Prediction of Radiation Induced Liver Disease Using Artificial Neural Networks

Ji Zhu¹,³*, Xiao-Dong Zhu²*, Shi-Xiong Liang², Zi-Yong Xu¹, Jian-Dong Zhao¹,³, Qi-Fang Huang², An-Yu Wang², Long Chen², Xiao-Long Fu¹,³ and Guo-Liang Jiang¹,³

¹Department of Radiation Oncology, Fudan University Cancer Hospital, Shanghai, ²Department of Radiation Oncology, Cancer Hospital, Guangxi Medical University, Nanning and ³Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Received June 7, 2006; accepted August 10, 2006; published online October 26, 2006

Objective: To evaluate the efficiency of predicting radiation induced liver disease (RILD) with an artificial neural network (ANN) model.

Methods and Materials: From August 2000 to November 2004, a total of 93 primary liver carcinoma (PLC) patients with single lesion and associated with hepatic cirrhosis of Child–Pugh grade A, were treated with hypofractionated three-dimensional conformal radiotherapy (3DCRT). Eight out of 93 patients were diagnosed RILD. Ninety-three patients were randomly divided into two subsets (training set and verification set). In model A, the ratio of patient numbers was 1:1 for training and verification set, and in model B, the ratio was 2:1.

Results: The areas under receiver-operating characteristic (ROC) curves were 0.8897 and 0.8831 for model A and B, respectively. Sensitivity, specificity, accuracy, positive prediction value (PPV) and negative prediction value (NPV) were 0.875 (7/8), 0.882 (75/85), 0.882 (82/93), 0.412 (7/17) and 0.987 (75/76) for model A, and 0.750 (6/8), 0.800 (68/85), 0.796 (74/93), 0.261 (6/23) and 0.971 (68/70) for model B.

Conclusion: ANN was proved high accuracy for prediction of RILD. It could be used together with other models and dosimetric parameters to evaluate hepatic irradiation plans.

Key words: primary liver carcinoma – three-dimensional conformal radiation therapy – radiation induced liver disease – artificial neural networks

INTRODUCTION

Primary liver carcinoma (PLC) is a common cancer and is the fifth most common malignancy in men and the eighth in women worldwide. The number of new cases is estimated to be 564,000 per year, including 398,000 in men and 166,000 in women (1). Especially in China and other Asia countries, PLC has a higher incidence because of the prevalence of hepatitis B (2,3).

Surgical resection is the choice of treatment for early stages of PLC with 5-year survival of 30–50% (4,5). However, most PLC patients are technically unresectable or medically inoperable at the time of diagnosis. As a palliative treatment, radiotherapy had been used to treat PLC since the middle of the twentieth century, but the outcome was quite disappointing in the early years (6). Since the last decade, three-dimensional conformal radiation therapy (3DCRT) had been introduced to PLC radiotherapy, which yielded encouraging results with 2-year survival rates of 33–41% (7–9). As irradiation doses to tumors increased, another problem was brought forward, i.e. higher incidence of radiation induced liver disease (RILD). As reported recently by our study and others, RILD was an almost fatal irradiation induced complication associated with as much as 76% of mortality (9,10). Unfortunately, there has been no effective treatment for RILD to date, so it is most critical to prevent RILD from occurring. Thus, questions were raised as to what the RILD risk variables were, what was hepatic radiation tolerance and how can one predict RILD. In the literature, some studies have been carried out to predict RILD probability by a normal tissue complication probability (NTCP) model, or to look for hepatic tolerance in terms of three-dimensional dosimetric parameters (11,12). However, a few authors tried a statistical approach using artificial neural networks.
networks (ANN) to predict the probability of radiation injuries. ANN mimics the biological neural network and is carried out by computers. It was originally described by McCulloch and Pitts in 1943 (13), but only after the 1990s has ANN achieved successful clinical application after a gradual learning process. In the current study we investigated ANN to see whether it could correctly predict RILD for PLC patients treated by 3DCRT, or combined with transcatheter arterial chemoembolization (TACE).

