Scientific Research and Essays Vol. 7(23), pp. 2076-2088, 21 June, 2012
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE11.2193
ISSN 1992-2248 ©2012 Academic Journals

Full Length Research Paper

An intelligence system approach using artificial neural networks to evaluate the quality of treatment planning for nasopharyngeal carcinoma

Tsair-Fwu Lee[1*#], Pei-Ju Chao[1,2#], Chang-Yu Wang[3,4], Wei-Luen Huang[1], Chiu-Ching Tuan[5], Mong-Fong Horng[1], Jia-Ming Wu[6], Shyh-An Yeh[6], Fu-Min Fang[2] and Stephen Wan Leung[4]

[1]Medical Physics and Informatics Laboratory, Department of Electronics Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan.
[2]Department of Radiation Oncology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan.
[3]Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan.
[4]Department of Radiation Oncology, Yuan's General Hospital, Kaohsiung City, Taiwan.
[5]Department of Electronics Engineering, National Taipei University of Technology, Taipei Taiwan.
[6]Department of Radiation Oncology, E-Da Hospital /I-Shou University, Kaohsiung, Taiwan.

Accepted 30 May, 2012

The quality of the nasopharyngeal carcinoma (NPC) treatment plans evaluation using three types of artificial neural networks (ANNs) are instructed by three different training algorithms. Three ANNs including Elman (ANN_{-E}), feed-forward (ANN_{-FF}), and pattern recognition (ANN_{-PR}) were trained by using three different models, that is, leave-one-out ($Train_{-loo}$), random selection ($Train_{-random}$), and user defined ($Train_{-user}$) method. One hundred sets of NPC treatment plans were collected as the input data of the neural networks. The conformal index (CI) and homogeneity index (HI) were used as the characteristic values and also to train the neurons. Four grades (A, B, C, and D) were classified in degrading order. The over-training issue is considered between the train data and the number of neurons. The receiver operating characteristic (ROC) curves were obtained to evaluate the performed accuracies. The optimal numbers of neurons for ANN_{-E}, ANN_{-FF}, and ANN_{-PR}, in the loo method are 6, 24, and 9; in the random-selection method, they are 26, 22, and 4; and in the user-defined method they are 12, 8, and 11 neurons, respectively. The optimal size of train data is 92% of total inputs in the cases of ANN_{-E} and ANN_{-FF} and 76% in the case of ANN_{-PR}. The networks with higher accuracy are $ANN_{-PR-loo}$ (93.65 ± 3.60%), $ANN_{-FF-loo}$ (88.05 ± 5.84%), and ANN_{-E-loo} (87.55 ± 5.86%), respectively. The networks with shorter training time are $ANN_{-PR-random}$ (0.55 ± 0.11 s), $ANN_{-PR-user}$ (0.59 ± 0.08 s), and $ANN_{-PR-user}$ (1.07 ± 0.16 s), respectively. The ROC curves show that the $ANN_{-PR-loo}$ approach has the highest sensitivity, which is 99%. $ANN_{-PR-loo}$ reduces the amount of trail-and-error during the iterative process of generating inverse treatment plans. It is concluded that the $ANN_{-PR-loo}$ is an excellent model among the three for classifying the quality of treatment plans for NPC.

Key words: Artificial neural networks (ANNs), dose-volume histogram (DVH), intelligence system, nasopharyngeal carcinoma (NPC).

INTRODUCTION

Clinically, intensity modulated radiation therapy (IMRT) is the most common technique to deliver radiation doses to nasopharyngeal carcinoma (NPC) patients, because IMRT is capable of delivering a high dose to the irregular tumors, and prevents organs at risk (OARs) and normal tissues from being exposed to radiation. However, it is usually difficult to complete a suitable IMRT plan at one time because both the patient's condition and some complex formulas need to be considered simultaneously. The IMRT technique greatly benefits NPC patients, offering much higher treatment quality. The IMRT technique combines several different radiation fields to produce steep dose-volume histogram (DVH) and isodose curves for the planned target. These steep curves mean that the dose gradient at the border between cancerous and normal tissue varies rapidly. Usually the acceptable dose distribution can be produced by using seven to nine fields (Lee et al., 2008; Oldham et al., 2008).

