance, and that chromosomes were best chosen to survive based on a normalized geometric ranking scheme [30] where, for example, the chance that the fittest chromosome out of 20 would survive each generation was 15%, falling to just 0.4% for the least-fit chromosome. Since in practice the final population in the algorithm did not completely converge to a single "fittest" chromosome but contained slight variants, the op-

timal torso indices for Cobb angle prediction were deemed to be those that were selected in more than 50% of the most-fit chromosomes from the final population.

Neural-Network Architecture and Training. The ''fitness function'' used by the genetic algorithm was a customized ANN designed to predict the Cobb angle from torso input indices using the Levenberg-Marquardt back-propagation training algorithm from the Matlab Neural Network Toolbox (v. 4.0, 2000). While ANN link weights could themselves have been derived by genetic algorithm, this process would have been very slow [26]. Here, we used the GA solely to select appropriate ANN inputs, preferring much faster back-propagation for internal optimization of the ANN link weights. The ANN was trained to recognize input-output relations using a ''learning set'' comprising 80% of the training set ($n=71$), with the remaining 20% ($n=18$) kept aside as a ''validation set.'' To ensure that the ANN results could be generalized to new data, ANN training in each iteration was halted once error in the validation set began to increase [31]. The validation set was randomly extracted from the training set such that input indices did not differ significantly ($p>0.05$) between it and the remaining learning set. To minimize the possible effect of coincidental relations, the training set was re-split randomly into learning and validation sets every five generations of the GA. ANN performance varied even when re-using the same learning/validation split, since each ANN run began with a different random initialization of network link weights. To minimize this variability, we used Bayesian regularization to adjust link weights [31] and chose the best-performing ANN from multiple repeats of each training session. The more ANN repeats were performed, the more likely it was for at least one to escape local minima and approach an optimal solution. We investigated the effects on network performance of varying the number of ANN repeats, of using either one or two hidden layers each with a varying number of nodes, and of changing the learning/validation split.

Performance Measurement. Since in this study our goal was to generate ANN predictions as close as possible to the actual Cobb angle, we defined the ''fitness'' of each chromosome in the GA based on the accuracy of the ANN using the input indices indicated by that chromosome. Since scoliotic curve progression is considered to occur with a 5–10° increase in Cobb angle [4,32], we evaluated ANN performance in this range of accuracy. Because the fitness function was needed only for purposes of chromosome ranking, results were not sensitive to the exact weights used, so we simply chose weights empirically to be 40% for predictions accurate within 5°, 40% for 7.5°, and 20% for 10°. For example, if an ANN used the set of indices from a given chromosome to estimate 60%, 75% and 90% of Cobb angles within 5°, 7.5° and 10° respectively, fitness was calculated as $0.4\times 60\% + 0.4\times 75\% + 0.2\times 90\% = 0.72$. We used this weighted-average measure of ''fitness'' instead of the standard error of estimation of the Cobb angle (S.E.E.) because in case there were a few poorly-predicted outliers (e.g., the sometimes-observed ''non-standard rotation'' subgroup [5]), the weighted-average ''fitness'' would be more sensitive to small changes in accuracy of prediction within the bulk of the group than the S.E.E., which would be dominated by the few outliers. Once the GA was complete, ANN performance using the optimal ''chromosome'' of torso input indices was assessed in more detail. In the training and test sets we measured the fitness, S.E.E., percentage of predictions within 5° and 10°, and Pearson correlation coefficient between actual and estimated Cobb angles. Sensitivity (SN), specificity (SP), and positive and negative predictive values (PPV, NPV) were computed for a hypothetical clinical test where a test-positive result was an ANN prediction of a Cobb angle greater than a specified threshold (e.g., 30°), and a true-positive result was a Cobb angle greater than that threshold. ANN performance was also assessed in subgroups of patients (pooling the training and test sets due to small sample sizes) with differing age, curve severity, curve type and bracing status.

