

Applying Informatics in Tissue Engineering

Jie Xu¹, Xiaolin Zhou², Daping Yang³, Haiyan Ge¹, Qi Wang⁴, Kang Tu⁴, Tiefang Guo³

¹Department of General Surgery, The Shanghai Tenth People's Hospital of Tongji University, Shanghai, P. R. China

²Department of Neurology, The Shanghai First People's Hospital of Shanghai Jiao Tong University, Shanghai, P. R. China

³Department of Orthopaedics, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, P. R. China

⁴Department of Bioinformatics, Harbin Medical University, Harbin, Heilongjiang, P. R. China

Summary

Objective: To facilitate tissue engineering strategies determination with informatics tools.

Methods: Firstly, tissue engineering experimental data were standardized and integrated into a centralized database; secondly, we used data mining tools (e.g. artificial neural networks and decision trees) to predict the outcomes of tissue engineering strategies; thirdly, a strategy design algorithm was developed, and its efficacy was validated with animal experiments; lastly, we constructed an online database and a decision support system for tissue engineering.

Results: The artificial neural networks and the decision trees respectively predicted the outcomes of tissue engineering strategies with the predictive accuracy of 95.14% and 85.26%. Following the strategies generated by computer, we cured 18 of the 20 experimental animals with a significantly lower cost than usual.

Conclusion: Informatics is beneficial for realizing safe, effective and economical tissue engineering.

Keywords

Tissue engineering, informatics, database, artificial intelligence, machine learning

Methods Inf Med 2005; 44: 38–43

1. Introduction

The opportunity that tissue engineering provides for medicine is extraordinary. In the United States alone, over half-a-trillion dollars are spent each year to care for patients who suffer from tissue loss or dysfunction [1]. Currently used alternatives are not intended to integrate into the host, or subjected to wear upon long-term implantation [2]. Tissue engineering is an interdisciplinary field that applies the principles of engineering and life sciences toward the development of biological substitutes that restore, maintain or improve tissue or organ function [3]. Cells, biomaterials and biomolecules constitute tissue engineered products, which regenerate or repair tissue by differentiation, cell-to-cell signaling, biomolecule production, and formation of extracellular matrix. The functionality of an engineered tissue may be structural (e.g., bone, cartilage, and skin) or metabolic (e.g., liver, pancreas), or both. Tissue engineering is promising to become an ideal alternative to tissue or organ transplantation [4]. Time magazine has predicted that tissue engineering will be the number one career by 2010 [5].

Many engineering principles have been applied to make tissue engineering progress rapidly, but the great value that informatics can bring to this field hasn't been widely recognized. The essence of tissue engineering strategy is to choose the best combination of cells, biomaterials, biomolecules and operational processes according to specific clinical occasions. Much importance has been attached to this subject, for it is the key of realizing safe, effective and economical tissue engineering [6, 7]. There

are hundreds of possible combinations of different cells, biomaterials, biomolecules and operational variables, for the four elements are all highly diverse (e.g., cells may be embryonic stem cells, embryonic germ cells, adult stem cells and tens of their differentiated lineages).

Provided with enough experimental data, we would be able to train classifiers and determine strategies successfully by predicting their outcomes. Unfortunately, current tissue engineering data are in irregular forms, scattered in numerous laboratories all over the world. No computer-aided tissue engineering strategy has been introduced, leaving the adoption of strategies at the discretion of diverse experimenters.

In the present study, we constructed a tissue engineering database to integrate experimental data, and applied data mining tools (artificial neural networks and decision trees) to predict the outcomes of tissue engineering strategies. In addition, a tissue engineering strategy design algorithm was developed, and its efficacy was validated with animal experiments.

2. Materials and Methods

2.1 Database Construction and Data Preparation

The tissue engineering database currently comprises six tables, which respectively contain the data of human, quadrumanas, rabbits, sheep, dogs and pigs. Each table has 67 common variables and some species-specified variables (e.g. "addictions" for

human, etc.), describing the features of recipients, tissue engineering strategies and results (Table 1).

