Neuroradiology (1995) 37: 89–93

P. S. Christy
O. Tervonen
B. W. Scheithauer
G. S. Forbes

Use of a neural network and a multiple regression model to predict histologic grade of astrocytoma from MRI appearances

Received: 21 August 1993
Accepted: 3 March 1994

This work was presented as a poster at the Society for Magnetic Resonance in Medicine Tenth Annual Meeting, San Francisco, California, USA, 1991

P. S. Christy · O. Tervonen (✉)[1] ·
B. W. Scheithauer · G. S. Forbes
Mayo Clinic, 200 First Street Southwest,
Rochester, MN 55905, USA

Present address:
[1] Department of Diagnostic Radiology,
Oulu University Central Hospital,
Kajaanintie 50, SF-90220 Oulu, Finland

Abstract Several MRI features of supratentorial astrocytomas are associated with high histologic grade by statistically significant p values. We sought to apply this information prospectively to a group of astrocytomas in the prediction of tumor grade. We used 10 MRI features of fibrillary astrocytomas from 52 patient studies to develop neural network and multiple linear regression models for practical use in predicting tumor grade. The models were tested prospectively on MR images from 29 patient studies. The performance of the models was compared against that of a radiologist. Neural network accuracy was 61 % in distinguishing between low and high grade tumors. Multiple linear regression achieved an accuracy of 59 %. Assessment of the images by a radiologist yielded 57 % accuracy. We conclude that while certain MRI parameters may be statistically related to astrocytoma histologic grade, neural network and linear regression models cannot reliably use them to predict tumor grade.

Key words Astrocytoma · Neural network · Magnetic resonance imaging · Brain neoplasms

Introduction

Studies have been undertaken to correlate the histologic appearance of cerebral glial neoplasms with MRI features such as mass effect, tumor heterogeneity and cerebral edema [1–3] (Fig. 1). With the knowledge of the relationships of such features to tumor grade, neuroradiologists successfully predicted tumor grade on a three tiered grading scheme approximately 80 % of the time using "gestalt" impressions [1]. It might therefore be expected that an objective model which employs the statistical relationships between MRI features and tumor grade could be developed to predict astrocytoma tumor grade with similar reliability. Our objective is to create models using linear and nonlinear calculation schemes to predict histologic grade consistently and objectively.

Multiple linear regression model

Multiple linear regression provides a starting point for developing a model to predict tumor grade. Each feature is graded and plotted against tumor grade and the slopes of the individual plots used to develop a formula which predicts tumor grade. While this model is useful in many cases, it has important limitations. First, it requires a linear relationship between the MRI parameters and histologic grade, and second, that the features be independently predictive of grade. When a radiologist offers an impression regarding the implications of a set of imaging features, the interpretation is not restricted by such constraints and a model so restricted might be reasonably expected to fail in this role.

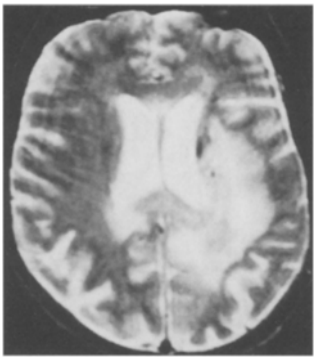

Fig. 1 MRI of a representative cerebral astrocytoma, illustrating moderate tumor heterogeneity, mild mass effect, moderate edema, poor circumscription and vascular flow voids. These features are all suggestive of high histologic grade