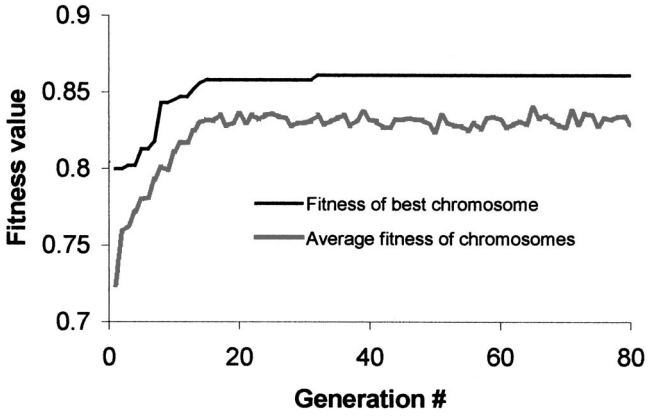

Fig. 4 Genetic algorithm (GA) training showing typically rapid convergence. This example GA tested 20 individuals each with chromosomes of 24 input indices. Fitness was already high in the initial population due to "seeding" of the population with individuals whose indices correlated well to the Cobb angle.

Statistical Methods. Pearson correlation coefficients between pairs of input indices and between input indices and the Cobb angle were computed and tested for significance using Fisher's Z transformation. For comparison with the GA, a stepwise linear regression model was generated to pick the set of indices that best predicted the Cobb angle from the 47 available torso input indices (entry on probability of $F<0.05$, removal on probability of $F>0.10$, modeled using SPSS v. 9.0, SPSS Inc., Chicago, Ill., 1999).

Results

Neural-Network Training. GA analysis was computationally intensive. On a 600 MHz IBM PC, a typical ''seeded'' GA required ~20 generations to converge (Fig. 4), at ~90 min per generation when using 40 chromosomes and 10 ANN repeats per chromosome. We used 10 ANN repeats per chromosome as a compromise between training time and reliability of results. Accuracy of Cobb angle estimation from 35 training sessions using the best of 10 ANN repeats varied by 0.7% (SD), compared to 5.2% for 3 ANN repeats. We reduced training time by ~20% with little effect on ANN performance by minimizing network size. Use of one hidden layer with 3 nodes in 30 ANN trials (10 repeats each) caused network ''fitness'' to vary by only 3.3% (SD) from that of an ANN using 2 hidden layers with 11 and 6 nodes. ANN performance was insensitive to the specific split in the data set, as fitness varied by 2.6% (S.D.) between 10 ANN trials (10 repeats each) using different random splits of the training set into learning and validation sets. Variation in performance due to differing splits would likely be further reduced in a larger data set.

GA-ANN Performance. The best GA used a 40-chromosome population and 40 generations (reaching 95% of the final fitness value after 9 generations) to select 17/47 indices for optimal ANN Cobb angle estimation, including the 4 indices used in the comparable stepwise regression model (Table 1). Several selected indices did not correlate significantly with the Cobb angle (Table 1) and may have been used by the GA-ANN to sort data into groups or to ''flag'' non-spinal causes of asymmetry. The ANN using these 17 indices was 30% more accurate than stepwise regression and 26% more accurate than a ''naïve'' ANN using the inputs derived from stepwise regression, estimating the Cobb angle within 5° in 65% of the test set and 84% of the training set (and within 10° in 85% and 99% respectively) (Table 2). The actual

Table 1 Torso asymmetry indices selected by genetic algorithm & regression for optimal Cobb angle estimation

Index Category & Description	Unit	Zone	Used by: linreg?*	G.A.?**	"r" to Cobb***
Principal Axis (PAX) Rotation					
Maximum magnitude of PAX rotation	°	T9-L4		^	0.72
Maximal clockwise PAX rotation	°	T11-L4	^	^	0.33
Range of PAX rotation	°	T12-L5		^	0.80
Back Surface Indices					
Maximum magnitude of back surface rotation (BSR)	°	T5-T12		^	0.56
Range of net rib hump	mm	T10-S1	^	^	0.79
"Quasi-Cobb angle" of spinous process line	°	T10-L2		^	0.53
Range of spinous process lateral deviation	mm	T12-L4		^	0.50
Torso centroid					
Range of centroid lateral deviation	mm	T7-L4	^	^	0.77
Asymmetry of left & right half-areas					
Range of left-right difference in aspect ratios	. . .	T7-L2			0.59
Range of half-centroid AP asymmetry	mm	T11-S1	^	^	0.83
"Quasi-Cobb angle" of left-right difference in lateral inertia, measured with respect to PAX axis.	. . .	T11-S1			0.62
Clinical indices					
Body mass index	kg/m^2	. . .		^	n.s.
Diagnosis (1=AIS, 2=JIS, 3=OS)	. . .	. . .		^	n.s.
T1-L5 torso height	mm	. . .		^	n.s.
Indices that identify outliers					
Minimum cross-section eccentricity	. . .	T8-L5		^	0.20
AP imbalance of torso centroid line	mm	T1-L5		^	n.s.
Range of difference between PAX and BSR	°	T11-L5		^	0.71