The results are evaluated from different points of view, including overall, structural, mechanical, metabolic, biocompatibility, toxicity, etc. The overall result can be described as “cured”, “improved”, “inefficacious” or “deteriorated” as following: If the lost tissue is completely repaired (restored as normal), the result will be defined as “cured”. Otherwise, the result will be defined as “improved”, “inefficacious” or “deteriorated”, comparing with the condition before treatment.

The criteria is appropriate for both human and animal, for “cure” is often easily recognized, and the other three are all relative changes compared with the initial state of the receiver itself, which can be both human and animal. Now, we are developing human-specified evaluations, such as life quality, psychological state, etc.

Current records in the database are mainly elicited from the experiments accomplished in our laboratory (the Tissue Engineering Laboratory of Harbin Medical University), during the years of 1998 to 2004.

In this work, we adopted the “rabbit” table, which had more records and less missing values than other tables. The dataset contained 1,430 records about rabbits treated with engineered bones, cartilage and tendons. There are 67 variables in the dataset, including 7 for “physiological features”, 13 for “pathological features”, 12 for “cells”, 13 for “biomaterials”, 11 for “biomolecules”, 5 for “operational procedures”, 1 for “overall results”, and 5 for “specified results”. The “specified results” were unavailable in some of the records, so the relevant five variables were eliminated from the dataset. We used the “overall result” to denote the outcomes of tissue engineering strategies, and all the other 61 variables were adopted as features of recipients and tissue engineering strategies.

There were three types of values in the dataset: nominal values (e.g. types of cells and biomaterials, etc.), rank values (e.g. nutrition, overall result, etc.), and numerical values (e.g. volume of lost bone, pore size of biomaterial, etc.). Before artificial neural networks and decision trees were trained, a

Table 1 Variables in the database

Objects		Variable examples
Recipients	Physiological features	Gender, age, weight, development, nutrition, etc.
	Pathological features	Extent of tissue loss (volume of bones, area of skin, length of tendons or ligaments), organ function, tumor, infection, wound, immunodeficiency, metabolic disorders, etc.
Tissue engineering strategies	Cells	Embryonic stem cells, embryonic germ cells, adult stem cells or their differentiated lineages, gene-modified cells, etc.
	Biomaterials	Component: polymeric substances, hydroxyapatite, collagen, etc; Structure: porosity, pore size, etc; Shape: membrane, capsules, etc.
	Biomolecules	Transforming Growth Factor (TGF), Bone Morphogenetic Protein (BMP), Epidermal Growth Factor (EGF), Vascular Endothelial Cell Growth Factor (VEGF), Platelet Derived Growth Factor (PDGF), Neutrophil Regeneration factors (NRF), etc; origin, quantity, etc.
	Operational procedures	Cell culture, vascularization, implantation, etc.
Results	Overall results	Cured, improved, inefficacious, deteriorated.
	Specified results	Structural, mechanical, metabolic, biocompatibility, toxicity, etc.

1-of-C encoding (one binary variable per class) was applied to the nominal values. For example, the biomolecules of VEGF, PDGF and BMP were denoted with 001, 010 and 100, respectively. The rank values were denoted with numbers. For example, “nutrition status” was denoted with 0, 1, 2 and 3 instead of “abiotrophy”, “innutrition”, “eutrophy” and “supernutrition”. The variable “overall result” was denoted with 0 and 1, according to the scheme: cured or improved \rightarrow 0 (negative); inefficacious or deteriorated \rightarrow 1 (positive).

Because the input variables have very different orders of magnitude, they were rescaled to be included within the interval [0, 1] by using the following equation:

$$V_n = \frac{V_o - v_{\min}}{V_{\max} - V_{\min}} \quad (1)$$

in which V_o and V_n are respectively the old and new value of the variable, V_{\min} and V_{\max} are the minimum and maximum values of that variable in the original dataset.

