Reprinted from INVESTIGATIVE RADIOLOGY, October, 1998
Vol. 33, No. 10
Lippincott Williams & Wilkins Printed in U.S.A.

Focal Calvarial Bone Lesions

Comparison of Logistic Regression and Neural Network Models

ESTANISLAO ARANA, MD, PHD.* LUIS MARTÍ-BONMATÍ, MD, PHD,† ROBERTO PAREDES, BSC‡, AND DANIEL BAUTISTA, MD§

Arana E, Martí-Bonmatí L, Paredes R, Bautista D. Focal calvarial bone lesions: Comparison of logistic regression and neural network models. Invest Radiol 1998;33:738–745.

RATIONALE AND OBJECTIVES. **To assess the accuracy of logistic regression (LR) and artificial neural networks (NN) in the diagnosis of calvarial lesions using computed tomography (CT) and to establish the importance of the different features needed for the diagnosis.**

METHODS. **One hundred sixty-seven patients with calvarial lesions as the only known disease were enrolled. The clinical and CT data were used for LR and NN models. Both models were tested with the leave-one-out method. The final results of each model were compared using the area under receiver operating characteristic curves (A_z).**

RESULTS. **Of the lesions, 73.1% were benign and 26.9% were malignant. There was no statistically significant difference between LR and NN in differentiating malignancy. In characterizing the histologic diagnoses, NN was statistically superior to LR. Important NN features needed for malignancy classification were age and edge definition, and for the histologic diagnoses matrix, marginal sclerosis, and age.**

CONCLUSION. **NNs offer wide possibilities over statistics for the study of calvarial lesions other than their superior diagnostic performance.**

KEY WORDS. **Skull neoplasms; computed tomography, head; computers; neural networks; receiver operating characteristic curve; statistics; logistic regression.**

From the *Department of Radiology, Hospital Casa de Salud, Valencia, Spain; the †Department of Radiology, Doctor Peset University Hospital, Valencia, Spain; the ‡Technical Institute of Informatics, Polytechnic University, Valencia, Spain; and the §Public Health Center, Denia, Alicante.

Awarded the 1997 Annual Prize by the Royal Academy of Medicine and Surgery, Seville.

Reprint requests: Estanislao Arana, MD, PhD, Department of Radiology, Hospital Casa Salud, Manuel Candela, 41, 46021 Valencia, Spain.

Received, and accepted for publication .

MOST BONE LESIONS can be accurately detected and differentiated with plain films. However, focal bone lesions in flat bones are better observed with computed tomography (CT) because of its superiority in depicting cortical bone, as in the cranial vault.

The exact incidence of calvarial lesions on imaging examinations is unknown, although it must be low. Because of the lack of familiarity with these rare lesions and their usually nonspecific presenting symptoms, accurate diagnosis is frequently difficult when they are observed in CT studies. Further, most articles on the CT appearance of calvarial lesions deal only with case reports; series comprising more than ten cases are rare.[1–4]

This study compares an artificial neural network (NN) with a logistic regression (LR) model for predicting the malignancy and histology of calvarial lesions. Several methods have been used in developing diagnostic aids for focal bone lesions,[5–7] but only recently has NN been described for this task.[8,9] Neural networks are nonlinear algorithms that can fit quite complex models. Extensive applications of these computer programs in radiology can be found elsewhere.[5,8,9]

The main purpose of our study was not to develop a tool to diagnose calvarial lesions, but to validate the potential effectiveness of NN over LR as a diagnostic aid. We also attempted to identify the minimal subset of features that would yield an accurate diagnosis.

Methods

Patients

To be included in this study, all calvarial lesions had to be histologically proved or followed with clinical and radiologic studies for at least 2 years to be included in this study.