**MATERIALS AND METHODS**

**Patients’ Selection**

The criteria of patient selection were as follows: (i) PLC with single lesion; (ii) associated with hepatic cirrhosis of Child–Pugh grade A; (iii) treated by 3DCRT with hypofractionated irradiation; (iv) 3DCRT dosimetric parameters available; (v) minimum follow-up time of >4 months for patients without RILD.

In the patient database of the Cancer Hospital at Guangxi Medical University, Nanning, China, from August 2000 to November 2004, we found 93 eligible patients. All of them were technically unresectable as a result of locally advanced lesions, medically inoperable because of associated severe hepatic cirrhosis, or unwilling to undergo surgery. Their clinical characteristics were shown in Table 1. There were 85 cases of male and eight female with median age of 45 years. Fifty-nine patients were staged as T3N0M0 (UICC/AJCC, 1997), and 34, T4N0M0. All patients received 3DCRT, but 32 patients, got TACE prior to irradiation with a median TACE session of two (range, 1–4). The interval was 2–3 weeks between TACE and 3DCRT.

**Radiation Therapy Technique**

The technique of 3DCRT had been described earlier (10). Briefly, prior to CT simulation, patients were immobilized by a stereotactic body frame and trained to breathe as shallowly as possible with the assistance of a tight abdominal belt to limit breathing. All patients underwent a planning CT scan to facilitate three-dimensional treatment planning using Topslane planning system (Topslane Medical Corp., Shanghai, China). The gross tumor volume (GTV) was delineated on a CT scan. Planning target volume (PTV) was determined by adding 0.5–1.5 cm to GTV. Liver motion was taken into account and additional margins were added to account for lesion motion owing to respiration. With the help of beam’s eye view, four to eight coplanar or non-coplanar fields were designed. The dose was prescribed to isocenter as 100% without inhomogeneity tissue correction, and PTV should be covered by 90% of isodose curve. Hypofractionated irradiation was applied, i.e. three fractions per week (Monday, Wednesday, Friday). The fraction size ranged from 4 to 8 Gy, and the median total dose was 53 Gy (range 40–68 Gy). However, the principle of ‘trial and error’ had been observed during the study. At the very beginning, we applied a large fraction, 7–8 Gy, but more acute adverse effects and complications were noticed. Therefore, to improve patients’ tolerance we decreased fraction size gradually, but the total doses still remained high. Our late experience demonstrated that 4–5 Gy/fraction, three fractions per week to a total of about 50 Gy was tolerable. Thus, this fractionation was used for all patients in the later period of the study.

**TACE**

TACE was performed with infusion of a mixture of 5 ml of lipiodol, 50–60 mg/m² of epiadriamycin, and 30–40 mg/m² of cis-platinum, or 5 ml of lipiodol, 6–7 mg/m² of mitomycin C and 10–15 mg/m² of 10-hydroxycamptothecin (HTCP), followed by gelfoam embolization (Jinling Pharmaceutical Co. Ltd., Nanjing, China).

**Definition of RILD**

RILD was defined by Lawrence’s criterion (14,15). It occurred within 4 months after the completion of irradiation and manifested as either anicteric elevation of alkaline phosphatase (AKP) level of at least twice the upper normal level and non-malignant ascites (classic RILD), or elevated transaminases of at least five times the upper limit of the normal or of pre-treatment level (Grade 3 or 4 hepatic toxicity of CTC) (non-classic RILD). The diagnoses of hepatic injuries should be in the absence of documented progressive liver carcinoma.

**Follow-up**

All patients were followed up at least 4 months after irradiation, except RILD occurring within 4 months. Each visit included history, physical examination, complete blood count, serum AFP, blood chemistry, abdominal ultrasound and CT scan.

### Table 1. Patient demographics and clinical characteristics with or without RILD

<table>
<thead>
<tr>
<th>Variables</th>
<th>Without RILD vs. With RILD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD, years)</td>
<td>47.1 ± 11.6 versus 45.3 ± 6.9</td>
<td>0.665*</td>
</tr>
<tr>
<td>Sex (male:female)</td>
<td>77:8 versus 8:0</td>
<td>1.000b</td>
</tr>
<tr>
<td>TACE (yes: no)</td>
<td>31:54 versus 1:7</td>
<td>0.256b</td>
</tr>
<tr>
<td>T stage* (T3:T4)</td>
<td>57:28 versus 2:6</td>
<td>0.048b</td>
</tr>
<tr>
<td>PVT (yes: no)</td>
<td>16:69 versus 3:5</td>
<td>0.353b</td>
</tr>
</tbody>
</table>

RILD, radiation induced liver disease; SD, standard error; TACE, transcatheter arterial chemoembolization; PVT, portal vein thrombus, T stage*, UICC/AJCC staging system (1997).