In this work, three accuracy measures were used: the true Positive Rate (PR), also known as sensitivity, the true Negative Rate (NR), also known as specificity, and the Predictive Accuracy (PA), which gives an overall evaluation. The accuracy measures can be computed with True Positive (TP), True Negative (TN), False

Positive (FP), and False Negative (FN) as following:

$$PR = \frac{TP}{FN + TP} \times 100(\%)$$

$$NR = \frac{TN}{TN + FP} \times 100(\%)$$

$$PA = \frac{TN + TP}{TN + FP + FN + TP} \times 100(\%) \quad (2)$$

A cross-validation method (K-fold cross-validation) was used to prepare training and test datasets. The 1430 records were randomly cut into five (i.e. $K = 5$) parts, with four of five parts used as training data and the remaining part used as test data each time. For example, for the first set, parts one to four were used for the training data and part five for test data; and for the second set, part one, two, three, and five were used for training data and part four for test data, and so on. The final Predictive Accuracy (PA) was the average of the five test samples. Standard deviations (SD) of the five test samples were also showed. The dataset was randomly cut for 20 times, and each time a cross-validation was implemented. The PA with the smallest SD best indicated the performance of the classifier. In the statistical analysis, a value of $P < 0.05$ was accepted as significant.

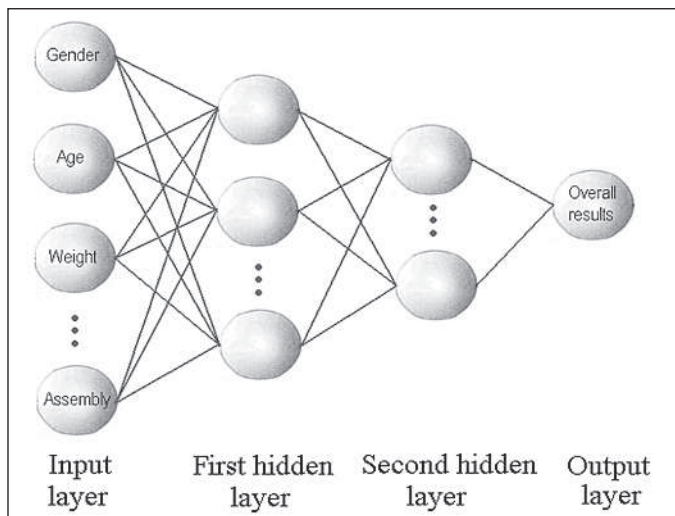


Fig. 1
 Illustration of a four-layered neural network with one input layer, two hidden layers and one output layer

2.2 Data Mining Classification Techniques

The two categories of data mining classification techniques used to predict the outcomes of tissue engineering strategies were carried out with one widely available commercial software program (Matlab), and a shareware tool in the Computational Sta-

tistics Toolbox [8] was also used to infer decision trees.

2.2.1 Artificial Neural Networks (ANNs)

The artificial neural networks (ANNs) with backpropagation algorithm has been utilized in most of the neural networks applications [9-11]. In this study, different neural network architectures were tested to obtain

the best model configuration for the prediction of the tissue engineering strategies. The modeling method was based on the principles of the backpropagation algorithm [12]. A backpropagation network typically comprises three types of neuron layers: an input layer, one or more hidden layers and an output layer each including one or several neurons. As shown in Figure 1, nodes from one layer are connected to all nodes in the following layer, but no lateral connections within any layer, nor feedback connections are possible. We compared the predictive accuracy of ANNs that had different hidden layers, neuron numbers and training epochs.