A treatment plan that results in a higher planning target volume (PTV) coverage and reduce the complications in normal tissues is preferred, and this may be done after several attempts using trial and error. The inverse calculation is one kind of algorithm that is embedded in treatment planning systems (TPS). It is generally used in the optimization procedure for IMRT. The inverse calculation adopts iterative operation and an optimal algorithm to produce varied intensity of treatment beams. This allows IMRT to find a dose that compromises between the PTV and critical organs (Webb, 2004; Leung et al., 2007). The interactive interface is also supported in modern planning systems. The dose-volume based weighting and the priority of the critical organs can be set. Therefore, planners can define some limits for PTV and OARs, which is called constraint-based optimization.

In order to find the solution during the optimization procedure, three steps are performed: (1) determine the constraints and priority setting making up an objective function by the planner, (2) work out the objective function, and (3) evaluate the quality of the treatment plan with the prescribed dose and criteria. These three steps are executed sequentially or iteratively until an optimal solution is reached (Stieler et al., 2009). However, the quality of a final plan depends on the planners' experiences, which may be learned from others' experience or published journals (Deasy et al., 2007; Wilkens et al., 2007). It is very time-consuming for a planner to fine-tune for individual optimal solutions. Technically the final result obtained is usually not an optimal solution, but a sub-optimal one. Generally, if we want to find an optimal solution, we have to consider not only a minimum objective function, but also the individual clinical conditions and many parameters not included in the objective function. If an expert knowledge-based system is applied to learn and to accumulate those experiences, then the time taken to create an optimal treatment plan will be reduced. This knowledge-based system is especially effective for complex treatment plans, such as NPC plans.

Artificial neural networks (ANNs) are widely used in the modern sciences (Wu et al., 2009; Bahi et al., 2006; Vasseur et al., 2008). There are some major researches on ANN models that were used to predict the side effects after radiation therapy. For instant, the leave-one-out (loo), random-selection, and user-defined methods were applied to train the feed-forward ANN model introduced by Su et al. (2005) which is used to predict the probability of pneumonitis after treatment. The sensitivity obtained by the three different methods are 0.95, 0.57, and 0.71 and their respective accuracies are 0.94, 0.88, and 0.90. The ANN is also used to calculate the probability of developing radiation pneumonitis, as proposed by Chen et al. (2007). Based on Chen's model, the receiver operating characteristic (ROC) curves show that the sensitivity is 0.67, the specificity is 0.69, $p = 0.020$. Obviously, involvement of the aforementioned non-dose characteristics makes ANNs more generalized. Moreover, Mathieu et al. (2005a) adopts an ANN to optimize the dose distribution. It is applied in treatment plans to make sure the time taken for calculation is acceptable and the error is less than 2%. Isaksson et al. (2005) also uses the feed-forward network to predict the motion of a tumor in the lung during radiation therapy. Results show that this method is better than the conventional one and the self-adaptive filter. Many kinds of treatment techniques introduced by Bortfeld and Webb (2009) are used to reduce the treatment time effectively. Our preliminary result (Chao et al., 2010) shows that a back-propagation model using dose parameters and dose indices can produce high accuracy in evaluating NPC plans. Some applications of ANNs were used in the past to implement 3DCRT plans effectively and efficiently, and a few of them were applied in the expert system of IMRT.

Whether a treatment plan is acceptable or unacceptable, it usually depends on the planner's experience. It is time-consuming to evaluate the calculated results and fine-tune the weightings by trial and error. In this study, three types of ANNs are instructed by three different training algorithms to effectively evaluate the quality of the NPC treatment plans. A better match for ANN and the training algorithm will be chosen and established. We aim to help to make an intelligence judgment that reduces the amount of interaction between planner and TPS during the iterative process of generating inverse treatment plans. It can decide whether a plan is acceptable and ranking the quality of treatment plans automatically, therefore, providing an improvement suggestion when the plan was not acceptable.

MATERIALS AND METHODS

Three different neural network models, namely, the Elman (ANN_{-E}), feed-forward (ANN_{-FF}), and pattern recognition (ANN_{-PR}) models, are adopted. Each model is worked with three selection methods, named the leave-one-out (loo), random-selection, and user-defined methods, to train the neurons. The overall system flowchart is as shown in Figure 1. The data of DVHs are imported into the untrained ANNs and the training method is selected. We want to

*Corresponding author. E-mail: tflee@kuas.edu.tw or lwan@ms36.hinet.net.

