

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Diagnostic Prediction of Vertebral Column Using Rough Set Theory and Neural Network Technique

Mei-Ling Huang, Yung-Hsiang Hung and Den-Ming Liu

Department of Industrial Engineering and Management, National Chin-Yi University of Technology,  
No. 57, Sec. 2, Zhongshan Rd., Taiping, Taichung, 41170, Taiwan, Republic of China

---

**Abstract:** Numerous physical diseases come with age, many of which are inherited and can be alleviated through appropriate therapies if they are recognized in the early stage. The vertebral column supports the head and trunk; it enables us to stand erect, protects the nerves governing physical activities and sensations and is the main transfer passage of nerves. This study employed the artificial neural network and rough set theory to conduct experimental tests on the medical diagnosis of the vertebral column from University of California at Irvine (UCI) database. To avoid long operation time and large weight differences, the data were simplified by standardization and classification accuracy and area under receiver operating characteristic curve were used to assess diagnostic performance. This study used back-propagation neural network (BPNN), rough set theory GA Algorithm (RST-GA) and rough set theory Johnson Algorithm (RST-JA) to test a medical database of vertebral columns. The experimental results showed that the BPNN is superior to RST-GA and RST-JA and has higher accuracy on medical diagnosis of vertebral column. The classification accuracy and under the ROC curve (AURC) of BPNN were 90.32 and 99.42%, respectively. This study developed intelligent diagnostic methods to assist doctors on assessing vertebral column patients, thus, providing patients with more suitable and earlier medical care. With high diagnostic accuracy, the experimental results showed that BPNN classification prediction model was reliable on assessing the disease of vertebral column to facilitate a better lifestyle in a timely fashion for patients. However, without practical verification, this prediction model may not be applicable to other diseases.

**Key words:** Artificial neural network, rough set theory, genetic algorithm, medical diagnosis

---

### INTRODUCTION

Various physical diseases come with age, of which many are inherited and inevitable. The vertebral column plays an important role in body motion, which supports the head and trunk and protects the nerves governing body activities and sensations. In addition, it is the main transfer passage of the nerves (Xie *et al.*, 2012).

The vertebral column comprises vertebra and inter-vertebral disc, which main function is protecting the spinal cord and supporting weight. In terms of biomechanics and anatomy, the vertebral column is composed of vertebral body, inter-vertebral disc, facet joints and ligaments, all of which have their own mechanical roles. The vertebral body and inter-vertebral disc sustains pressure, the facet joint, the shear force, axial torque force and the ligament the tension (Allaoui and Artiba, 2009).

There are many data mining technologies, in which Artificial Neural Network (ANN), Genetic Algorithm (GA)

and Support Vector Machine (SVM) are the most widely used at present (Cheng and Jhan, 2013). This study employed the back-propagation neural network (BPNN) and rough set theory (RST) to construct an intelligent prediction method suitable for medical diagnosis of the vertebral column. The proposed method is expected to assist doctors and examiners on diagnosing diseases and provide with faster and better medical care and other services to patients.

Lumber-back pains have many symptoms, which may be caused by muscular injuries or non-traumatic diseases. Lumber-back pain and diseases afflicting other parts of the body in turn, strikes the muscles or other lower lumbar structures. The pain may be caused by the neural system, or the tension in it.

Scoliosis strikes 2% females and 0.5% males out of the total population. The disease has many causes, such as congenitally, hereditarily, neural muscle and unequal limb length, as well as other causes, including rachischisis, spinal muscular atrophy and tumor

(Kramer, 1981). Over 80% of scoliosis cases are idiopathic, meaning that their causes are unknown. In terms of age, idiopathic scoliosis can be divided into four kinds, which are under 3 years old, 3-9 years old, 10-18 years and after skeletal maturation. The most common kind is 10-18 years old, accounting for 80% of all cases.

The human central nervous system (CNS) comprises the brain and spinal cord. The brain is protected by solid skull, while the long spinal cord is protected by the vertebra, which is the keel. Between the vertebrae there is cartilage serving as a buffer, called the inter-vertebral disc. If the inter-vertebral discs are deformed, displaced, or cracked, they may press the adjacent spinal cord and spinal nerves, which is called Herniated Inter-vertebral Disc (HIVD) (Tay and Shen, 2003).

Spondylolysis means that a vertebra shifts forward relative to its adjacent vertebrae, which can cause deformation in the lumbar vertebrae and spinal stenosis, with the most common symptom of lumbar pain. Vertebral degradation may also cause spondylolysis as the backbone becomes aged and worn with time and such changes affect normal and healthy vertebral arrangement. The facet joints are connections at the back of the backbone. Degraded inter-vertebral discs and facet joints can cause greater displacement in the vertebral body, resulting in loose facet joints and less support functions. Patients suffering from spondylolysis caused by degradation, which mainly produces dislocation in the fourth and fifth lumbar vertebrae, are generally over 40 years old female (Kramer, 1981).

**MATERIALS AND METHODS**

**Artificial neural network (ANN):** ANN is a heuristic algorithm for simulating the composition of nerve cells, which is comprised of nodes in three layers, which are Input Layer, Hidden Layer and Output Layers, with the network and node architecture as shown in Fig. 1. The product of the Hidden Layer node, where they are joined, is calculated by weight and threshold values, which are sent back to the network, in order to approach the expected value, raise classification accuracy and achieve better effects, as compared with traditional statistics (Li *et al.*, 2012).

**Rough set theory (RST):** The rough set theory (RST), first proposed by Polish mathematician, Pawlak (1982), is a mathematical theory used to process inaccurate, uncertain and incomplete data. RST defines knowledge as a division of data and each divided set is called a concept. A pair of approximation sets (upper and lower approximation sets) are selected from the known

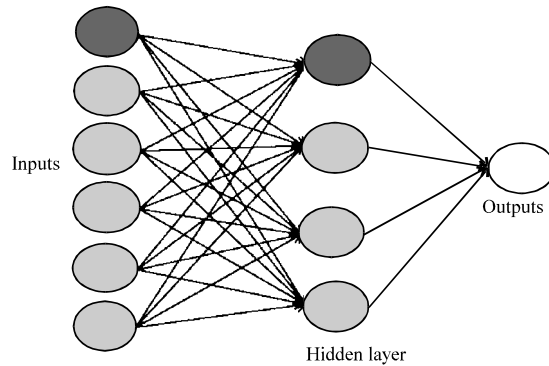


Fig. 1: Example of a neural network structure

divisions to provide an approximate description of a concept, which becomes the rule and can be applied to the classification and prediction of new data. As the application of RST to analysis not subjected to assumption, its flexibility is increased. The earliest use of the theory was to solve inaccuracy problems by determining the relationships between data attributes. Since then, it has been widely applied in many fields (Chen, 2001).