2.2.2 Decision Trees (DTs)

The most important feature of decision trees (DTs) is their capability to break down a complex decision making process into a collection of simpler decisions, thus extracting knowledge in a form closer to human perception [13-16]. The decision trees inference algorithm provided by the Computational Statistics Toolbox learns from categorized or numerical data with the parameter of “MaxN”, which denotes the maximal number of differently labeled cases permitted in one terminal node. In the cross-validation process, “MaxN” ranged from 1 to 10, resulting in different architectures of decision trees. The predictive accuracy was compared to acquire optimized architectures.

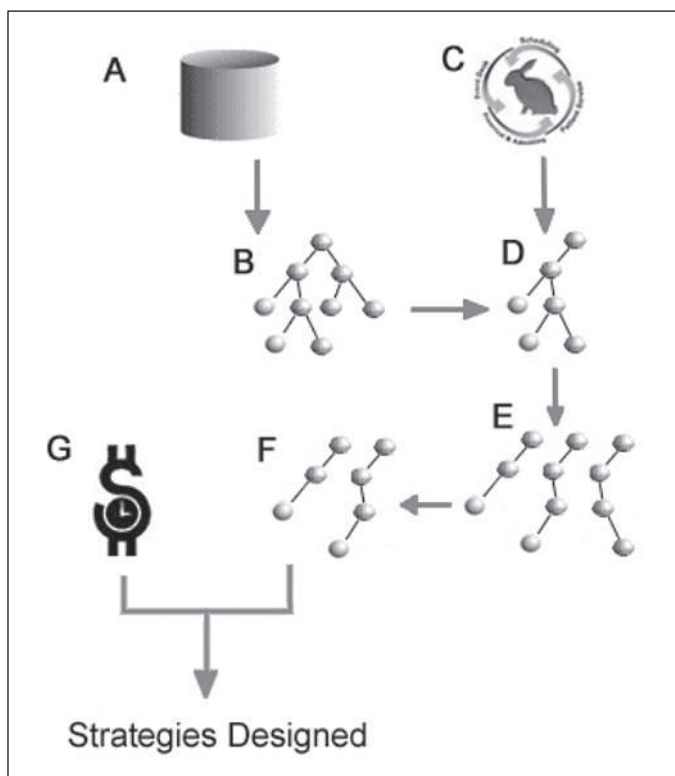


Fig. 2
 The flow chart of tissue engineering strategy design algorithm. (A) Prepare the training data. (B) Infer the decision tree. (C) Specify the condition of the subject. (D) Eliminate those nodes disaccord with the condition of the subject (e.g. If the subject is male and there is a node “gender”, the subsequent nodes after the decision of “female” would be eliminated). (E) Enumerate pathways from the root node to the remained terminal nodes. (F) Select those pathways with terminal nodes as “cured” or “improved”. (G) Estimate the cost of tissue engineering experiments in each node, and select the most cost-effective one as the final strategy.

2.3 Tissue Engineering Strategy Design Algorithm

In a pathway from the root node to one terminal node of the resultant decision tree, we may find both the features of certain clinical occasion and the combination of cells, biomaterials, biomolecules and operational processes. The strategy design algorithm accumulated the successful pathways (i.e. pathways with the terminal node as “cured” or “improved”) and put forward the most appropriate one according to the specified clinical condition. The costs of the strategies were also evaluated to make tissue engineer-

ing more economical (as shown in Fig. 2, the flow chart of the strategy design algorithm).

We applied the algorithm to generate appropriate strategies for bone repair in a femoral critical-sized segmental defect in New Zealand White (NZW) rabbits. The NZW rabbit model for a nonhealing (critical-sized) defect of the femur has been previously established [17]. Based on these studies and our pilot study, we used the following model: NZW rabbits with a unilateral femoral osteotomy gap of 1.2 to 2.0 cm created surgically under general anesthetic. The weight of all rabbits ranged from 2.0 to 3.5 kg before surgery. Two groups of 20 animals were investigated: In the first group, the tissue engineering strategies were generated automatically by the algorithm mentioned above; whereas in the second group, the strategies were provided by a researcher who had five years of experience in tissue engineering. All strategies were implemented by one group of experimenters, who were kept from the origin of the strategies. The curative effect was assessed at three months after surgery.