*linreg=stepwise regression model estimating Cobb angle (entry on probability of $F<0.05$, removal on probability of $F>0.10$).
**GA=genetic algorithm estimating Cobb angle by neural-network output ("J-Cobbangle").
***=Pearson coefficient of correlation of index to Cobb angle in the full data set ($n=115$).
"n.s."=not significant ($p>0.05$).
AP=antero-posterior. AIS, JIS=adolescent/juvenile idiopathic scoliosis. OS=other scoliosis.

and GA-ANN-estimated Cobb angles correlated strongly ($r=0.93-0.97$, $p<0.0001$, Fig. 5). The GA-ANN was highly sensitive and specific for Cobb angles above 20°, for example detecting Cobb angles >30° in the test set (prevalence=39%) with SN=100%, SP=94%, PPV=91%, and NPV=100% (Fig. 6). Regression and ANN models performed better on the training set than on the test set (Fig. 5), an "overfitting" effect that is common with small sample sizes and would likely be reduced as more patients are added to the database [31].

The GA-ANN outperformed stepwise regression overall (Table 2) and particularly in certain subgroups such as older patients with large curves (age >12 yr, Cobb angles >30°, n=37), in which GA-ANN and regression estimates of the Cobb angle were within 5° in 89% and 54% of the test set, respectively. GA-ANN estimates of the Cobb angle varied in accuracy between subgroups, being within 5° of the clinical Cobb angle in 87% of unbraced patients (n=52, all subgroups pooled between training and test sets), 75% of braced patients (n=63), 73% of young patients (age <13 yr, n=60), 87% of older patients (age >12 yr, n=55), 79% of patients with mild curves (Cobb angle <30°, n=57), 81% of patients with large curves (Cobb angle >30°, n=58), 77% of patients with a rightward primary curve (n=91) and 92% of patients with a leftward curve (n=24). These variations supported the idea that each patient group had distinct characteristics and ideally should be analyzed separately. While we did not have enough data to train a "local" ANN effectively for each subgroup, when we trained "local" regression models (which did not require as much data as an ANN) in each subgroup, we found substantial improvements in performance. For example, a stepwise regression model trained only on patients with leftward scoliotic curves (n=24) estimated 88% of Cobb angles in that group within 5°, compared to 63% for the overall stepwise regression model trained on all patients and tested on the leftward-curve group.

Discussion

The surface asymmetry of scoliosis seems to be caused mechanically by deformity of the underlying scoliotic spine and rib cage, though this complex spine-surface relation is extremely difficult to model directly (e.g., through a finite element model). In this paper we have developed a genetic algorithm-neural network (GA-ANN) to relate spine and surface deformity based on the assumption that the mechanical links (through ribs and intervening tissues) between spine and surface give rise to common patterns of surface-spine relation that enable consistent estimation of spinal deformity from changes in surface shape. Since a scoliosis curve is considered to have progressed with a >5° increase in the Cobb angle between X-rays [32], with some suggesting a threshold of 10° [4], successful Cobb angle estimation should be accurate within 5–10°. The GA-ANN analysis of torso asymmetry indices presented here was accurate in the test set within 5° in two-thirds of scans and within 10° in six out of seven scans. This accuracy is on the threshold of clinical usefulness and may already be sufficient to predict a 5°–10° change in the Cobb angle of an individual patient, a hypothesis we will test as we gather more longitudinal data. We predicted more than twice as large a proportion of Cobb angles within 5° than was possible in a comparable study predicting the Cobb angle from indices of back surface asymmetry in 104 patients [5], and 30% more than in our pilot study that used ANN analysis (without GA) of a partial set of indices of torso asymmetry in 65 patients [10]. Particularly encouraging was the high sensitivity and specificity of the network for Cobb angles from 25°–40°, as the GA-ANN was able to