This study employed the UCI Machine Learning Repository Vertebral Column dataset, where the characteristic variables include 6 input variables: (1) Pelvic incidence, (2) Pelvic tilt, (3) Lumbar lordosis angle, (4) Sacral slope, (5) Pelvic radius and (6) Degree spondylolisthesis and 1 type of variable (1: Abnormal, 0: Normal), with a total of 310 instances in test data sets. In this study, BPNN, RST GA Algorithm (RST-GA) and RST Johnson Algorithm (RST-JA) were used for experimental testing and the data sets were standardized before constructing the model. Figure 2 schematically represents the experiment process steps evolving in such three ways.

**Standardization of data sets:** To avoid long operation time and large weight differences, the data must be simplified, meaning the data are narrowed for input values of the nerve network. The standardization equation is given below:

$$x^* = \frac{x_i - \bar{x}}{s^2} \tag{1}$$

where,  $x^*$  is the narrowed data,  $x_i$  is the  $i$ th original data,  $\bar{x}$  is the average value of original data and  $s^2$  is the sample variance of the original data. This study collected a total of 310 data instances and used 6 characteristic variables (i.e., pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and degree spondylolisthesis)

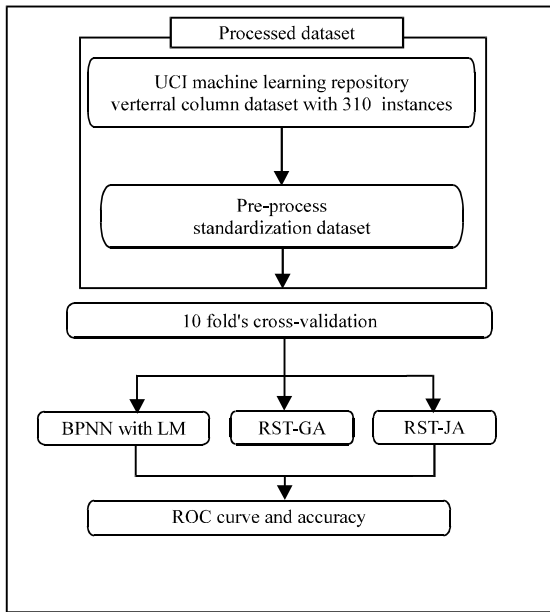


Fig. 2: Experimental paradigm for vertebral column experiments

to determine whether inter-vertebral discs suffer herniation or spondylolisthesis, for division into abnormal and normal status.

**Back-propagation neural network (BPNN):** BPNN is used for monitored learning, which is most frequently used in the artificial nerve network, as it has visible effects on solving classification problems (Romero *et al.*, 2011). BPNN parameters include structure parameters and learning parameters, where structure parameters include the hidden layer; and learning parameters include the learning rate (LR), initial weight range and inertia items. In this study, the non-linear transfer function used in the BPNN hidden layer is the Tan-Sigmoid Transfer function (tansig). The tansig function produces output ranging between -1 and 1 (Fig. 3) and responds to the input of neurons varying from negative infinity to positive infinity (Trappey *et al.*, 2006). The linear pureline transfer function was employed in the output layer (Fig. 4).

**Genetic algorithms (GA):** GA is an optimization algorithm developed from the “Survival of the Fittest in Natural Selection”, as described in the “Theory of Evolution” attributed to Darwin. In this algorithm, the initial population is randomly selected and the fitness function is evaluated in order to set the conditions for stopping operations, such as the number of generations to reach the best result (Chen *et al.*, 2010). To enable rapid

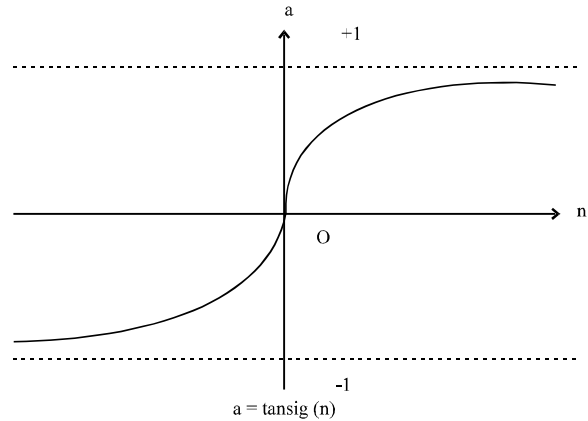


Fig. 3: Illustration of the tan-sigmoid transfer function

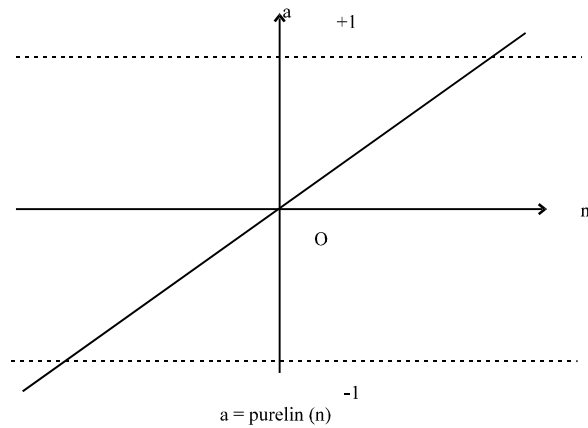


Fig. 4: Illustration of the pureline transfer function

calculation by a computer, Binary Genetic Algorithms (BGA) are used, thus, the encoding is required. The operational model is given as below:

- Produce offspring by random crossover of parent generation
- Stimulate mutation in offspring
- Use the roulette wheel method to select better genes

Crossover refers to double-point crossovers or uniform crossovers between two segments of chromosomes. The advantage of mutation is to produce offspring and avoid the restriction of a local value. The probability of crossover and mutation can be set between 0 and 1 (the latter always falls between 0.01 and 0.08). In the roulette wheel method, the best gene retained is sent to the crossover pool and GA convergence can be quickened reached to produce a new population of

offspring. According to the fitness function, poorer genes are removed and better genes are retained. Afterwards, the binary system is decoded back into the decimal system for a new population. The main advantage of GA is to avoid the restriction of a local best solution (Wong *et al.*, 2009). As the mutation capability enables GA to locate the best point within the entire region, this study used GA to reduce classification rules, thus, enhancing classification accuracy and increasing the area under receiver operating characteristic curve.