2.4 Online Database and Decision Support System for Tissue Engineering

We have made the tissue engineering database available at the website: <http://210.46.89.170/tissue/>. The software environment was based on common Microsoft web tools [18-20] and the Matlab Webserver [21]. The Active Server Pages [22] scripts were used to manipulate data and to activate the data mining algorithms, which were executed in the Matlab Webserver.

3. Results

In the cross-validation procedure (which was carried 20 times in this study), the classification model was more precise as the standard deviation becomes smaller. Thus, the performance of specific classifier was denoted by predictive accuracy with the smallest standard deviation. As shown in

Table 2
The predictive accuracy achieved by ANNs and DTs

	Artificial Neural Networks		Decision Trees
	One hidden layer	Two hidden layers	MaxN = 6
PA \pm SD*	88.45 \pm 4.1%	95.14 \pm 3.7%	85.26 \pm 4.9%

*PA = predictive accuracy, SD = standard deviation

Table 3 ANNs with one hidden layer. The predictive accuracy varied with the number of hidden neurons and the maximum training epochs.

Training epochs	Number of hidden neurons							
	3	4	5	6	7	8	9	10
1000	0.6896	0.6981	0.7212	0.7515	0.7231	0.7094	0.7110	0.7177
2000	0.7158	0.7367	0.7552	0.7869	0.7925	0.7771	0.7565	0.7208
3000	0.7138	0.7598	0.7740	0.7663	0.8117	0.8387	0.7898	0.8240
4000	0.7094	0.7708	0.8296	0.7521	0.8162	0.8845	0.8219	0.8144
5000	0.6923	0.7344	0.7962	0.8129	0.8019	0.8447	0.7867	0.7963
6000	0.7046	0.7763	0.8150	0.8319	0.7758	0.8044	0.7735	0.7581
7000	0.7554	0.7725	0.7448	0.8123	0.7817	0.8190	0.7937	0.7983