Neural networks

Neural network models provide an analytic method which requires neither linearity nor independence of features. Such models are based loosely upon neurons [4]. A sample architecture used in this study is illustrated in Fig. 2, and each of the neurons or processors, indicated in the illustration can be generically represented as in Fig. 3. Numerical inputs are applied to the left side of the network. Each of the neurons in the input layer then calculates an output based upon its individual input using a sigmoid transfer function which serves as an approximation of the all or none behavior exhibited by biological neurons:

$$y_j = 1/(1 + e^{-x_j}) \quad (1)$$

The output from each input neuron is multiplied by its individual weight factor and sent to each neuron in the next layer. Each neuron in the second layer linearly sums its inputs and produces an output which is applied to the same transfer function, then multiplied by its own weight factor and sent to the next layer. The output layer performs identical calculations to produce the final output from the neural network. The weighting factors contain the information in the neural network and are calculated using the back propagation of errors method, which employs the generalized delta learning rule [4]. This is an iterative process by which input training sets are applied to the neural network and outputs calculated. The actual output is compared to the desired output and the weight factors adjusted to increase the probability of producing a correct output on the next iteration. Training is halted when an arbitrary percentage of correct outputs is obtained (training tolerance parameter). The time required to train the neural network is altered by several other parameters which include the steepness of the sigmoid transfer function (transfer function gain), the amount of error tolerated between neural network output and target value before revising the connection strength weightings (training tolerance), the amount by which the connection strengths are changed with each iteration (learning rate), and the direction of the weight changes in relation to the previous iteration (momentum).

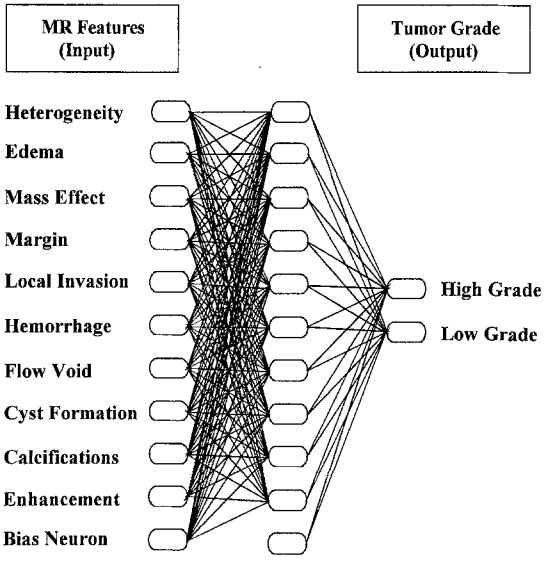

Fig. 2 Neural network architecture. Individual neurons or processors are represented by ovals. Bias neurons which provide a constant +1 input are placed in the input and second layers so that the network may produce a nonzero output in the case of all zero inputs. There are dual outputs. For a high grade tumor, the high grade neuron produces a +1 output and the low grade output neuron produces a −1 output. The values are reversed in the case of a low grade tumor

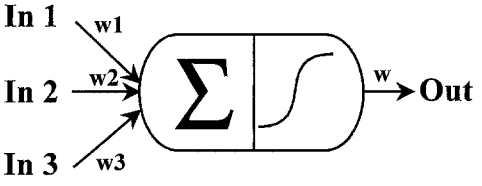

Fig. 3 Generic representation of a neural network's individual processors or neurons. Weighted inputs are summed linearly then applied to a sigmoid transfer function which simulates the all-or-none behavior of biological neurons. The output is weighted and then sent to the next layer of processors

Methods

Patient selection

Patients having biopsy-proven fibrillary astrocytomas whose pathological specimens and MRI studies were available were selected for development of the multiple linear regression and neural network models. Only those with our MR images obtained prior to biopsy, radiation therapy or surgery were included: 52 who underwent biopsy between 1987 and 1990. The multiple linear regression and neural network models were tested prospectively on an additional 29 patients selected according to the same criteria between 1990 and 1991.