**Johnson algorithm (JA):** Johnson algorithm (JA), also called Johnson's Rule, is usually applied to solve two-machine operation problems. It is able to optimize the sequence and check the sequence for ordering based on the ordered quantity of items (Sheu *et al.*, 2012). The four steps are as follows:

- **Step 1:** Let  $U = \{j|t_{j1} < t_{j2}\}$  and  $V = \{j|t_{j1} \geq t_{j2}\}$
- **Step 2:** Arrange the sequence from small to large in terms of  $t_{j1}$
- **Step 3:** Arrange the sequence from large to small in terms of  $t_{j2}$
- **Step 4:** The best sequence is from ordered set U to ordered set V

This study used JA to reduce RST classification rules and compared with RST+GA combinations to determine the better classification accuracy or area under receiver operating characteristic curve.

This study used JA to reduce RST classification rules and compared with RST+GA combinations to determine the better classification accuracy or area under receiver operating characteristic curve.

**Performance evaluation:** ROC (Receiver Operating Characteristic Curve) was published by Hanley and McNeil (1982). A greater area under the ROC curve (AURC) leads to a better effect. The ROC curve is drawn with a true positive rate (TP) and a false positive rate (FP) and the AURC value is determined.

## RESULTS AND DISCUSSION

This study divided the 310 data items into two groups, which are the training group (248 items), accounting for 80% and test group (62 items), accounting for 20%, for data preprocessing of BPNN and RST. The database was used to compare the accuracy and AURC values of the three methods, in order to find their classification effects regarding the database.

**Experimental results of BPNN:** In the training setting of BPNN, 5 learning rates (0.01, 0.05, 0.1, 0.5, 0.9) and 4 Hidden Layer numbers (3, 4, 5, 6), with 20 experimental combinations planned are used for comparative analysis. The most suitable prediction module is chosen according to the AURC experimental results. In the following BPNN performance, results are given in detail when the neuron numbers are  $k = 3$ ,  $k = 4$ ,  $k = 5$  and  $k = 6$ .

**Experimental combination 1:** When the neuron number in Hidden Layer  $k = 3$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the accuracy and AURC experimental results of BPNN are as shown in Fig. 5a-e. The experimental results in Fig. 6 show that BPNN has the best accuracy, but a greater variance of  $87 \pm 4\%$  when  $LR = 0.5$ . When  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the average classification accuracy of BPNN is  $87 \pm 4\%$ ; the highest AURC value is 0.98695 and the lowest value is 0.9645 when  $LR = 0.9$ .

**Experimental combination 2:** When the neuron number in Hidden Layer  $k = 4$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the accuracy and AURC experimental results of BPNN are as shown in Fig. 7a-e. The experimental results in Fig. 8 show that BPNN has the best accuracy when  $LR = 0.1$ . When  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the average classification accuracy of BPNN is 89%; the highest AURC value is 0.99419 when  $LR = 0.9$  and its lowest value is 0.97024 when  $LR = 0.5$ .

**Experimental combination 3:** When the neuron number in Hidden Layer  $k = 5$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the accuracy and AURC experimental results of BPNN are as shown in Fig. 9a-e. The experimental results in Fig. 10 show that BPNN has the best accuracy when  $LR = 0.05$ . When  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the average classification accuracy of BPNN is 87%; the highest AURC value is 0.9922 when  $LR = 0.9$  and its lowest value is 0.96465 when  $LR = 0.05$ .

**Experimental combination 4:** When the neuron number in hidden layer  $k = 6$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the accuracy and AURC experimental results of BPNN are as shown in Fig. 11a-e. The experimental results in Fig. 12 show that the accuracy variance is greatest, the full range from the maximum to minimum value is 4% and BPNN has best accuracy when  $LR = 0.05$ . When  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , the average classification accuracy of BPNN is 87%; the highest AURC value is 0.97872 when  $LR = 0.05$  and its lowest value is 0.94048 when  $LR = 0.9$ .

The results of 20 experimental tests, with  $k = \{3, 4, 5, 6\}$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$ , show that the lowest and highest AURC value of BPNN are 0.94186 and