Table 2 Genetic algorithm-neural network and regression model estimation of the Cobb angle

	ANN		Regression	
Performance measure	Training set ($n=89$)	Test set ($n=26$)	Training set ($n=89$)	Test set ($n=26$)
"r" actual vs. desired	0.97	0.93	0.93	0.89
S.D. actual vs. desired (°)	3.9	6.7	5.9	9.3
Percent correct within 5°	0.84%	0.65%	0.69%	0.46%
Percent correct within 10°	0.99%	0.85%	0.88%	0.73%
Fitness	0.899	0.723	0.760	0.577

S.D.=standard deviation. ANN=neural-network trained using genetic-algorithm-selected indices. Regression=stepwise regression model. Fitness=(weighted) sum of percent correct within 5° (40%), 7.5° (40%) and 10° (20%).

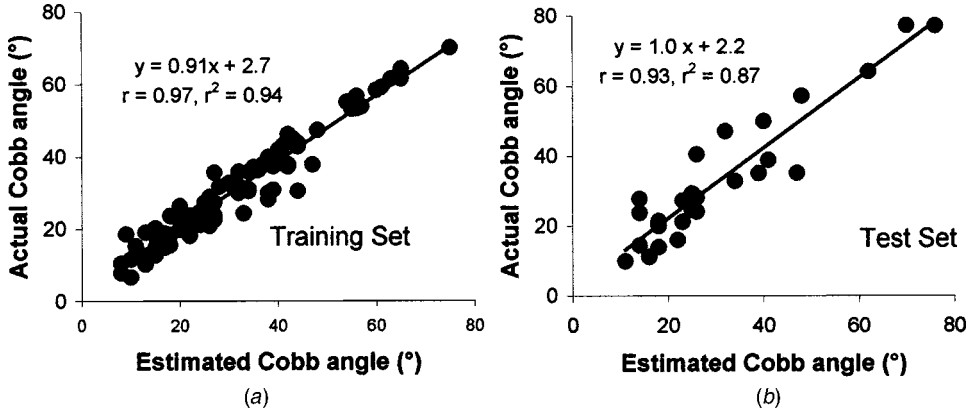

Fig. 5 Actual Cobb angle vs. ANN-estimate of Cobb angle, in (*a*) training set and (*b*) test set. ANN results were better in the training set; this "over-fitting" would be reduced by use of a larger data set.

distinguish accurately between curves in this important zone where decisions to apply bracing or surgical treatment are made. The network predictions were robust, varying by $<1\%$ (S.D.) with repeated ANN trials, by 3% with different internal network architectures, and by $<3\%$ with different splits of the data into learning and validation sets. The ANN accuracy was reduced by problems specific to certain torso indices: our visual estimate of the spinous process line was less reliable than that of others who palpated and marked the spinous processes [4,33], and our measurement of back surface rotation was sensitive to inaccuracy in the location of specific surface points related to calibration of our four-scanner torso scan system (3-D error in location: 2.4 mm). We expect accuracy to improve with correction of these issues and particularly with the scanning of more patients, since a larger sample size should reduce ANN "overfit" and raise ANN performance in the test set closer to that in the training set.

GA-ANN performance varied by clinical subgroups, with Cobb-angle estimation most accurate in older patients (perhaps due to more consistent posture), in leftward curves, and in unbraced patients (perhaps due to alteration of the normal surface-spine relation in braced patients). Accuracy was reduced for curves $<25°$, perhaps because of a reduced ratio of "signal" (torso asymmetry due to spinal deformity) to "noise" (torso asymmetry due to non-spinal causes). "Local" regression models trained and tested within a given subgroup produced better Cobb-angle predictions than the "global" model trained on the entire data set; we expect "local" ANN models to have similarly improved accuracy.