Table 1 Neural network definition

Input data	10 MRI astrocytoma features, graded 0–4, scaled from –1–+1 for input, average values for missing data
Output data	"High grade", range –1–+1, "low grade", range –1–+1
Architecture	3 layer, 10 neuron hidden layer, feedforward, fully connected
Training algorithm	Back propagation, generalized delta learning rule
Transfer function	Sigmoid, 0 minimum, +1 maximum, gain 1.0
Initial connection weights	Gaussian distribution, range –0.5–+0.5, standard deviation 0.17
Training tolerance	10 %
Learning rate	1.0
Momentum	0.0

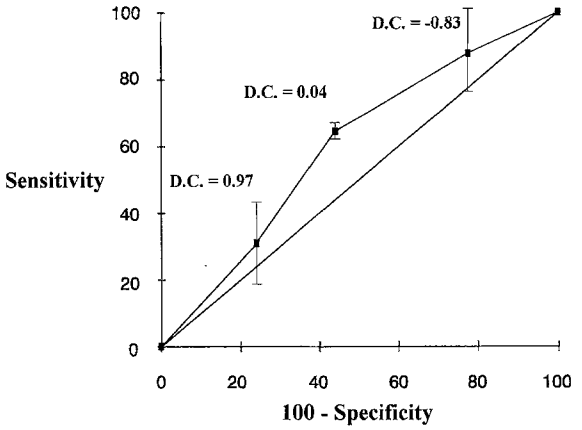

Fig. 4 ROC for neural network performance in detection of high-grade astrocytomas (*D. C.* decision criterion, Figs. 4–6)

Imaging technique

All images were obtained at 1.5 Tesla with a 192 × 256 or 256 × 256 acquisition matrix, a 24 cm field of view, 1 or 2 excitations and a 5 mm slice thickness. Partial saturation (T1-weighted) images were obtained using a spin-echo (SE) pulse sequence with repetition time (TR) 400 or 600 ms and echo time (TE) 20 ms, or an inversion recovery pulse sequence with TR 2000 ms, an inversion time (TI) 700 ms and TE 20 ms. T2-weighted images were obtained using an SE pulse sequence with TR 2000 or 2250 ms and TE 80 ms. Proton density images were obtained from a first echo at 20 ms during the T2-weighted sequence. Intravenous gadolinium enhancement was employed in 42 of 52 model development cases and in 24 of 29 prospective cases.

Grading of MRI features

All cases used in the development of the mathematical models were reviewed and consensus obtained between two neuroradiologists (G. S. F., O. T.), regarding the grading of 10 MRI features. Neither radiologist was aware of clinical history or biopsy results prior to grading, although both were aware that the cases were astrocytomas. The features were (1) tumor mass heterogeneity, (2) surrounding parenchymal edema, (3) mass effect, (4) indistinctness of the tumor margin, (5) local invasion of adjacent structures, (6) hemorrhage within the tumor, (7) presence of one or more flow voids, (8) presence of cysts within the mass, (9) calcification and (10) degree of contrast enhancement. Features were graded on a scale from 1 to 4, 0 indicating that the feature was absent. For the prospective portion of the study, the 10 features were graded for each patient by a radiologist (O. T.) who also predicted tumor grade. The radiologist's prediction of tumor grade employed knowledge of the statistical relationships between MRI features and histologic grade determined in a prior study in which contrast enhancement and degree of edema were most strongly directly related to tumor grade [3].

Histopathologic grading

Hematoxylin and eosin stained biopsy specimens were graded on a scale from 1 to 4 by a neuropathologist (B. W. S.) in a blinded fashion. Grading was performed by the same pathologist in the development of the models and in the prospective study. Grading criteria conformed to those of the new World Health Organization (W. H. O.) classification of tumors of the central nervous system [5, 6]: grade 1 tumors lack atypia, which is shown by grade 2 lesions; grade 3 tumors display atypia and mitotic figures and grade 4 tumors show in addition endothelial proliferation or necrosis, or both.

Mathematical models and performance analysis

The multiple linear regression model was developed using a statistics software package. Parameters whose values were unavailable (for example, the degree of contrast enhancement in patients who had only unenhanced studies) were set to the average value of the rest of the data set for that parameter. The equation so derived is as follows:

$$\text{Grade} = 1.196 - 0.028^{*}\ \text{heterogeneity} + 0.424^{*}\ \text{Edema} + 0.009^{*}\ \text{Mass\_Effect} + 0.330^{*}\ \text{margin\_sharpness} - 0.117^{*}\ \text{anatomic\_invasion} + 0.550^{*}\ \text{hemorrhage} + 0.004^{*}\ \text{flow\_void} - 0.172^{*}\ \text{cyst\_formation} + 0.382^{*}\ \text{calcification} + 0.139^{*}\ \text{contrast\_enhancement} \quad (2)$$

The neural network model was developed using commercially available software; the parameters employed are shown in Table 1.