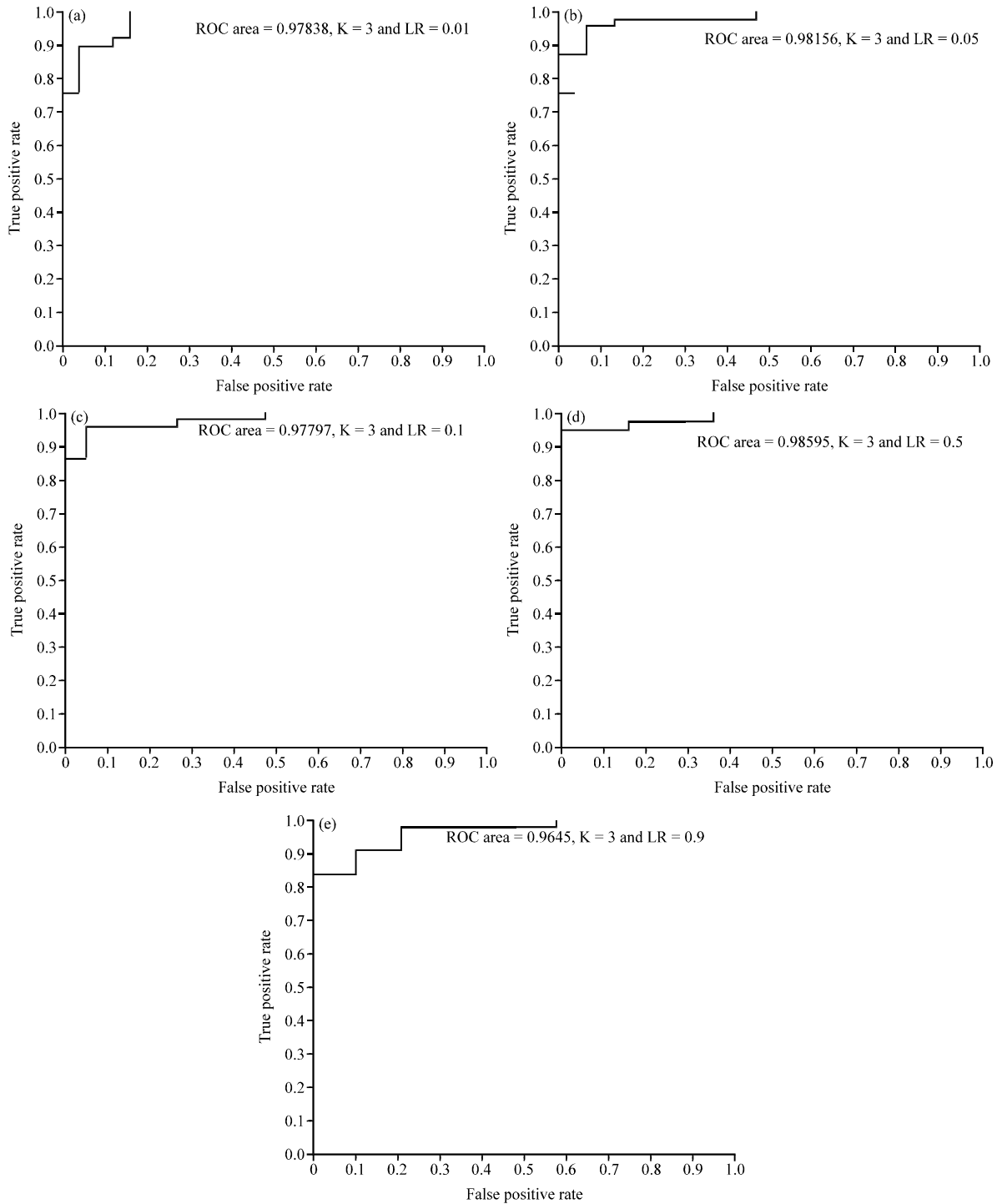


Fig. 5(a-e): (a) AURC result of BPNN:  $k = 3, LR = 0.01$ , (b) AURC result of BPNN:  $k = 3, LR = 0.05$ , (c) AURC result of BPNN:  $k = 3, LR = 0.1$ , (d) AURC result of BPNN:  $k = 3, LR = 0.5$  and (e): AURC result of BPNN:  $k = 3, LR = 0.9$

0.99369, respectively, thus, BPNN has excellent classification accuracy performance. The experimental

results also show that a greater or smaller neuron number does not lead to better accuracy or AURC value and the

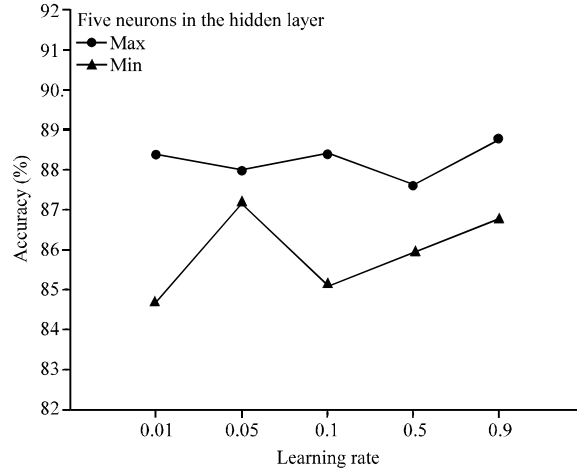


Fig. 6: Comparison of the AURC results for  $k = 3$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$