Because the data for this study were collected as part of our ongoing longitudinal study of curve progression, the data set consisted of multiple scans of the same patients (115/48=an average of 2.4 scans per patient). While scans were separated by intervals of at least 6 months and involved differences in patient growth and posture, they were still not entirely independent. Thus, while we cannot be sure from this study that a GA-ANN can accurately estimate any and all scoliotic deformity from a single surface scan of a new patient (though this may be possible, given a sufficiently large training database), we can make the clinically valuable observation that a GA-ANN accurately estimated each patient's current Cobb angle from surface asymmetry when trained using a diverse data set that generally included 1–2 previous scans of that patient.

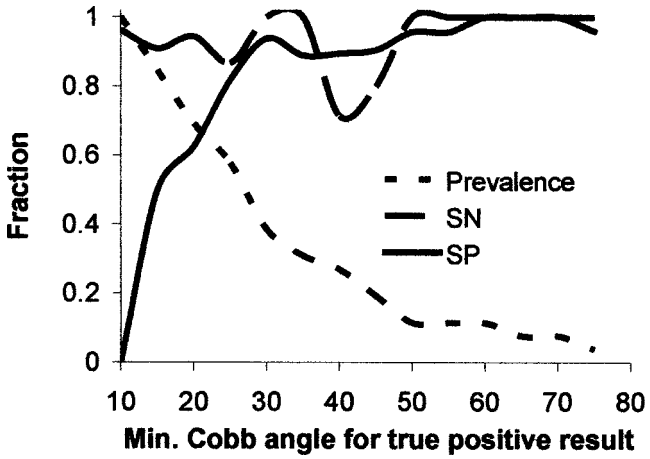

Fig. 6 Sensitivity (SN) and specificity (SP) of the prediction that a patient's Cobb angle would be greater than the specified minimum value, as fractions of the maximum possible value of 1.0. From the prevalence line (the proportion of the data set having Cobb angles greater than a given value) it is clear that 70% of our patients had curves between 20°–50°. In this zone, both SN and SP were greater than 0.7, each peaking near 0.95 at thresholds of 25°–35°. Estimation of severe curves >50° was also highly sensitive and specific.

Conclusion

Combined genetic algorithm-neural network (GA-ANN) analysis of 360° torso asymmetry in 115 idiopathic scoliosis patients estimated the Cobb angle within 5° in 65%–84% of cases, on the threshold of clinical usefulness for prediction of scoliosis progression. GA-ANN analysis was 30% more accurate than stepwise regression, and while the GA was much more time-consuming than regression, the extra processing time required for genetic-algorithm index selection would be unimportant in clinical use of the system since the GA would not need to be repeated for each new scan. Based on these results, we envision a clinical procedure where a new patient would be X-rayed, diagnosed and assigned to the appropriate subgroup based on age, sex and curve parameters, then monitored by torso scans analyzed by a GA-ANN specifically trained on other patients in that subgroup and on any available previous scans of the same patient. Repeat X-rays would be given only if the GA-ANN estimated that scoliosis progression has occurred. The longitudinal study for which we are currently gathering data will be critical to validation of this procedure, since it will test whether the GA-ANN is sufficiently accurate to detect progression in an individual patient over time. Future work will also include improvements in measurement of the spinous process line and back surface rotation, training of "local" neural-network models in subgroups of patients with different clinical character-

istics, and assessment of the relations between other features of surface and spinal deformity and of the effects of brace treatment on torso asymmetry.

Acknowledgments

The authors thank Dr. C.-E. Aubin (Ecole Polytechnique, Montreal), Dr. N.G. Shrive (Civil Engineering, University of Calgary), Dr. C. Frank (Surgery, University of Calgary), and R. Dudley (University of Calgary) for their advice on surface-asymmetry indices, Dr. G. Clynch (Clynch Technologies Inc., Calgary) for supporting our use of the laser scan system, Ms. N. Tardif (Ecole Polytechnique, Montreal) for evaluation of torso scanner accuracy, Dr. R. Dewar and Dr. E. Joughin (Alberta Children's Hospital, Calgary) for their clinical assistance, Dr. P.H. Gu (Mech. Eng., University of Calgary) and Dr. F. Cheriet (Ecole Polytechnique, Montreal) for neural-network advice, and the following organizations for their financial support: Alberta Children's Hospital Foundation, Alberta Heritage Foundation for Medical Research, Arthritis Society of Canada, Canadian Institutes of Health Research, Fraternal Order of Eagles (Alberta & Saskatchewan), Hospital for Sick Children's Foundation, and Natural Sciences & Engineering Research Council of Canada.