#Both author contribution equally. Part of this study was presented at the International Conference on Agricultural, Food and Biological Engineering (ICAFBE 2012)

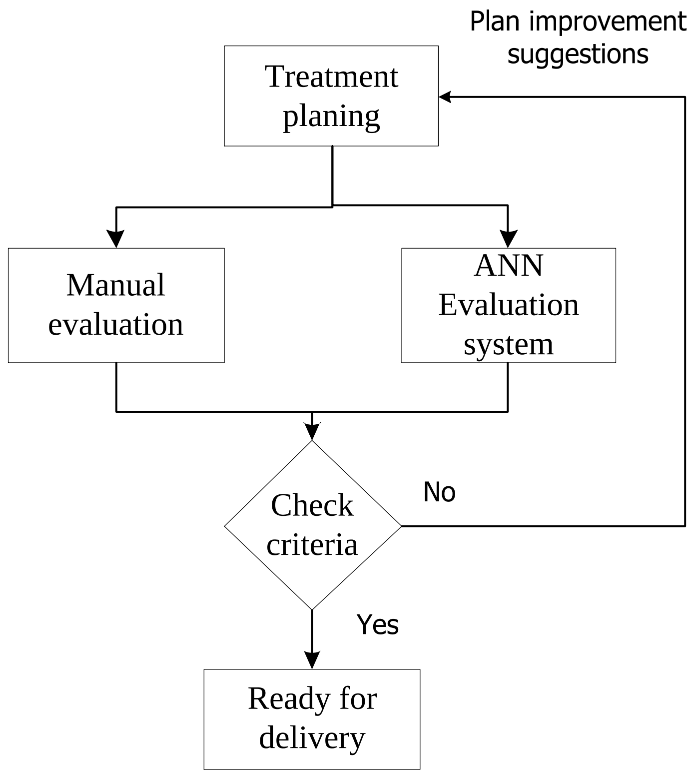


Figure 1. System flowchart for plan quality evaluation and improvement suggestions; ANN: artificial neural networks.

find an ANN model that matches a specific training method to produce the highest accuracy by consideration a mong the conditions of the training time, the number of neurons, the size of training data population, and ROC curves (Chen et al., 2007; Mathieu et al., 2005b). Then, the model can be taken as the best one to evaluate the treatment plans. The basic neural networks structure of the selection procedure is as shown in Figure 2. In the following, parameters and neural networks used are described.

Input parameters

According to the International Commission on Radiation Units and Measurements (ICRU) Report 62, the planning organ-at-risk volumes (PRVs) were defined as a safety margin around the OARs, particularly for a high-dose gradient area. In this study, the PRV of the spinal cord was determined by adding a 3D margin of at least 5 mm to the delineated spinal cord. The PRVs of the brain stem and chiasm were defined through addition of a 3D margin of at least 1 mm around the delineated structures. According to the suggestions of Radiation Therapy Oncology Group (RTOG) 0225 (Lee et al., 2003), one hundred NPC samples ($N_P = N_{100}$) are collected as the inputs, where N_P denotes the dimension of sample space and the suffix p denotes the number of samples in that group. This study was approved by the institutional review boards of the hospitals involved (IRB 99-1420B). Eventually, all the samples will be separated into four ranking classes, named A, B, C, and D. Each class is described as follows:

A: the treatment plan is accepted by physicians.
B: the prescription dose calculated on parallel organs exceeds the criteria.
C: the prescription dose calculated on serial organs exceeds the criteria.
D: the coverage of PTV does not meet the criteria.

The energy selected is of a 6 MV photon beam and a seven-field IMRT plan is created (Chao et al., 2010). This leads to the reduction of treatment time and enhance the biological effect. Each NPC plan has its own dosimetric indices and parameters (Lee et al., 2010, Lee et al., 2011, Fang et al., 2010; Widesott et al., 2008; Leung et al., 2007), which are discussed subsequently.

Dosimetric parameters

Planning target volume

Three parameters are commonly used to evaluate the coverage of the PTV. V_{93} means 93% of the total dose is received by 97% of PTV. The parameter V_{100} means 100% of the prescription dose covers more than 95% of PTV. Similarly, V_{110} means that 110% of the dose covers less than or equal to 20% of the PTV.

Constraints for the organs at risk

1) Spinal cord (SC): The maximum dose ≤ 45 Gy or 1 cc of PRV ≤ 50 Gy;
2) Brain stem (BS): The maximum dose ≤ 54 Gy or 1% of PRV ≤ 60 Gy;
3) Chiasm: The maximum dose ≤ 54 Gy or maximum dose of PRV ≤ 60 Gy;
4) Parotid: The mean dose ≤ 26 Gy or V_{30Gy} ≤ 50%;