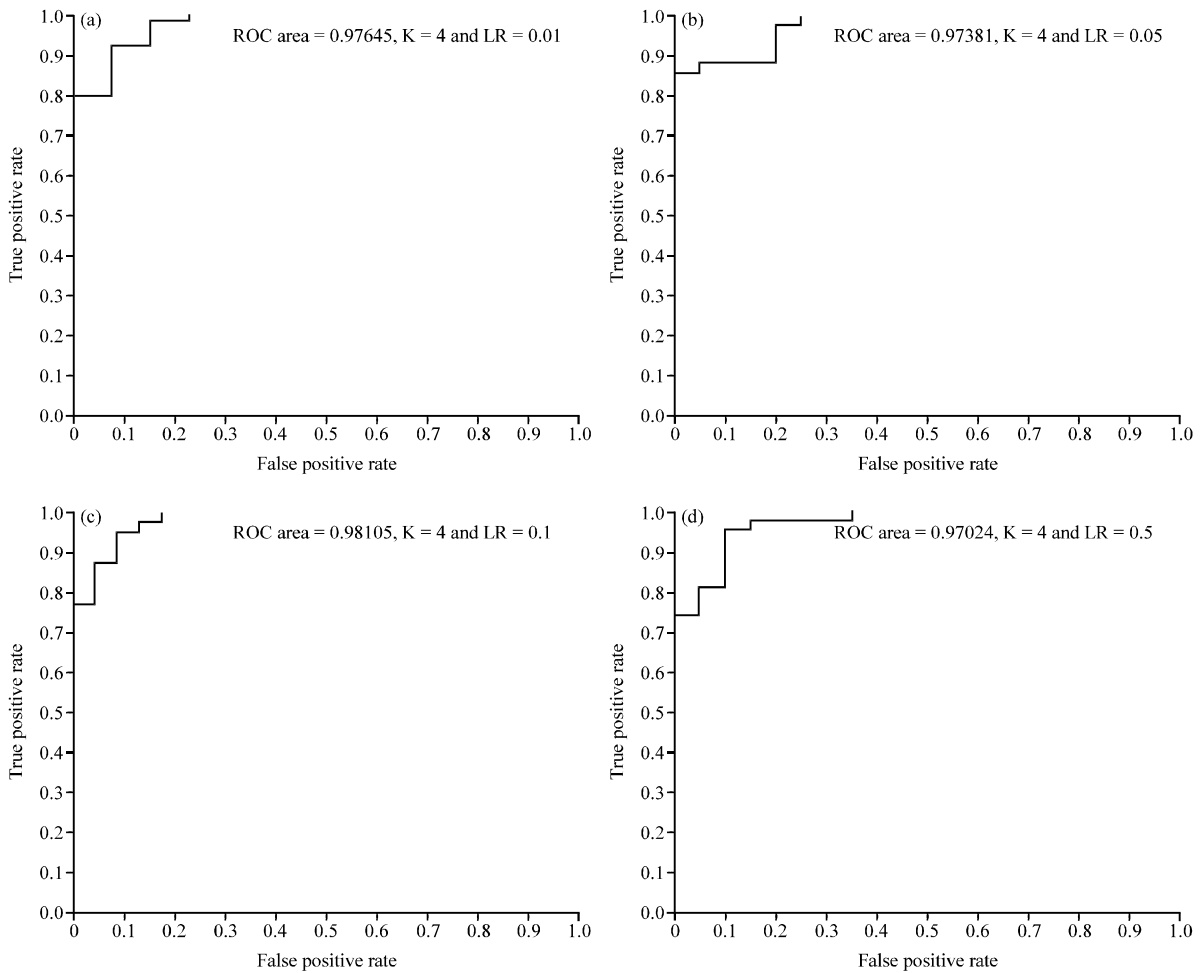


Fig. 7(a-e): Continue

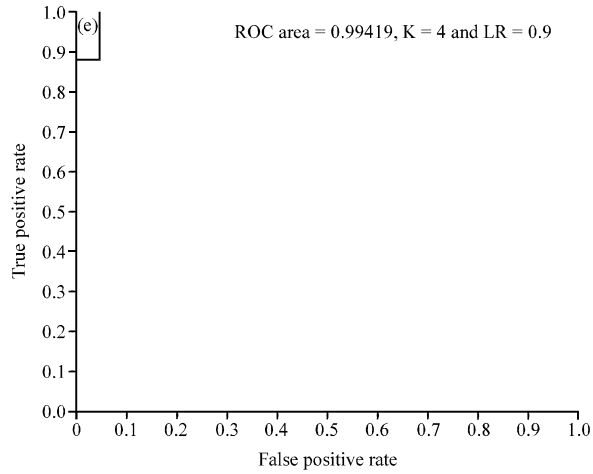


Fig. 7(a-e): (a) AURC result of BPNN: k = 4, LR = 0.01, (b) AURC result of BPNN: k = 4, LR = 0.05, (c) AURC result of BPNN: k = 4, LR = 0.1, (d) AURC result of BPNN: k = 4, LR = 0.5 and (e) AURC result of BPNN: k = 4, LR = 0.9

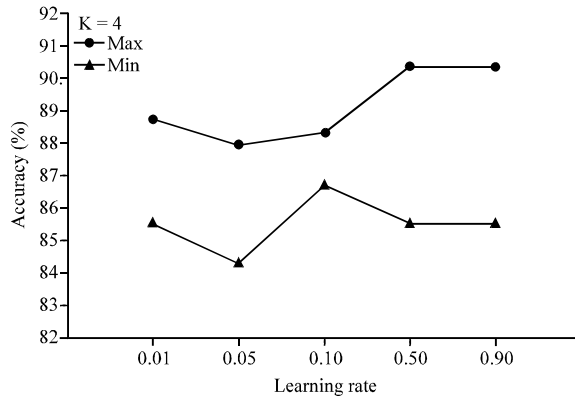


Fig. 8: Comparison of the AURC results for k = 4 and LR = {0.01, 0.05, 0.1, 0.5, 0.9}