Receiver operating characteristic (ROC) curves were used to evaluate performance of the grading methods. For neural network performance, output from the high-grade output neuron was evaluated at decision criteria throughout the range of the output neuron (–1 to +1). If the output was less than the decision criterion then the tumor was termed low grade; otherwise it was termed high grade. The output from the linear regression model was evaluated at decision criteria from 1.0 to 5.5 and the radiologist's predictions were from 0.5 to 4.5.

Results

Neural network performance is shown in Fig. 4, using an ROC curve. At all decision criteria the algorithm performs better than random guesses. At a decision criteri-

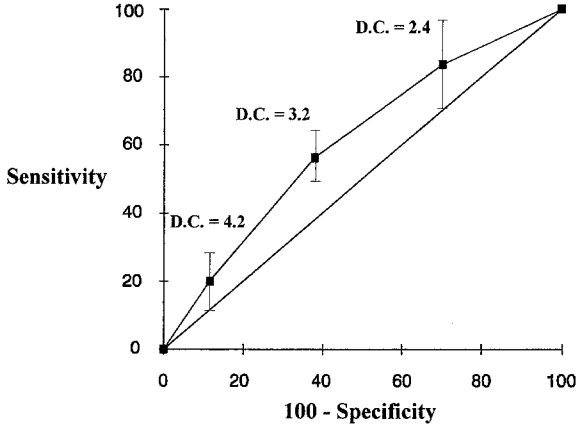

Fig. 5 ROC for multiple linear regression performance in detection of high-grade astrocytomas

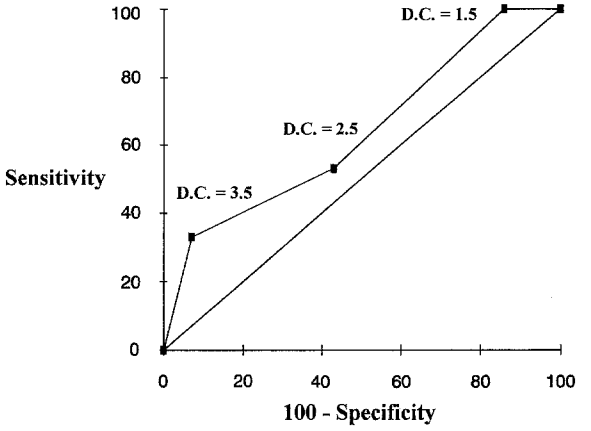

Fig. 6 ROC for radiologist's performance in detection of high-grade astrocytomas

on of 0.04 the neural network achieved an accuracy of 61 %. Figure 5 shows the ROC curve for the multiple linear regression model. Performance was again better than random, but not significantly different from that of the neural network. At a decision criterion of 3.2, and accuracy of 59 % was obtained. Figure 6 shows the ROC curve for the radiologist; his performance was comparable to that of the other two methods, with an accuracy was 57 %. Table 2 summarizes these results.

Discussion

Information regarding the grade of malignancy is essential for optimal clinical management of patients with astrocytomas. The ability of a radiologist to estimate tumor grade is based upon experience regarding the association of certain parameters such as edema and the degree of contrast enhancement with increasing malignancy. Experts have been able to make this determination with 80 % accuracy [1]. This is, however, not reliable enough for clinical decision marking nor is it completely objective or reproducible in general practice, as our study has shown. We therefore sought to provide an objective, reproducible means of relating MRI features to histologic grade.