TABLE 1. Recorded Variables with Descriptions

Variable	Value
Age	Years
Gender	Male, female
First symptom noticed	Tumor, pain, headache, incidental finding, others
Symptom length	Months
No. of lesions	Number
Bone	Frontal, parietal, occipital, sutures of fontanels
Centricity	Outer table, diploe, inner table
Maximal diameter	In millimeters
Shape	Circular, ovoid (image in plane of greatest diameter)
Character-appearance	Lytic permeative, lytic motheaten, lytic geographic, blastic, mixed blastic permeative, mixed blastic motheaten, mixed blastic geographic
Expansivity	No, mild, moderate, severe
Swell/mass	None, intracranial, subgaleal, both
Edge definition	Poor, moderate, well
Lobularity	Lobular, smooth
Marginal sclerosis	No, partial, rind (<2 mm), band (>2 mm)
Periosteal reaction	No, yes (any form)
Matrix	None, ground glass, calcified, ossified and sequestration
Cortical involvement	Diploe, internal, external, both corticals
Form of cortical involvement	None, thickened, thinned, broken

These lesions had to be incidentally discovered or the only known manifestation of an underlying entity, with no other signs or symptoms of systemic disease. Patients with congenital malformative or traumatic lesions, systemic disease, and previous resection were excluded. After these exclusions, a total of 167 patients with calvarial focal bone lesions were reviewed from 4012 head CT scans during a 4-year period.

All patients were examined with at least two plain radiographic projections and CT. Two CT scanners were used (9800 and Pace; GE Medical Systems, Milwaukee, WI). Computed tomographic sections were obtained with the window width and level settings that best allowed evaluation of soft-tissue structures and bones. The settings were individualized for each patient.

Lesions were divided into benign and malignant and were also classified according to histologic diagnosis. The benign group comprised all the histologically benign lesions except meningiomas. Meningiomas were included in the malignant group because they can be either benign or malignant at histologic examination, they behave aggressively and frequently recur, and their prognosis cannot be determined by histologic findings.

The radiographs were reviewed by two of the authors (EA, LMB) using guidelines modified from those established by Lodwick.[6] Nineteen structural imaging characteristics, as well as anatomic and demographic data, were evaluated without knowledge of the final diagnosis (Table 1). All findings were recorded on a spreadsheet and were used for both the LR model and the NN analysis. When a single feature had two or more findings in the same lesion, the most severe form was the one recorded.

Study Group

From 4012 head CT scans performed between 1993 and 1996, 359 calvarial lesions different from normal anatomic variations were observed. Histologic or clinical proof of disease could be obtained in 167 patients, resulting in a 4.16% incidence of cranial vault lesions. Surgical confirmation was obtained in 136 cases. In 31 patients the diagnosis was established by clinical and radiologic follow-up for at least 2 years.

The study group comprised 74 male patients (44.6%) and 93 female patients (55.4%), ranging in age from 0.5 to 81 years (31.1 ± 25.5 years, mean ± standard deviation). There were 122 benign lesions (73.1%); they occurred in patients with an age range of 5 to 80 years (26.1 ± 22.9). There were 45 malignant lesions (27%); they occurred in patients with an age range of 0.5 to 84 years (55.7 ± 18.6).

The most common symptoms at onset were palpable mass (74.7%), neurologic symptoms (15.9%), headache (5.3%), and local pain (4.2%). The duration of these symptoms ranged from 2 to 75.5 months (median 13 months). In 18 patients (10.8%), the cranial vault lesion was an incidental finding during the CT examination.

The most frequent lesions are described in Table 2. Lesions with less than 10 occurrences were classified as "others" and discarded from the LR and NN diagnoses. This was done because statistical power is lost with small samples and because we wanted to avoid developing small LR equations. Other benign lesions included dermoid cyst (n = 4), exostosis (n = 3), plexiform neurofibroma (n = 3), fibroma (n = 2), osteomyelitis (n = 2), aneurysmal bone cyst (n = 2), scalp angioma (n = 4), desmoplastic fibroma (n = 1), and nodular fasciitis (n = 1). Other malignant lesions included skin epidermoid cell carcinoma (n = 3),

TABLE 2. Most Common Histological Diagnoses

Diagnosis	*n*	%	Age range (yr)	Median
Langerhans' cell histiocytosis	31	18.6	0.5–35	4
Osteoma	22	13.2	5–84	51.5
Epidermoid cyst	22	13.2	1–80	25.5
Metastasis	21	12.6	5–80	62
Meningioma	17	10.2	29–74	51
Hemangioma	15	9	20–63	44
Fibrous dysplasia	10	6	8–64	18
Others benign	22	13.2	0.5–34	14
Ohters malignant	7	4.2	16–80	43

osteosarcoma ($n = 2$), plasmocytoma ($n = 1$), and dural sarcoma ($n = 1$).