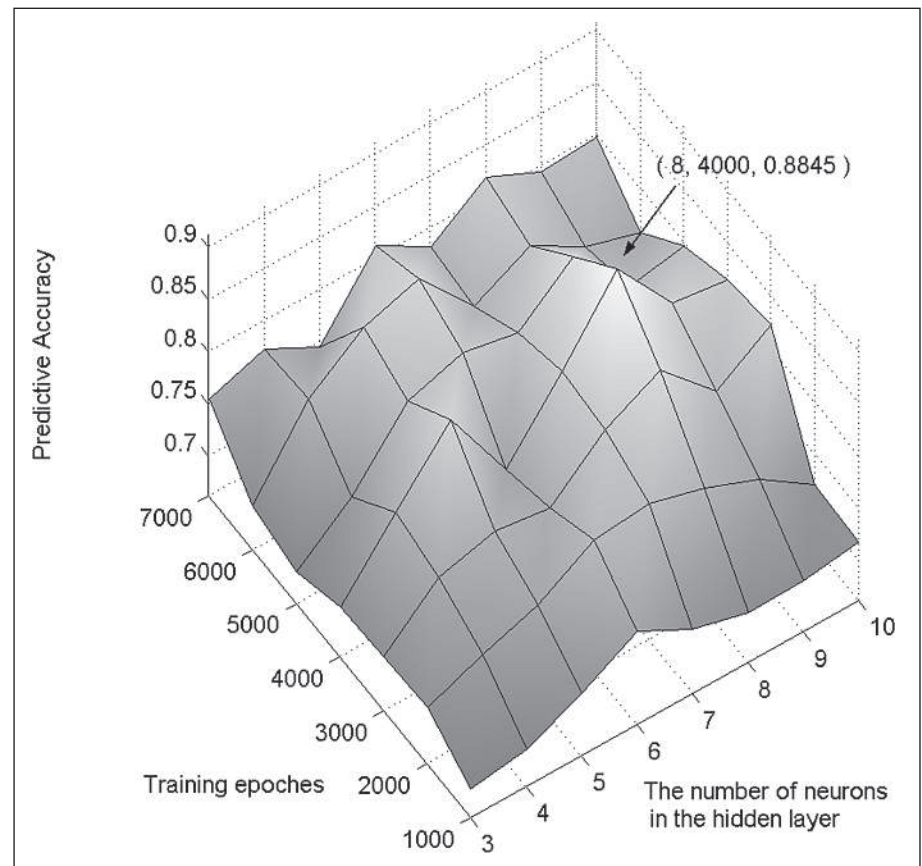


Fig. 3 ANNs with one hidden layer. The prediction accuracy varied with the number of hidden neurons and training epochs. It can also be seen that the neural network with 8 hidden neurons and 4,000 training epochs achieved the highest PA of 88.45% (as labeled by the arrow).

Table 4 ANNs with two hidden layers, the predictive accuracy varied with the numbers of neurons in the first and second hidden layers.

N_2^*	Number of neurons in the first hidden layer						
	2	3	4	5	6	7	8
2	0.7979	0.8071	0.8158	0.825	0.8938	0.8854	0.8438
3	0.7937	0.8113	0.8313	0.8467	0.9514	0.8974	0.8854
4	0.8229	0.7925	0.8442	0.8938	0.9275	0.9146	0.8975
5	0.8354	0.8043	0.9067	0.9296	0.9354	0.8942	0.8683
6	0.8171	0.8267	0.8896	0.8592	0.8871	0.8792	0.8762
7	0.8175	0.8329	0.8433	0.8478	0.8292	0.8479	0.8536
8	0.7925	0.7833	0.8208	0.8395	0.8571	0.8333	0.8454

* N_2 : the number of neurons in the second hidden layer.

Table 2, the artificial neural networks with two hidden layers performed best and achieved the predictive accuracy of 95.14%, which was significantly higher than that of decision trees and artificial neural networks with one hidden layer (t-test, $P < 0.05$).

3.1 Optimization of the Artificial Neural Networks with One Hidden Layer

The predictive accuracy varied with the number of hidden neurons and the maximum training epochs (as shown in Table 3 and Fig. 3). After experimental trails, the

number of hidden neurons was set to 8, being the maximum number of epochs set to 4000. The resultant predictive accuracy was 88.45% (the highest in all groups), and the relevant sensitivity (PR) and specificity (NR) were 89.27% and 87.61%, respectively.

3.2 Optimization of the Artificial Neural Networks with Two Hidden Layers

Neural networks with different training epochs and different numbers of neurons in the first and second hidden layers were respectively validated. The numbers of neurons in the two hidden layers all ranged from 2 to 8, resulting in 49 network architectures. The most appropriate training epoch for specific architecture was searched in the range of 2,000 to 20,000. After experimental trails, a neural network with respectively six and three neurons in the first and second hidden layers achieved the highest predictive accuracy of 95.14% (as shown in Table 4 and Fig. 4). The training epoch was set to 13,000, and the relevant sensitivity (PR) and specificity (NR) were 95.32% and 94.71%, respectively.

3.3 Decision Trees

The cross-validation procedure was implemented to the decision tree classifier. As shown in Table 5, the predictive accuracy depended on the value of MaxN. When MaxN was set to 6, the relevant decision tree performed best and achieved a predictive accuracy of 85.26%. The relevant sensitivity and specificity were 84.63% and 86.12%, respectively.