*a* test.

*b* Fisher test.
ANN MODELS AND STATISTICS

ANN models were constructed by use of neural-network software (Statistic Neural Networks, version 4.0). Multilayer perceptrons (MLP) was applied in this study. The architecture of MLP consisted of three layers, the input, hidden and output layers (Fig. 1). The input layer simply fed information as well as related predictive factors into the network, while nodes in the hidden and output layers processed information. During the training, the corresponding known outputs of the system were held in the output nodes to compare with the results produced by the network. The nodes in the hidden layer had no prescribed initial values and helped to allow complex relationships between the input and output nodes to evolve. Information was transported from the input layer to the output layer by calculating the sum on each node, which was derived from combining all the nodes in the previous layer.

MLP incorporated a non-linear activation function, allowing them to learn non-linear relationships. This flexibility was useful when trying to learn complex relationships between biological features and clinical outcomes. In this study the activation function of $\phi$ was sigmoidal, linear over a small range of values close to zero, but saturated for large values (equation 1).

$$\phi(v_j(n)) = \frac{1}{1 + \exp(-\alpha v_j(n))}$$  \hspace{1cm} (1)

where $v$ was the sum on node $j$ for case $n$, and $\alpha$ was a constant value of $>0$.

Once the values on the output node had been calculated, they were compared with the desired values and a back propagation algorithm was used to adjust the weights to decrease the difference between the actual and desired predictions. The process was repeated iteratively using all cases in the training set until it met the least mean square error (MSE) between the target and actual output values (equation 2).

$$MSE = \sqrt{\sum_{i=1}^{n}(O - T)^2}$$  \hspace{1cm} (2)

where, $O$ is actual outcome and $T$, predictive outcome, $i$ is the number of patients.

Before ANN model study, the total of 93 patients was proportionally randomized divided into two sub-datasets: training set and verification set. We established two models (A and B). In model A, the ratio of patient numbers was 1:1 for training (47 cases) and verification set (46 cases), and in model B, the ratio was 2:1 (62 cases for training set and 31 cases for verification set). Five clinical variables and 11 dose–volume variables were selected as the predictive factors to train ANN model, which included age, sex, TACE, T stage, portal vein thrombus (PVT), GTV, normal liver volume (NLV), mean dose to normal liver (MDTNL) and $V_5$ to $V_{40}$. $V_5$ was the percentage of normal liver volume that received a dose of $\geq 5$ Gy in the total normal liver volume (Tables 1 and 2). The other $V$ with suffixes expressed the same meaning, but the suffix numbers represented the doses received. The above 16 variables were taken as input