Logistic Regression Analysis

We developed an LR model for malignancy and seven other equations for the most common histologic diagnoses. For each dependent variable (malignancy and diagnosis), LR was performed in a backward form using the unconditional maximum likelihood estimation with all the significant variables by univariate LR analysis ($P < 0.05$). The magnitude of the effect was estimated by the adjusted odds ratio and 95% confidence interval (CI). The LR was chosen instead of a discriminant analysis because there were both qualitative and quantitative variables, and not every variable was normally distributed.[10] The only disadvantage is that the LR approach had several equations for the different diagnoses. The statistical analysis was performed with the SPSS statistical software package (SPSS Inc., Chicago, IL).

Network Architecture

The developed NNs had three layers, with a feed-forward architecture, being trained by the back-propagation algorithm with the sigmoid activation function.[11] The network software was custom-written in Visual C++ language and executed on a 150-MHz compatible personal computer.

The structure of the NN included 61 input units, which corresponded to the clinical and CT data and their possible values. These 61 categories matched the 19 independent variables used in the LR model.

No well-established theoretic method exists for designing an ideal NN,[12] and the optimal number of hidden nodes and iterations is unknown. The best designs are typically determined through trial and error.[13] Although several methods can be used to evaluate the performance of NNs as a diagnostic tool, the area (A_z) under the receiver operating characteristic (ROC) curve gives the best measure of the global accuracy of the model.[14] The performance of every NN was plotted as an ROC curve using the LABROC1 program developed by Metz et al.[15]

Two sets of networks were evaluated, one for malignancy (benign versus malignant) and the other for the different histologies. The output of the NN reflected the likelihood of malignancy, and the diagnosis NN had 7 output nodes, 1 for each of the main diseases. To obtain our NN for malignancy, we adjusted sets of networks with 2 to 10 hidden neurons, combined with different iterations, ranging from 30 to 400. For the best diagnosis NN, we also developed sets of networks with 18 to 26 hidden neurons iterated 50 to 400 times. The best diagnosis NN obtained was the one with the highest average A_z for the 7 main entities. Malignancy NN had 5 hidden neurons and 50 iterations. The best diagnosis NN had 22 nodes in the hidden layer and 150 iterations.

One of the few paradigms of neural network design is that at least one third of the data base should be used for testing;[12] this requirement can barely be fulfilled with a small data set. In these settings, an adequate design is better accomplished with the training of the network by the leave-one-out technique to prevent overfitting. Overtraining is an unwanted phenomenon produced when the NN trains either with the whole data set or with an excessive number of iterations. The network learns by memorizing every connection and weight from all the cases, and its ability to generalize in other patients is reduced.

Generalization of the Models

The aims of a predictive model are to generalize its rules to the general population and to evaluate the performance of the algorithms with a set of cases not used during construction of the model. For generalization of the models, the leave-one-out method was used to test the LR equations and NNs.[16] According to this method, all cases except one are used to develop the LR equation or NN; that model is then tested on the patient who is left out, obtaining the predicted probability for that excluded case. The process was repeated so that every patient is used to construct the model and to test it, but never simultaneously.

Simplification of the Input Features

One of the main purposes of the study was to identify the minimal subset of diagnostic features that would yield an accurate diagnosis, both in the malignancy and histologic analysis. Because of the large number of possible combinations, it was not feasible to investigate all of them by decreasing the number of input features; this would require the development of a new network and then the performance of a leave-one-out on each one. Instead, the input features were ranked according to their importance and then removed one by one until the network performance was significantly degraded.