Table 2 Comparative performance of grading algorithms in 29 cases (%)

	Neural network	Multiple linear regression	Radiologist
Accuracy	61	59	57
Sensitivity	64	56	53
Specificity	56	62	55

Multiple linear regression provides a first step toward the goal of histologic grade prediction and in this study was shown to be as reliable as the other means of interpretation. Its major problem, however, is that by definition it requires that the input features be linearly related to the output parameter. The margin indistinctness parameter is a good example of this problem. The coefficient of 0.330 used in the regression model implies that it is a relatively important parameter. Its p value, however, is not statistically significant, indicating that the heavy weighting is artifactual. The model might be improved somewhat by eliminating this parameter altogether. By doing this, however, we may be losing important information that may have some, as yet unknown, nonlinear relationship to the desired output parameter.

Neural network models provide an objective analytic method which requires neither linearity of data nor independence of parameters and as such, may provide a tool for radiologists in establishing diagnoses as well as in making diagnoses more specific. For example, if an input parameter has no detectable relationship to the output, its neural network connection strengths will be set to values near zero, while very complex, nonlinear, interdependent relationships can be represented by nonzero connection strengths. This type of analysis has recently been used successfully in interpretation of neonatal chest radiographs and differential diagnosis of interstitial lung diseases [7, 8]. Gross and coworkers [7] found good agreement between a neural network and radiologists in diagnosing neonatal lung disease. In work by Asada et al. [8], good neural network performance was obtained in distinguishing between 9 types of interstitial lung diseases on the basis of 20 clinical and radiographic parameters.

The present study indicates that neural network analysis can differentiate between low- and high-grade astrocytomas with a consistency similar to that of a radiologist but that it cannot significantly increase accuracy in predicting tumor grade. The accuracy of the neural network was slightly greater but not statistically

different from those of the multiple linear regression model or the radiologist's interpretation. Neural networks thus can very effectively provide an accurate output if there is reasonable correlation between input and output data (image parameters diagnosis), but accuracy is limited by the data available. In the case of chest radiographs, the correlation between radiographic parameters is sufficient to provide accurate results. However, in the case of astrocytomas, the relationship between MRI features and tumor grade is not sufficiently strong to allow means of tumor grade determination which are more accurate than those already available. This relates to the fact that while MR tissue characteristics may serve as markers of histologic parameters there are currently no MRI parameters which directly reflect tissue histology. Finally, stereotactic biopsy, performed in some of the case, is limited to sampling a relatively small region within a given tumor and may miss regions of more pronounced anaplasia.

Acknowledgement This work was funded under NIH Training Grant HL07269-14R.

References

1. Dean BL, Drayer BP, Bird CR, Flom RA, Hodak JA, Coons SW, Carey RG (1990) Gliomas: classification with MR imaging. Radiology 174: 411–415
2. Kelly PJ, Daumas-Duport C, Scheithauer BW, Kall BA, Kispert DB (1987) Stereotactic histologic correlations of computed tomography- and magnetic resonance imagingdefined abnormalities in patients with glial neoplasms. Mayo Clin Proc 62: 450–459
3. Tervonen O, Forbes GS, Scheithauer BW, Dietz MJ (1992) Diffuse "fibrillary" astrocytomas: correlation of MRI features with histopathologic parameters and tumor grade. Neuroradiology 34: 173–178
4. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructures of cognition, vol 1. MIT Press, Cambridge, pp 675–695
5. Kleihues P, Burger PC, Scheithauer BW (1991) World Health Organization classification of tumors of the central nervous system. Springer, New York
6. Daumas-Duport C, Monsaigneon V, Blond S, Munari C, Musolino A, Chodkiewicz JP, Missir O (1987) Serial stereotactic biopsies and CT scan in gliomas: correlative study in 100 astrocytomas, oligoastrocytomas and oligodendrocytomas. J Neuro-Oncol 4: 317–328
7. Gross GW, Boone JM, Greco-Hunt V, Greenberg B (1990) Neural networks in radiologic diagnosis II: interpretation of neonatal chest radiographs. Invest Radiol 25: 1017–1023
8. Asada N, Doi K, MacMahon H, et al (1990) Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study. Radiology 177: 857–860