Table 2. Patient dosimetric parameters with or without RILD

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean ± SD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without RILD versus with RILD</td>
<td></td>
</tr>
<tr>
<td>GTV (cm$^3$)</td>
<td>372.9 ± 34.0 versus 723.5 ± 72.1</td>
<td>0.003</td>
</tr>
<tr>
<td>NLV (cm$^3$)</td>
<td>1148.6 ± 289.5 versus 904.3 ± 267.5</td>
<td>0.024</td>
</tr>
<tr>
<td>MDTNL (Gy)</td>
<td>19.9 ± 5.4 versus 24.9 ± 3.9</td>
<td>0.008</td>
</tr>
<tr>
<td>$V_5$ (%)</td>
<td>79.9 ± 13.3 versus 90.4 ± 4.5</td>
<td>0.000</td>
</tr>
<tr>
<td>$V_{10}$ (%)</td>
<td>62.3 ± 14.2 versus 75.4 ± 13.2</td>
<td>0.027</td>
</tr>
<tr>
<td>$V_{15}$ (%)</td>
<td>50.4 ± 14.4 versus 64.4 ± 13.2</td>
<td>0.020</td>
</tr>
<tr>
<td>$V_{20}$ (%)</td>
<td>40.7 ± 13.8 versus 55.5 ± 11.8</td>
<td>0.009</td>
</tr>
<tr>
<td>$V_{25}$ (%)</td>
<td>32.6 ± 13.0 versus 44.6 ± 11.0</td>
<td>0.018</td>
</tr>
<tr>
<td>$V_{30}$ (%)</td>
<td>25.8 ± 12.2 versus 36.1 ± 9.1</td>
<td>0.015</td>
</tr>
<tr>
<td>$V_{35}$ (%)</td>
<td>20.2 ± 11.0 versus 29.9 ± 7.2</td>
<td>0.006</td>
</tr>
<tr>
<td>$V_{40}$ (%)</td>
<td>15.3 ± 9.8 versus 24.8 ± 6.1</td>
<td>0.002</td>
</tr>
</tbody>
</table>

RILD, radiation induced liver disease; SD, standard error; GTV, gross tumor volume; NLV, normal liver volume; MDTNL, mean dose to normal liver; $V_{n}$, the percentage of normal liver volume that received $\geq n$ Gy in the total normal liver volume. The other $V$ with suffixes expressed the same meaning, but the suffix numbers represented the doses received. $P$ value from $t$ test.
variables and RILD, as a single output variable. Hidden number of nodes was set at 4.

To evaluate efficacy of prediction of RILD by ANN, the following were calculated: sensitivity (the fraction of cases with RILD predicted correctly among total cases), specificity (the fraction of cases without RILD predicted correctly among total cases), accuracy (sum of correctly predicted patients divided by total case number), PPV (positive prediction value), and NPV (negative prediction value). Receiver-operating characteristic (ROC) curves were generated for ANN.

RESULTS

During the follow-up period, we found eight cases of RILD among 93 patients (9%) in 4 months after completion of 3DCRT. Three cases were classified as classic RILD, and five cases, non-classic. Elevations of AKP, or sGOT and sGPT appeared in these patients in association with non-malignant ascites in five cases, jaundice in three cases, hepatomegaly in three cases, encephalopathy in two cases and upper quadrant abdominal pain in three cases. Although we gave appropriate therapy to maintain their hepatic function, six RILD cases (75%) finally died of hepatic failures shortly after onset of this complication.

For prediction of RILD, the areas under ROC curves for training and verification set were 0.8897 and 0.8831 from model A and B, respectively (Fig. 2). It showed a good performance regardless of different ratios between training and verification sets (model A or B).

To evaluate the efficacy of prediction of RILD by ANN, we compared the outcome predicted by ANN (model A and B) with what actually occurred. The comparison demonstrated that ANN models, either model A or model B did show good consistency between the predicted RILD and the actual occurring RILD (Tables 3 and 4). For prediction of RILD by ANN, sensitivity, specificity, accuracy, PPV and NPV were 0.875 (7/8), 0.882 (75/85), 0.882 (82/93), 0.412 (7/17) and 0.987 (75/76) for model A, and 0.750 (6/8), 0.800 (68/85), 0.796 (74/93), 0.261 (6/23) and 0.971 (68/70) for model B.

DISCUSSION

For local advanced PLC, radiation therapy has played an important role in the management of PLC, but the problem puzzling us was how to predict the probability of RILD as early as possible. In our previous publications (11), we have found that association of hepatic cirrhosis was the most risky clinical variable for RILD. Therefore, when talking about hepatic irradiation tolerance, we had to separate PLC patients based on their liver cirrhosis, i.e. Child–Pugh grade A, B or C. In our previous study, we have gained hepatic irradiation tolerance in terms of three-dimensional dosimetric parameters. For PLC patients associated with Child–Pugh grade A, the tolerances were MDTNL of 23 Gy, $V_{5}$ of 0.86, $V_{10}$ of 0.68, $V_{15}$ of 0.59, $V_{20}$ of 0.49, $V_{25}$ of 0.35, $V_{30}$ of 0.28, $V_{35}$ of 0.25 and $V_{40}$ of 0.20 (11). These dosimetric parameters would be of great value as dose constraints for liver irradiation.