To rank the variables by importance, one input feature was removed with all its categories, and the resulted new NN was trained using all the remaining input features with the leave-one-out method. The performance of the new network was then measured by the A_z. The process was repeated, eliminating a different feature from the original NN until all features had been excluded once. The hypoth-

TABLE 3. Variables Included for the Logistic Regression Analysis of Malignant Lesions

Variable	OR	CI 95%	*P*
Age	1.0748	1.04–1.10	<0.001
Edge (vs. poor)			
Moderate	0.2	0.03–1.15	0.07
Defined	0.011	0.001–0.09	<0.001
Mass (vs. none)			
Intracranial	43.03	4.06–455.1	0.002
Sugaleal	10.94	0.99–120.7	0.05
Both	21.63	1.75–267.1	0.01
No. of lesions	6.2014	0.97–39.48	0.05

OR: odds ratio; CI: confidence interval.

esis was that the exclusion of an important feature would degrade the performance more than the exclusion of an unimportant variable. Once ranked, the input features were discarded in order from least to most important in an analogous way to backward analysis. The performance of this final simplified NN was compared with those of the original NNs and the LR models.

Comparison

The probabilities predicted for A_z with the NN and LR outputs were compared with a two-tailed, nonparametric approach.[17] The CLABROC algorithm was used for two reasons. First, it gives a maximum-likelihood estimation of the ROC curve based on continuous data, such as the output of the network and statistic models. Second, it accepts the case-sampling variation between the NN and the LR with each patient.

Results

Logistic Regression

The variables included in the LR equation for malignancy were age, edge definition, soft-tissue mass, and number of lesions (Table 3). The features that counted for histologic diagnosis are summarized in Table 4. In Langerhans' cell histiocytosis, the variables included in the LR model were lobularity, shape, and age. Osteoma had as important model variables centricity, age, diameter, and duration of symptoms. For epidermoid cyst, the unique selected variable was bone location. In metastasis, the chosen variables were edge

TABLE 4. Variables for the Diagnoses Logistic Regression Equations

Diagnosis	Variable	OR	CI 95%	*P*
Langerhans' cell histiocytosis	Lobularity (vs. lobulated)			
	Smooth	0.072	0.15–0.33	<0.001
	Shape (vs. circular)			
	Ovoid	0.226	0.07–0.71	0.011
	Age	0.886	0.84–0.93	<0.001
Osteoma	Centricity (vs. diploe)			
	Outer table	174.5	12.23–2490	<0.001
	Internal table	34.85	0.96–1263	0.053
	Age	1.05	1.01–1.09	0.003
	Diameter	0.93	0.88–0.98	0.017
	Symptom length	1.02	0.99–1.04	0.058
Epidermoid cyst	Bone (vs. frontal)			
	Parietal	0.42	0.13–1.35	0.149
	Occipital	1.8	0.56–5.70	0.318
	Sutures/fontanels	12	0.98–146.5	0.052
Metastasis	Edge definition (vs. poor)			
	Moderate	0.09	0.02–0.38	<0.001
	Well	0.01	0.01–0.11	<0.001
	Age	1.06	1.02–1.10	<0.001
	No. of lesions	4.36	1.10–17.28	0.036
Meningioma	Gender (vs. male)			
	Female	7.40	1.64–33.28	0.009
	Periosteal reaction (vs. no)			
	Yes	20.72	5.07–84.7	<0.001
Hemangioma	Shape (vs. circular)			
	Ovoid	0.17	0.04–0.65	0.01
	Soft tissue mass (vs. no)			
	Intracranial	0.03	0.003–0.36	0.005
	Subgaleal	0.21	0.04–1	0.051
	Both	1.10	0.18–6.76	0.912
	Periosteal reaction (vs. no)			
	Yes	9.79	1.65–57.96	0.012
Fibrous dysplasia	Diameter	1.03	1.01–1.06	0.003
	Age	0.96	0.93–0.99	0.038

OR: odds ratio; CI: confidence interval.