3.4 Implementation of the Strategy Design Algorithm

All rabbits were examined at the same time point (12 weeks after surgery). As shown in Table 6, more rabbits in the first group (computer-aided strategies group) were cured or improved, and the therapeutic cost

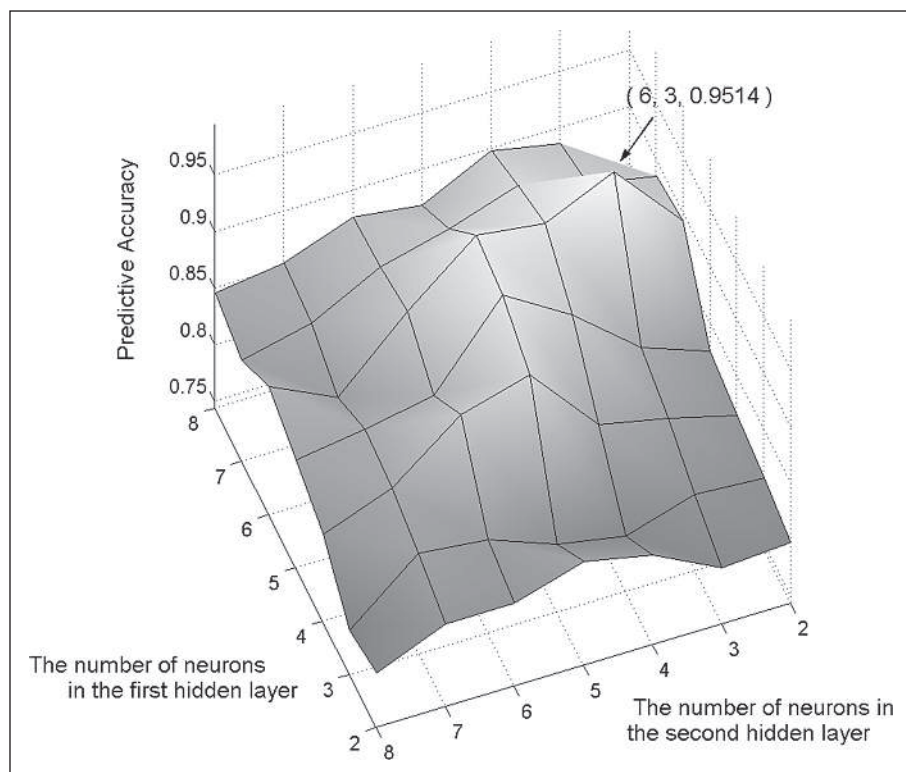


Fig. 4 ANNs with two hidden layers, the prediction accuracy varied with the numbers of neurons in the two hidden layers. The ANN with respectively 6 and 3 neurons in the first and the second hidden layers achieved the highest PA of 0.9514

Table 5 The predictive accuracy and the relevant MaxN

MaxN	2	3	4	5	6	7	8	9	10
PA	0.6032	0.7243	0.7558	0.8142	0.8526	0.8393	0.8053	0.7995	0.7837

Table 6

Results of tissue engineering strategies

	The first group	The second group
Origins of strategies	Generated by computer	Base on human experience
Cured or improved cases	18	17
Inefficacious or deteriorated cases	2	3
Cost per rabbit (dollars)	64.0 ± 5.1	73.8 ± 4.9

per rabbit in the first group was significantly lower than that in the second group (t-test, $P < 0.05$).

4. Discussion and Conclusions

Despite the rapid progress of tissue engineering in the 1990s, concerns on safety, effect reliability and therapeutic costs have been obstructing the extensive application of tissue engineering. In order to overcome these difficulties, one potential way is to analyze existing experimental data in an attempt to predict the outcomes, and to suggest therapies [23]. In this work, informatics tools were applied to rationalize tissue engineering strategies.

Firstly, the Internet and database were found useful for the acquisition, standardization and delivery of tissue engineering information. Currently, tissue engineering experiences are shared mainly through published literature, which usually describes only successful cases in very short and ragged forms. Using the Internet and a centralized database, we can establish an open data source, which has standardized format, high capacity and wide accessibility. Thus, the informational flow in this field can be expedited.