References

[1] Levy, A. R., Goldberg, M. S., Mayo, N. E., Hanley, J. A., and Poitras, B., 1996, ''Reducing the Lifetime Risk of Cancer from Spinal Radiographs Among People With Adolescent Idiopathic Scoliosis,'' Spine, **21**(13), pp. 1540–1548.
[2] Reamy, B. V., and Slakey, J. B., 2001, ''Adolescent Idiopathic Scoliosis: Review and Current Concepts,'' Am. Fam. Physician, **64**(1), pp. 111–116.
[3] Pope, M. H., Stokes, A. F., and Moreland, M., 1984, ''The Biomechanics of Scoliosis,'' Crit. Rev. Biomed. Eng., **11**(3), pp. 157–188.
[4] Goldberg, C. J., Kaliszer, M., Moore, D. P., Fogarty, E. E., and Dowling, F. E., 2001, ''Surface Topography, Cobb Angles, and Cosmetic Change in Scoliosis,'' Spine, **26**(4), pp. E55–63.
[5] Stokes, I. A., and Moreland, M. S., 1989, ''Concordance of Back Surface Asymmetry and Spine Shape in Idiopathic Scoliosis,'' Spine, **14**(1), pp. 73–78.
[6] Thulbourne, T., and Gillespie, R., 1976, ''The Rib Hump in Idiopathic Scoliosis. Measurement, Analysis and Response to Treatment,'' J. Bone Joint Surg. Br., **58**(1), pp. 64–71.
[7] Scutt, N. D., Dangerfield, P. H., and Dorgan, J. C., 1996, ''The Relationship Between Surface and Radiological Deformity in Adolescent Idiopathic Scoliosis: Effect of Change in Body Position,'' Eur. Spine J., **5**(2), pp. 85–90.
[8] Theologis, T. N., Fairbank, J. C., Turner-Smith, A. R., and Pantazopoulos, T., 1997, ''Early Detection of Progression in Adolescent Idiopathic Scoliosis by Measurement of Changes in Back Shape with the Integrated Shape Imaging System Scanner,'' Spine, **22**(11), pp. 1223–1227.
[9] Bunnell, W. P., 1984, ''An Objective Criterion for Scoliosis Screening,'' J. Bone Jt. Surg., Am. Vol., **66**(9), pp. 1381–1387.
[10] Jaremko, J. L., Poncet, P., Ronsky, J., Harder, J., Dansereau, J., Labelle, H., and Zernicke, R. F., 2001, ''Estimation of Spinal Deformity in Scoliosis from Torso Surface Cross Sections,'' Spine, **26**(14), pp. 1583–1591.
[11] Jaremko, J. L., Delorme, S., Dansereau, J., Labelle, H., Ronsky, J., Poncet, P., Harder, J., Dewar, R., and Zernicke, R. F., 2000, ''Use of Neural Networks to Correlate Spine and Rib Deformity in Scoliosis,'' Computer Methods in Biomechanics and Biomedical Engineering, **3**(3), pp. 203–213.
[12] Dayhoff, J. E., 1990, Back-Error Propagation, in *Neural Network Architectures: An Introduction*, New York, Van Nostrand Reinhold, pp. 58–79.
[13] Dayhoff, J. E., 1990, *Neural Network Architectures: An Introduction*, Van Nostrand Reinhold, New York.
[14] LeCun, Y., Bottou, L., Orr, G. B., and Muller, K. R., 1998, Efficient Backpropagation, in *Neural Networks: Tricks of the Trade*, Orr, G. B., and Muller, K. R., eds., New York, Springer-Verlag, pp. 9–50.
[15] Cross, S. S., Harrison, R. F., and Kennedy, R. L., 1995, ''Introduction to Neural Networks,'' Lancet, **346**(8982), pp. 1075–1079.
[16] Friedman, J., 1991, ''Multivariate Adaptive Regression Splines (MARS),'' Annals of Statistics, **19**, pp.1–141.
[17] Hertz, J., Krogh, A., and Palmer, R. G., 1991, Principal Component Analysis, in *Introduction to the Theory of Neural Computation*, New York, Addison-Wesley, pp. 204–210.