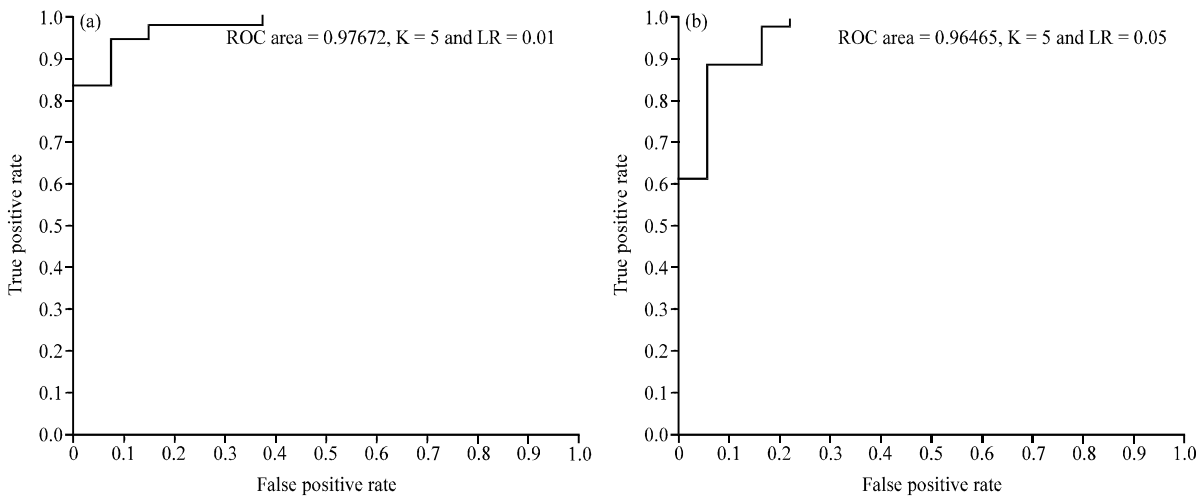


Fig. 9(a-e): Continue



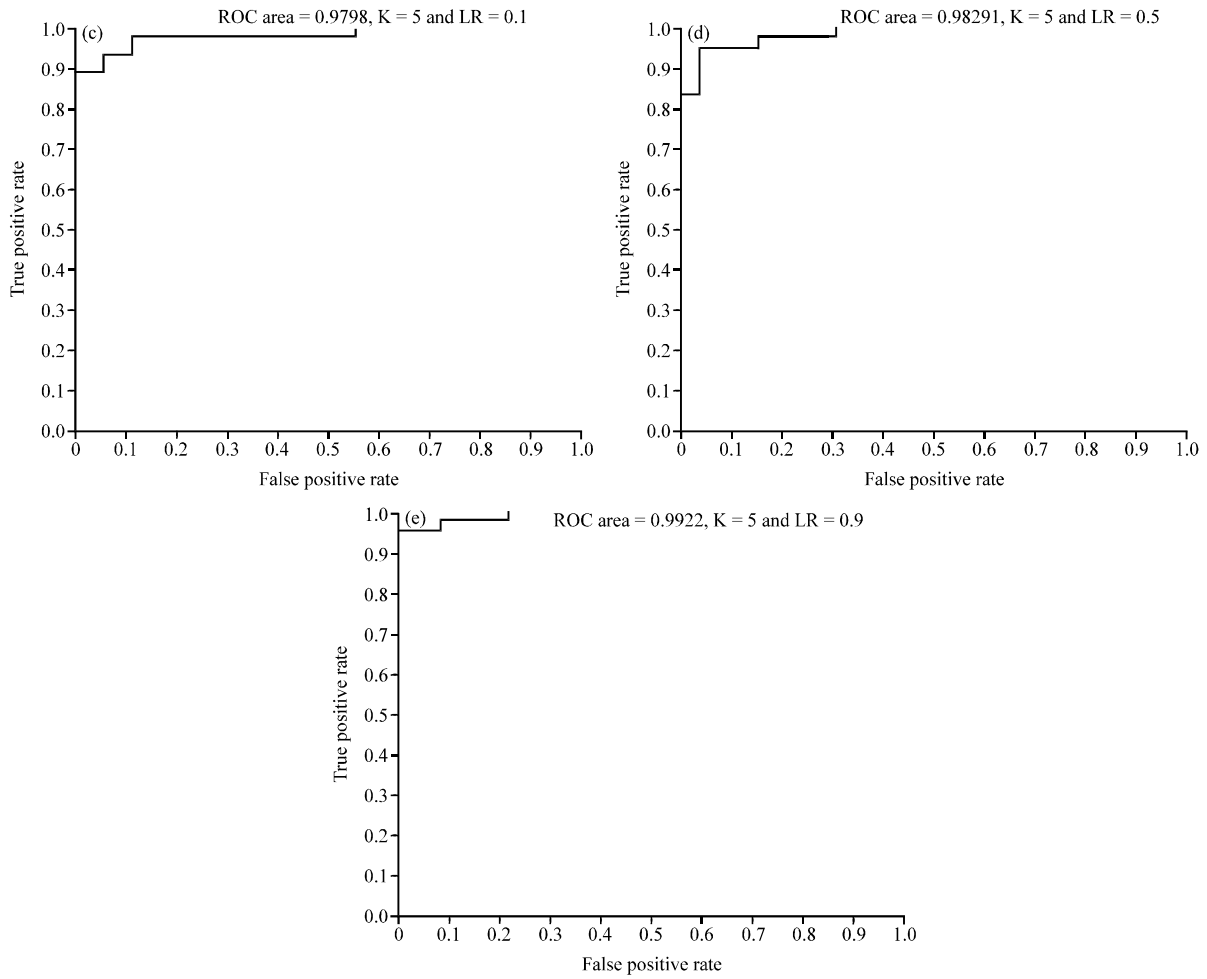


Fig. 9(a-e): (a) AURC result of BPNN:  $k = 5$ ,  $LR = 0.01$ , (b) AURC result of BPNN:  $k = 5$ ,  $LR = 0.05$ , (c) AURC result of BPNN:  $k = 5$ ,  $LR = 0.1$ , (d) AURC result of BPNN:  $k = 5$ ,  $LR = 0.5$  and (e) AURC result of BPNN:  $k = 5$ ,  $LR = 0.9$

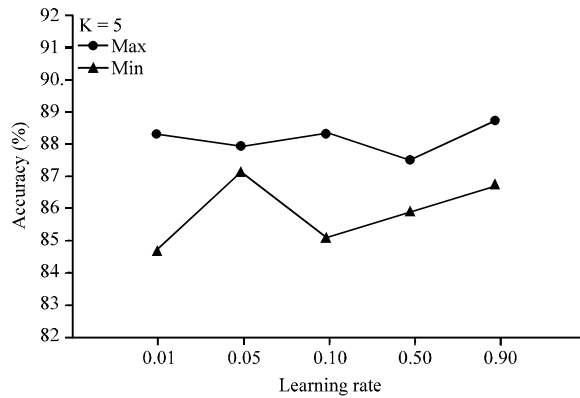


Fig. 10: Comparison of the AURC results for  $k = 5$  and  $LR = \{0.01, 0.05, 0.1, 0.5, 0.9\}$