TABLE 5. Ranking of Input Features by Single Exclusion for Malignancy and Diagnoses in the Neural Network Analysis*

	Malignancy		Diagnosis	
Rank	Feature	A_z	Feature	A_z
1	Age	0.9125	Matrix	0.9411
2	Edge	0.9198	Marginal sclerosis	0.9432
3	Matrix	0.9250	Age	0.9435
4	Soft tissue mass	0.9262	Soft tissue mass	0.9484
5	Gender	0.927	Edge	0.9516
6	Symptoms	0.9287	Periosteal reaction	0.9523
7	Bone	0.9295	Lobularity	0.9523
8	Expansivity	0.9315	Type cortical invovlement	0.9549
9	Shape	0.9318	Shape	0.9550
10	Type of cortical involvement	0.9323	Diameter	0.9551
11	Lobulation	0.9327	Length	0.9556
12	Centricity	0.9333	No lesions	0.9559
13	Character	0.9338	Symptoms	0.9561
14	Periosteal reaction	0.9346	Gender	0.9568
15	No lesions	0.9353	Cortical involvement	0.9581
16	Symptom length	0.9354	Expansivity	0.9598
17	Cortical involvement	0.9358	Character	0.9615
18	Marginal sclerosis	0.9369	Bone	0.9623
19	Diameter	0.9379	Centricity	0.9629
	None excluded	0.9617	None excluded	0.9673

A_z: area under the ROC curve.
* The variables were excluded individually and the network performance calculated with the missing feature. A large decrease of the A_z implied the exclusion of an important variable. The features are ranked from most (1) to least (19) important.

definition, age, and number of lesions. Gender and periosteal reaction counted for meningioma. Hemangioma diagnosis relied on shape, soft-tissue mass, and periosteal reaction. Fibrous dysplasia diagnosis depended on diameter and age, as selected by the LR equation.

Neural Network

In the malignancy NN, the ordered variables were headed by age and edge definition (Table 5). Exclusion of the variables, from bottom to top, provided a similar A_z until the exclusion of edge definition. When edge definition was excluded, a significant decrease in the A_z was found in the NN with age as the only feature (0.8131 ± 0.0331 versus 0.9129 ± 0.0031, $P < 0.05$). Therefore, the minimal variables needed for a confident diagnosis of malignancy are age and edge definition. Table 5 also summarizes the ranked features for the histologic NN. Table 6 shows the A_z in the diagnosis NN for each disease. The elimination of the features did not decrease the A_z until age was removed; then, the NN provided a decrease in the A_z for Langerhans' cell histiocytosis ($P < 0.001$), meningioma ($P < 0.001$), and osteoma ($P < 0.05$), with no significant A_z loss in the remaining diagnoses. This means that at least three variables—age, marginal sclerosis, and matrix—are needed for the diagnosis NN.

Comparison

The comparison of A_z for malignancy between LR and NN models was not statistically significant (0.9392 versus 0.9617, $P = 0.2$). However, for histologic diagnosis, the NN

TABLE 6. Results, Expressed as A_z, of the Diagnoses Neural Network with All Input Features, Only Matrix, Matrix and Marginal Sclerosis, and Matrix, Marginal Sclerosis, and Age*

Diagnosis	All input nodes	Matrix	Matrix, marginal sclerosis	Matrix, marginal sclerosis and age
Langerhans' cell histiocytosis	0.9854 ± 0.018	0.6385 ± 0.0591 (<0.001)	0.7949 ± 0.051 (<0.001)	0.9852 ± 0.0153 (NS)
Osteoma	0.9998 ± 0.002	0.8808 ± 0.0484 (<0.05)	0.9431 ± 0.0348 (NS)	0.9361 ± 0.0368 (NS)
Epidermoid cyst	0.9961 ± 0.002	0.6347 ± 0.0682 (<0.001)	0.9396 ± 0.0358 (NS)	0.9580 ± 0.0302 (NS)
Metastasis	0.9944 ± 0.003	0.5225 ± 0.0692 (<0.001)	0.7663 ± 0.0633 (<0.05)	0.9446 ± 0.0351 (NS)
Meningioma	0.9771 ± 0.025	0.5921 ± 0.0768 (<0.001)	0.7297 ± 0.0726 (<0.001)	0.8580 ± 0.0581 (NS)
Hemangioma	0.9409 ± 0.004	0.6558 ± 0.0804 (<0.001)	0.8460 ± 0.0644 (NS)	0.9409 ± 0.0428 (NS)
Fibrous dysplasia	0.9601 ± 0.041	0.8804 ± 0.071 (NS)	0.9070 ± 0.0639 (NS)	0.7937 ± 0.086 (NS)

NS: nonsignificant decrease of A_z compared with all-features diagnoses neural network.
* At least three variables are needed for a correct performance in the diagnoses neural network. P values are in parentheses.