[18] Natale, C. Di, Macagnano, A., D'Amico, A., and Davide, F., 1997, ''Electronic-Nose Modelling and Data Analysis Using a Self-Organizing Map,'' Measurements in Science & Technology, **8**, pp. 1236–1243.
[19] Dayhoff, J. E., 1990, The Kohonen Feature Map, in *Neural Network Architectures: An Introduction*, New York, Van Nostrand Reinhold, pp. 163–191.
[20] Naes, T., Kvaal, K., Isaksson, T., and Miller, C., 1993, ''ANNs in Multivariate Calibration,'' Journal of Near-Infrared Spectroscopy, **1**, pp. 1–11.
[21] Forrest, S., 1993, ''Genetic Algorithms: Principles of Natural Selection Applied to Computation,'' Science, **261**(5123), pp. 872–878.
[22] Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan.
[23] Dybowski, R., Weller, P., Chang, R., and Gant, V., 1996, ''Prediction of Outcome in Critically Ill Patients Using Artificial Neural Network Synthesized By Genetic Algorithm,'' Lancet, **347**(9009), pp. 1146–1150.
[24] Jefferson, M. F., Pendleton, N., Lucas, S. B., and Horan, M. A., 1997, ''Comparison of a Genetic Algorithm Neural Network with Logistic Regression for Predicting Outcome After Surgery for Patients with Nonsmall Cell Lung Carcinoma,'' Cancer, **79**(7), pp. 1338–1342.
[25] Jefferson, M. F., Pendleton, N., Mohamed, S., Kirkman, E., Little, R. A., Lucas, S. B., and Horan, M. A., 1998, ''Prediction of Hemorrhagic Blood Loss with a Genetic Algorithm Neural Network,'' J. Appl. Physiol., **84**(1), pp. 357–361.
[26] Jefferson, M. F., Pendleton, N., Lucas, C. P., Lucas, S. B., and Horan, M. A., 1998, ''Evolution of Artificial Neural Network Architecture: Prediction of Depression After Mania,'' Methods Inf. Med., **37**(3), pp. 220–225.
[27] Poncet, P., Delorme, S., Ronsky, J. L., Dansereau, J., Harder, J., Clynch, G., Dewar, R. O., Labelle, H., Gu, P. H., and Zernicke, R. F., 2000, ''Reconstruction of Laser-Scanned 3D Torso Topography and Stereo-Radiographical Spine and Rib-Cage Geometry in Scoliosis,'' Computer Methods in Biomechanics & Biomedical Engineering, **4**(1), pp. 59–75.
[28] Labelle, H., Dansereau, J., Bellefleur, C., and Jequier, J. C., 1995, ''Variability of Geometric Measurements from Three-Dimensional Reconstructions of Scoliotic Spines and Rib Cages,'' Eur. Spine J., **4**(2), pp. 88–94.
[29] Jaremko, J. L., Poncet, P., Ronsky, J. L., and Zernicke, R. F., 2000, ''Estimation of Vertebral Levels from Torso Surface Data,'' Arch. Physiol. Biochem., **108**(1–2), pp. 198.
[30] Houck, C. R., Joines, J., and Kay, M., 1996, ''A Genetic Algorithm for Function Optimization: A Matlab Implementation,'' ACM Transactions on Mathematical Software, http://www.ie.ncsu.edu/gaot.
[31] Demuth, H., and Beale, M., 2000, *Neural Network Toolbox User's Guide (v. 4.0 for Matlab v. 6.0)*, The Math Works, Natick, MA.
[32] Weinstein, S. L., and Ponseti, I. V., 1983, ''Curve Progression in Idiopathic Scoliosis,'' J. Bone Jt. Surg., Am. Vol., **65**(4), pp. 447–455.
[33] Drerup, B., and Hierholzer, E., 1996, ''Assessment of Scoliotic Deformity from Back Shape Asymmetry Using an Improved Mathematical Model,'' Clin. Biomech. (Los Angel. Calif.), **11**(7), pp. 376–383.