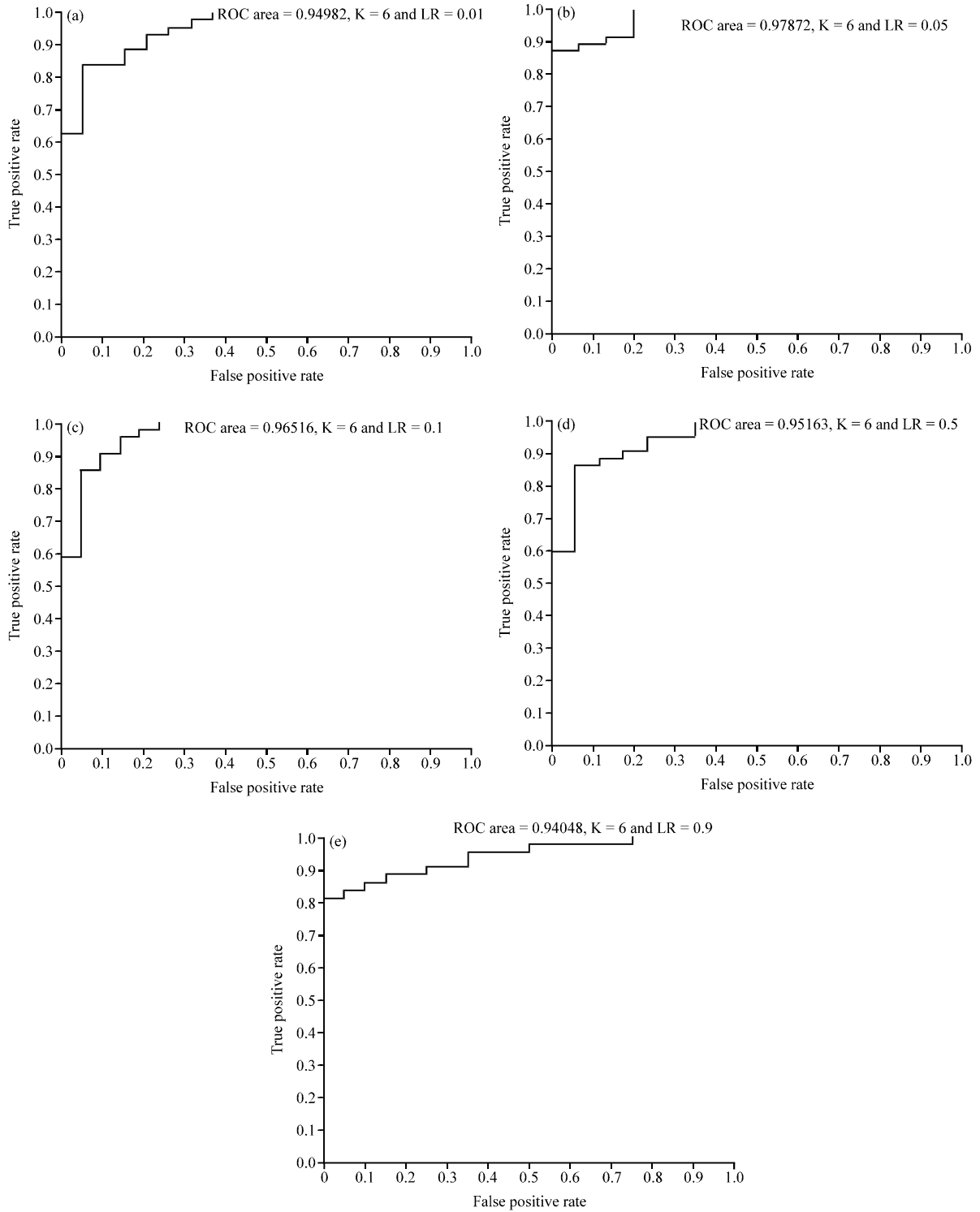


Fig. 11(a-e): (a) AURC result of BPNN:  $k = 6$ ,  $LR = 0.01$ , (b) AURC result of BPNN:  $k = 6$ ,  $LR = 0.05$ , (c) AURC result of BPNN:  $k = 6$ ,  $LR = 0.1$ , (d) AURC result of BPNN:  $k = 6$ ,  $LR = 0.5$  and (e) AURC result of BPNN:  $k = 6$ ,  $LR = 0.9$

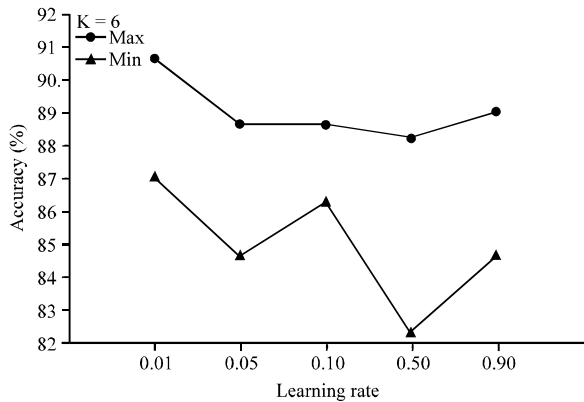


Fig. 12: Comparison of the AURC results for k = 6 and LR = {0.01, 0.05, 0.1, 0.5, 0.9}

Table 1: Accuracy of classification for RST-GA

No.	1	2	3	4	5	6
Boundary region	Default	Default	Default	0.95	0.80	0.70
Hitting fraction	0.45	0.35	0.50	0.45	0.45	0.45
Accuracy	80.65	80.65	80.65	79.03	82.26	75.81
AURC	75.21	74.60	75.21	78.52	88.74	87.27

set value of a characteristic value usually adopted by scholars, plus type number and divided by two, is the ideal neuron number. This study also found that the hidden neuron number of  $(6+2)/2 = 4$  result in more stable and accurate AURC values.

**Experimental results of RST-GA and RST-JA:** In the RST method, this study chose GA and JA to reduce the rules. In the parameter settings of GA, the crossover rate is 0.3 and the mutation rate is 0.05, the boundary region (BR) and the Hitting Fraction (HF) are adjusted for experimental tests. BR varies between 0 and 1 and HF varies between 0 and 0.5. The experiment is divided into two phases. In the first phase, BR is set as a default value and different HF values are tested. After full region simulation, the experimental results show that three experimental combinations (HF = {0.35, 0.45, 0.5}) have better results, of which the experimental combination HF = 0.45 produces the best result. In the second phase, HF is set as 0.45 and different BR values are tested. The experimental results show that three experimental combinations (BR = {0.95, 0.8, 0.7}) have better results, of which combination #5 (BR = 0.8, HF = 0.45) produces the best result. The six parameter combinations are as shown in Table 1 and comparative analysis of the experimental results of classification accuracy and AURC is presented in Fig. 13.

The JA experiment involves BR and HF. The experimental comparison of RST-JA and RST-GA is divided into two phases. In the first phase, BR is set at

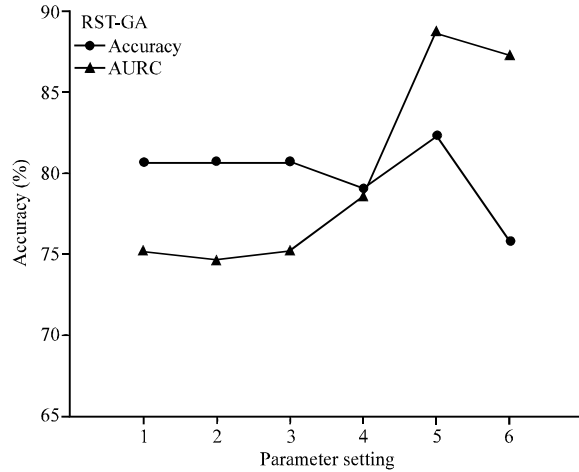


Fig. 13: Experimental results of RST-GA

Table 2: Accuracy of classification for RST-JA

No.	1	2	3	4	5	6
Boundary region	0.95	0.95	0.95	0.80	0.70	Default
Hitting fraction	0.45	0.35	0.50	0.45	0.45	0.45
Accuracy	79.03	72.58	79.03	79.03	80.65	77.42
AURC	76.75	75.13	76.75	78.27	75.21	75.63