TABLE 7. Comparison of Logistic Regression and Neural Networks and Receiver Operating Characteristic Areas

	Logistic regression		NN		
Group	$A_z \pm SD$	S/S*	$A_z \pm SD$	S/S*	*P*
Malignancy	0.9392 ± 0.051	79.7/95.2	0.9515 ± 0.042	86.6/90.3	0.2
Langerhans' cell histiocytes	0.9117 ± 0.071	92.4/46.9	0.9854 ± 0.018	95.9/97.8	0.001
Osteoma	0.9029 ± 0.037	81.5/83	0.9998 ± 0.002	98.1/98.1	<0.001
Epidermoid cyst	0.5441 ± 0.781	77.7/12.3	0.9961 ± 0.002	99.9/97.8	<0.001
Metastasis	0.9192 ± 0.091	82.7/85	0.9944 ± 0.003	98.8/98.8	0.007
Meningioma	0.6903 ± 0.078	39.9/87.2	0.9771 ± 0.025	95.6/98.8	<0.001
Hemangioma	0.7211 ± 0.075	75.4/53.4	0.9409 ± 0.004	93/93	<0.001
Fibrous dysplasia	0.8372 ± 0.049	98/7.5	0.9601 ± 0.041	98/91	<0.001

A_z: area under the ROC curve; SD: standard deviation; S/S*: sensitivity/specificity (%).

yielded a significantly higher A_z than LR (Table 7). The sensitivity and specificity of each model at the highest point of accuracy in every ROC curve was also calculated. Although the ROC curve can provide these values for every threshold cutoff point, we chose the highest point because it was the most efficient one (prior probability = 0.5). An example of a lesion correctly classified by both methods is shown in Figure 1, and a lesion identified only by the NN is displayed in Figure 2.

Discussion

With plain films, the incidence of nonanatomic variations presenting as calvarial lytic lesions was about 7%.[18] However, the precise incidence of calvarial lesions has not been estimated using CT or magnetic resonance imaging. The distribution of the most common lesions encountered in our series differs somewhat from the reported incidence[3,4] because of different selection criteria. A possible bias is that the performance of our NN designs should be tested on a more generalized set of lesions.[18–21]

There is general agreement that the appearances of most calvarial lesions on CT and MR scans are difficult to interpret (Keesling C, presented at the 1995 annual meeting of the Radiological Society of North America). Although some authors claim that most lesions are predictable according to the patient's age, the radiologic diagnosis of these lesions is usually uncertain because of overlapping radiologic and clinical appearances.[22] As with other lesions in the skele-

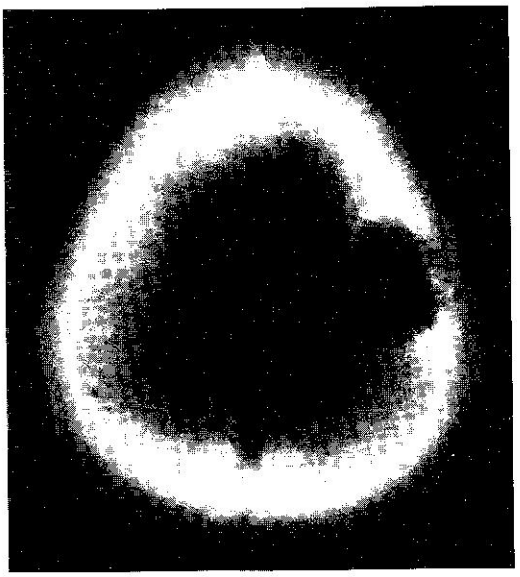

Figure 1. Langerhans' cell histiocytosis. Computed tomography with bone window shows a lytic lesion in the left parietal bone. Logistic regression and neural networks correctly identified it as Langerhans' cell histiocytosis.