Besides, mathematical models had also been developed to predict RILD probability for a hepatic irradiation plan. In the literature, the Lyman NTCP model and logistic regression model have been reported (11,12). The aim of the predictive model was to foretell the risk of RILD and evaluate the

<table>
<thead>
<tr>
<th>RILD</th>
<th>No RILD</th>
<th>RILD</th>
<th>No RILD</th>
<th>RILD</th>
<th>No RILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>3</td>
<td>39</td>
<td>4</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Wrong</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>43</td>
<td>4</td>
<td>42</td>
<td>8</td>
</tr>
</tbody>
</table>

RILD, radiation induced liver disease.

Table 4. The result of predictive model B (the ratio of patient number was 2:1 for training and verification set)

<table>
<thead>
<tr>
<th>RILD</th>
<th>No RILD</th>
<th>RILD</th>
<th>No RILD</th>
<th>RILD</th>
<th>No RILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>4</td>
<td>45</td>
<td>2</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Wrong</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>57</td>
<td>3</td>
<td>28</td>
<td>8</td>
</tr>
</tbody>
</table>

RILD, radiation induced liver disease.
safety of a hepatic irradiation plan. In our previous publication, we had applied the Lyman model to predict the probability of RILD (12). Despite the very high accuracy in prediction of RILD by our Lyman model, we still wished to try ANN, which was based on the following considerations. The Lyman NTCP model had shortcomings in prediction. In the Lyman NTCP model, only physical doses were transferred to radiobiological effect, while all clinical features were ignored, though some of them might be critical to the outcome. Whereas, both clinical and dosimetric variables could be put into the model together in a logistic regression model, the linear feature, limited by multicollinearity, restricted the application of model. Multicollinearity meant that different variables containing highly relative information, such as V5–V40, could not be put into the model together. However, as a non-linear model, the ANN model skipped these problems that the Lyman NTCP and logistic regression models faced. Any known information that might associate with predictive outcome could be included in the ANN model regardless of whether those variables correlated with each other. This was to say that compared with the former two models, ANN could make use of information more efficiently. Thus, ANN was more rational, and worth trying.

When an ANN model is applied for prediction of RILD, we met the problem of over-fitting. With the character of wide acceptance, the ANN model had large weight parameters, which were prone to over-fitting (16,17). This meant that as the training set seemed to improve during the course of training, it actually worsened in another dataset (e.g. the verification set). To avoid over-fitting, the technology of ‘cross verification’ was applied as an ‘extra-validation’, which meant that the performance of the verification set was observed concurrent with the process of training, to determine when to stop the process. If the predictive performance of the training set was opposite to that of the verification set, the process of learning was stopped to avoid over-fitting (Fig. 3).

How did we select these 16 variables to build an ANN model? Between patients with RILD and without RILD, all dosimetric variables and T stage were significantly different (P < 0.05), but four other variables, including age, sex, TACE and PVT, were not (Tables 1 and 2). Clearly, we did not have enough reasons to select the latter. Nevertheless, we still chose them because we thought that they might be potential predictive factors.

According to our plan, a continuous extra-validation will be used in the future. Every new PLC patient with Child–Pugh grade A will be evaluated by an ANN model before radiotherapy, then patients will be followed up within at least 4 months to observe the occurrence of RILD as an extra-validation compared to that predicted. If the ANN model is proved to be effective, the data of these patients will be used to re-compute the ANN model. By updating the dynamic model, the accuracy of the model increases with enlargement of the sample size.

In the summary, the ANN model established with our patients could predict the probability of RILD correctly. The ANN model has already been used in a wide range of clinical areas, such as diagnostic radiology, risk assessment and outcome prediction (18–21). Recently, it was applied to predict radiation complication and showed high accuracy for prediction of radiation pneumonitis (22,23). Because of its character of non-liner and huge flexibility, we believe that the ANN model will play a more important role in a wide variety of clinical areas, including radiation oncology.

Acknowledgment

This study was sponsored by Grant No. 2004-826 from Ministry of Public Health, People’s Republic of China.

References