Table 3: Comparison of AURCs for BPNN, RST-GA and RST-JA models

Method	BPNN	RST-GA	RST-JA
Conditions one (No. hidden neurons or boundary region)	4.00	0.45	0.45
Conditions two (learning rate or hitting fraction)	0.90	0.80	0.70
Accuracy (%)	90.32	82.26	80.65
AURC	99.42	88.74	75.21

0.95 and different HF values are tested. After full region simulation, the experimental results show that three experimental combinations (HF = {0.35, 0.45, 0.5}) have better results, of which the combination of HF = 0.45 produces the best result. In the second phase, HF is set at 0.45 and different BR values are tested. The experimental results show that three experimental combinations (BR = {0.7, 0.8, Default}) have better results, of which the combination of #5 (BR = 0.7, HF = 0.45) produces the best result. The six parameter combinations are as shown in Table 2 and comparative analysis of the experimental results of classification accuracy and AURC is presented in Fig. 14.

The experimental results show that in the RST-GA model, when BR = 0.8 and HF = 0.45, RST-GA is better than RST-JA and classification accuracy and AURC value are 82.26 and 88.74%, respectively. In RST-JA model, when BR = 0.7 and HF = 0.45, accuracy is 80.65% and AURC is 75.21% and its performance is poorer than that of RST-GA (Table 3).

This study used BPNN, RST-GA and RST-JA to test a medical database of vertebral columns. The experimental

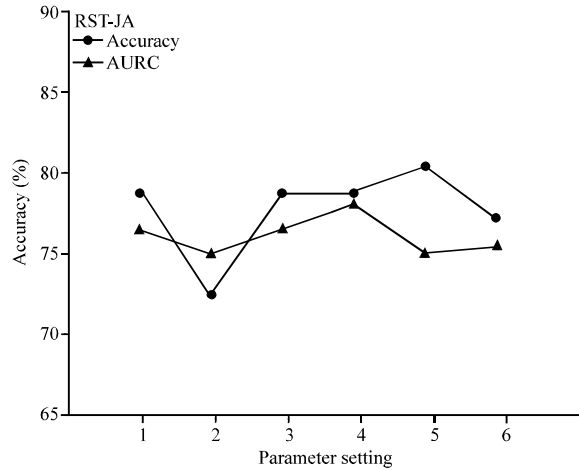


Fig. 14: Experimental results of RST-JA

results are as shown in Table 3. As seen, BPNN is superior to RST-GA and RST-JA in training results and has higher accuracy (90.32%) in medical diagnosis of vertebral column. Although the operating efficiency of RST is far higher than that of BPNN, its classification accuracy and AURC are lower than those of BPNN.

**CONCLUSION**

This study developed a diagnostic method aimed at assisting doctors in diagnosing vertebral column patients, thus, providing patients with more suitable and earlier medical care and other services. The results showed that BPNN has better classification accuracy (90.32%) and AURC (99.42%) value in vertebral column diagnosis than RST-GA and RST-JA. The BPNN classification prediction model could provide doctors and patients with an intelligent assessment system and enable the patients to receive suitable treatment and assistance and improved medical service quality, thus, facilitating a better lifestyle in a timely fashion. With the diseases of an aging population, it is important to assess the allocation of medical resources available within the medical system in order to improve medical services and care.

This research first introduces Fuzzy C-means (FCM) by explaining the formula structure and functions.

**REFERENCES**

Allaoui, H. and A. Artiba, 2009. Johnson's algorithm: A key to solve optimally or approximately flow shop scheduling problems with unavailability periods. *Int. J. Prod. Econ.*, 121: 81-87.

Chen, C.S., 2001. Biomechanical analysis of the lumbar spinal fusion. Ph.D. Thesis, National YANG-MING University, Taipei, Taiwan.

Chen, J.S., J.L. Hou, S.M. Wu and Y.W. Chang-Chien, 2010. Constructing investment strategy portfolios by combination genetic algorithms. *Expert Syst. Appl.*, 36: 3824-3828.

Cheng, W.C. and D.M. Jhan, 2013. A self-constructing cascade classifier with AdaBoost and SVM for pedestrian detection. *Eng. Appl. Artificial Intell.*, 26: 1016-1028.

Hanley, J.A. and B.J. McNeil, 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve *Radiology*, 143: 29-36.

Kramer, J., 1981. *Intervertebral Disk Diseases: Causes, Diagnosis, Treatment and Prophylaxis*. Thieme Medical Publishers Inc., Stuttgart, Germany, ISBN: 9783135824017, Pages: 221.

Li, H., C.X. Hu and Y. Li, 2012. Application of the purification of materials based on GA-BP. *Energy Procedia*, 17: 762-769.

Pawlak, Z., 1982. Rough sets. *Int. J. Comput. Inform. Sci.*, 11: 341-356.

Romero, F.T., J.C.J. Hernandez and W.G. Lopez, 2011. Predicting electricity consumption using neural networks. *IEEE Latin Am. Trans.*, 9: 1066-1072.

Sheu, T.W., C.P. Tsai, J.W. Tzeng, T.L. Chen, H.J. Chiang, W.L. Liu and M. Nagai, 2012. Using rough set and grey structural model to investigate the structure of the misconceptions' domain. *J. Grey Syst. Assoc.*, 15: 205-220.

Tay, F.E.H. and L. Shen, 2003. Fault diagnosis based on rough set theory. *Eng. Appl. Artif. Intell.*, 16: 39-43.

Trappey, A.J.C., F.C. Hsu, C.V. Trappey and C.I. Lin, 2006. Development of a patent document classification and search platform using a back-propagation network. *Expert Syst. Appl.*, 31: 755-765.

Wong, A., A. Mishra, J. Yates, P. Fieguth, D.A. Clausi and J.P. Callaghan, 2009. Intervertebral disc segmentation and volumetric reconstruction from peripheral quantitative computed tomography imaging. *IEEE Trans. Biomed. Eng.*, 56: 2748-2751.

Xie, Z., Y. Zhang and C. Jin, 2012. Prediction of coal spontaneous combustion in goaf based on the BP neural network. *Proc. Eng.*, 43: 88-92.