Figure 2. Hemangioma. Computed tomography with bone window, detail of a right parietal lesion. Logistic regression failed in the diagnosis, properly classified by the neural network.

ton,[23] a more accurate diagnostic modality is clearly needed to avoid the implications of misdiagnosis.

Clinical prediction rules are useful in several fields, including diagnostic radiology, to aid in the medical decision-making process.[24] One of the pioneers in computer-assisted diagnosis of focal bone lesions was Lodwick and his bayesian approach.[6] The main drawbacks of this analysis is that it does not learn from new cases: it requires a very large number of patterns to provide a statistically valid probability matrix,[25] and it performs better with common diseases.[26] Logistic regression has emerged as the statistical analysis of choice for predicting dichotomous outcomes.[27] Recently, the use of vector analysis[7] and NNs[8] has been also advocated in recognizing radiologic patterns.[28,29] One of the arguments against the widespread use of NNs is that their results are rarely compared with statistical methods to test relative performance.[27] The best method to determine whether NNs are better for a given problem than a statistical procedure is to compare their performance in the same training and validation groups.[27]

We found that a simple three-layer back-propagation NN can be trained to distinguish between patterns of calvarial lesions associated with benignity and malignancy and the most common cranial vault lesions. The similar performance of the NN and LR models to differentiate benign from malignant lesions is probably the result of the good adaptation of the statistical model to the sample size. The most relevant variables selected by both the LR and NN models were age and edge definition. Age is recognized as the single most important clinical variable in the diagnosis of focal bone lesions.[25] Edge definition is one of the classic radiologic variables that determines the aggressiveness of the lesion. However, in a study of calvarial lytic lesions, edge definition was not important for diagnosis, although no statistical method was used.[18]

As the sample size in the histologic diagnosis groups decreases, the predictive ability of the LR equation also declines. Because all classifying methods depend on training, better results in larger samples must be expected. For this reason, the diagnoses with fewer than ten occurrences were discarded.[8] Surprisingly, the NN performed extremely well compared with LR, even with the smaller groups. The leave-one-out method, used to generalize the LR analysis, may be responsible for the low performance of the statistical model.

There were four variables needed for both the malignancy and histologic diagnosis NNs: age, edge definition, marginal sclerosis, and matrix. These features are known to be of great relevance in clustering bone lesions.[6] It is beyond the scope of this study to describe the most important single features in each diagnosis; in fact, we did not build different NNs for each entity. This was one of the shortcomings of LR, which required different equations and variables for each diagnosis.

The superiority of NNs over LR reflects the network's ability to obtain complex relations between variables, even in small samples. This result is achieved by NNs as a result of their parallel architecture, a cardinal difference with the statistical methods.[30] The best performance of the network could not be the result of overtraining, because the methods used in its design were specifically chosen to avoid this bias. Further, although overtraining is one of the commonly assumed explanations for the best performance of the NNs compared with other classifying methods, statistical methods are also prone to overfitting.[27]

An analysis using NNs on solitary bone lesions was reported by Reinus et al,[8] but that study lacked comparison with a statistical counterpart. This paper had no historical data but indicated that little training is needed for patterns with small variation. That model was slightly more sensitive than ours (89.8% versus 86.6%) but much less specific (75.1% versus 90.3%). These differences may result from the fact that their series included all skeletal lesions exclusively studied with radiographs. Because these films cannot demonstrate all the radiologic characteristics of a lesion, several input variables obtained in our series from CT images were not considered in their work. Although they used cross-validation for NN training, the different performance must be related to the wide spectrum of lesions studied. Their good results in small samples with distinctive appearances imply that the quality of the data seems more important than the quantity. This is shown by the low performance of NNs in larger samples with variegated radiographic appearances.

The main problem in classifying bone lesions, including calvarial ones, is the considerable overlapping of imaging findings;[19] this renders impractical most techniques used for their classification, such as vectorial analysis.[7] Neural networks have potential advantages over traditional multivariate analysis in classifying data that have nonlinear boundaries among different classes. The superiority of NNs over statistical approaches seems clear in complex pattern-recognition problems such as those in this study, even with small data sets.

In conclusion, NNs can be used to diagnose most lesions in the cranial vault.