Furthermore, data mining tools have been applied to determine tissue engineering strategies. In the computational experiments based on a set of animal data, the artificial neural networks could detect un-

successful strategies with the predictive accuracy of 95.14%, so they can be used to give an alarm to improper strategies. The higher accuracy of artificial neural networks may be due to its capability of generalization of nonlinearly separable problems, while the decision trees have the advantage of extracting knowledge in a form closer to human perception. Even if the experimenter has no prefigured strategy, the design algorithm would serve as a consultant, providing both effective and economical solutions.

It is important to stress the main goal of this work: to show that it is possible to facilitate tissue engineering strategic determination with informatics tools. The results obtained so far (a predictive accuracy of 95.14%), although not for all species, clearly back this claim. In addition, the proposed approach opens room for the automation of tissue engineering strategies.

In future research, we plan to improve the database by adjusting the variables and acquiring more experimental data. Another interesting direction is based on the application of further data mining tools to more species (especially human beings). Finally, it would be meaningful to improve the online data submission and decision support system.

References

- Lanza R, Vacanti JP, Langer R, editors. Principles of Tissue Engineering Second Edition. Academic Press; 1997.
- Maguire JK, Cosca MF, Lynch MH, editors. Clin Orthop 1987; 216: 213-23.

- Langer RS, Vacanti JP. Tissue engineering: the challenges ahead. *Sci Am* 1999; 280 (4): 86-9.
- Chapekar MS. Tissue engineering: challenges and opportunities. *J Biomed Mater Res* 2000; 53 (6): 617-20.
- Rawe J. "What will be the ten hottest jobs?" *Time*, 2000; 155: 54.
- Yang ZM, Spector M, editors. Tissue Engineering. China Chemical Industry Press; 2002.
- Mason C. Automated tissue engineering: a major paradigm shift in health care. *Med Device Technol* 2003; 14 (1): 16-8.
- Martinez WL, Martinez AR, editors. Computational Statistics Handbook with MATLAB. Chapman & Hall/CRC Press; 2001.
- Robert J. Schalkoff, editors. Artificial Neural Networks, Addison-Wesley Press; 1997.
- Bose NK, Liang P., editors. Neural Network Fundamentals with Graphs, Algorithms, and Applications. McGraw-Hill Press; 1996.
- Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995; 346: 1135-8
- Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press; 1986: 45-7.
- Swain P, Hauska H. The decision tree classifier design and potential. *IEEE Trans Geosci Electron* 1977; 15: 142-7.
- Quinlan JR. Induction of decision trees. *Machine Learning* 1986; 1: 81-106.
- Breiman L, Friedman JH, Olshen RA, editors. Classification and Regression Trees. Chapman & Hall; 1984.
- McKenzie DP, McGorry PD, Wallace CS, et al. Constructing a minimal diagnostic decision tree. *Methods Inf Med* 1993; 32: 161-6.
- Concannon MJ, Boschert MT, Puckett CL. Bone induction using demineralized bone in the rabbit femur: a long term study. *Plast Reconstr Surg* 1997; 99: 1983-8.
- Haravu LJ. Web-enabled databases: an overview of technology and problems. *Information Today and Tomorrow* 2000, 19 (1): 7-10.
- William H, Mary C. Web-based access to locally developed databases. *Library Computing* 1999; 18 (1): 51-8.
- Homer A., Sussman D. B., editors. Active Server Pages 3.0, Wrox Press; 1999.
- <http://www.mathworks.com/products/webserver/>
- <http://msdn.microsoft.com/asp/>
- McQuatt A, Sleeman D, Andrews PJ, et al. Discussing anomalous situations using decision trees: a head injury case study. *Methods Inf Med* 2001; 40 (5): 373-9.

Correspondence to:

Jie Xu
 NO.119, Bei Xin Street
 Dao Wai District
 Harbin
 Hei Long Jiang
 China 150036
 E-mail: wwwdbwww@163.com