References

1. Wecht DA, Sawaya R. Lesions of the calvaria: Surgical experience with 42 patients. Ann Surg Oncol 1997;4:28–36.
2. Arana E, Latorre FF, Revert A, et al. Intradiploic epidermoid cysts. Neuroradiology 1996;38:306–311.
3. Ruge JR, Tomita T, Naidich TP, Hahn YS, McLone, DG. Scalp and calvarial masses of infants and children. Neurosurgery 1988;22: 1037–1042.
4. Steinmeier R, Huk W, Fahlbusch R. Tumoren und raumfordernde Fehlbildungen der Schädelkalotte. Radiologe 1988;28:134–140.
5. Boone JM. Neural networks at the crossroads. Radiology 1993;189: 357–359.
6. Lodwick GS. A probabilistic approach to the diagnosis of bone tumors. Radiol Clin North Am 1965;3:487–497.

7. Reinus W, Wilson A. Quantitative analysis of solitary lesions of bone. Invest Radiol 1995;30:427–432.
8. Reinus WR, Wilson AJ, Kalman B, Kwasny S. Diagnosis of focal bone lesions using neural networks. Invest Radiol 1994;29:606–611.
9. Piraino D, Richmond B, Uetani M. Application of an artificial neural network in radiographic diagnosis. J Digit Imaging 1991;4:226–232.
10. Press SJ, Wilson S. Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 1978;73:699–705.
11. Hinton G. How neural networks learn from experience. Sci Am 1992; 207:144–151.
12. Miller A, Blott B. Review of neural network application in medical imaging and signal processing. Med Biol Eng Comput 1992;30: 449–464.
13. Fisher RE, Scott J, Palmer E. Neural networks in ventilation-perfusion imaging. Part I. Effects of interpretative criteria and network architecture. Radiology 1996;198:699–706.
14. Penny W, Frost D. Neural networks in clinical medicine. Med Decis Making 1996;16:386–398.
15. LABROC1 [computer program] developed by Metz C, Herman B, Wang P, et al. Department of Radiology, University of Chicago, 1993; IBM-PC Fortran *URL* ftp://random.bsd.uchicago.edu.
16. Dorfman D, Berbaum K, Metz C. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992;27:723–731.
17. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.
18. Thomas J, Baker H. Assessment of roentgenographic lucencies of the skull: A systematic approach. Neurology 1975;25:99–106.
19. Greenfield G, Arrington J. Metastatic tumors, differential diagnosis and bone marrow. In: Greenfield G, Arrington J, eds. Imaging of bone tumors: A multimodality approach. Philadelphia: JB Lippincott, 1995;333–444.
20. Dahlin D. Introduction and scope of study. In: Dahlin D, ed. Bone tumors: General aspects and data on 6,221 cases. 3d ed. Springfield: Charles C. Thomas, 1978;3–16.
21. Kilpatrick SE, Wenger DE, Gilchrist GE, Shives TC, Wollan PC, Unni KK. Langerhans' cell histiocytosis (histiocytosis X) of bone. Cancer 1995;76:2471–2484.
22. Lane B. Erosions of the skull. Radiol Clin North Am 1974;12:257–282.
23. Di Rocco C, Jannelli A, Fileni A, Moschini M. Solitary osteolytic lesions of the skull vault in children. Neurochirurgia Stuttg 1982; 25:57–61.
24. D'Orsi C, Getty D, Swets R. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. Radiology 1992;184:619–622.
25. Lodwick G, Hodes P. Radiologic concepts. In: Lodwick G, Hodes P, eds. The bones & joints: An atlas of tumor radiology. Chicago: Year Book Medical Publishers, 1973;1–82.
26. de Dombal F. Computer-aided diagnosis and medical decision support are not synonymous. Meth Inform Med 1995;34:361–368.
27. Tu J. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996;49:1225–1231.
28. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. Radiology 1993;187: 81–87.
29. Baker JA, Kornguth PJ, Lo JY, Floyd CE Jr. Artificial neural network: Improving the quality of breast biopsy recommendations. Radiology 1996;198:131–135.
30. Kattan M, Becj J. Artificial neural networks for medical classification decisions. Arch Pathol Lab Med 1995;119:672–677.