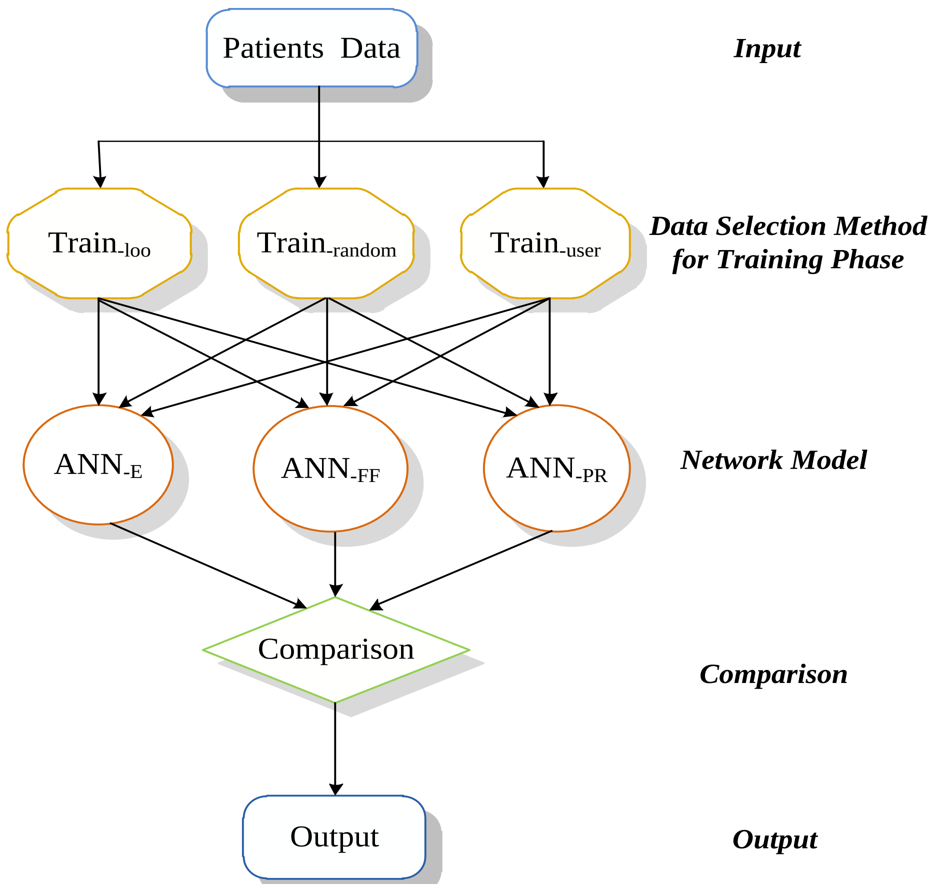


Figure 2. Basic neural networks structure. ANN: artificial neural networks; ANN-$_{E}$: the Elman network; ANN-$_{FF}$: the feed-forward network; ANN-$_{PR}$: a pattern recognition network; Train-$_{loo}$: ANN with leave-one-out method for training data selection; Train-$_{random}$: ANN with random selection method for training data selection; Train-$_{user}$: ANN with user-defined method for training data selection.

5) Lens: The maximum dose must be ≤ 10 Gy and as low as possible;
6) Eyes: the maximum dose must be ≤ 45 Gy;
7) Mandible: The maximum dose must be ≤ 70 Gy or 1 cc of PRV and cannot exceed 75 Gy;
8) Oral cavity excluding PTV: the mean dose must be ≤ 40 Gy;
9) Healthy tissue: the mean dose must be ≤ 30 Gy or no more than 1% or 1 cc of the tissue outside the PTV will receive ≥ 110% of the dose prescribed to the PTV.

Dosimetric indices

Conformal index (CI)

This is used to estimate the coverage of PTV (Feuvret et al., 2006).

$$CI = V_{PTV} \times \frac{V_{TV}}{TV_{PV}{}^{2}}$$

where V_{TV} is the treatment volume of prescribed isodose lines, V_{PTV} is the volume of PTV, and TV_{PV} is the volume of V_{PTV} within V_{TV}. The best conformal case is the value of *CI* equal to 1.

Homogeneity index (HI)

This index describes how the homogeneity varies within the PTV.

$$HI = \frac{D_{5\%}}{D_{95\%}}$$

where $D_{5\%}$ and $D_{95\%}$ are the minimum doses delivered to 5 and 95% of the PTV. A higher HI indicates poorer homogeneity.

Therefore, there are 14 indices included in D = [V_{93}, V_{100}, V_{110}, SC, BS, rt Parotid, lt Parotid, Lens, rt Eye, lt Eye, Oral, Mandible, CI, HI] which are presented in this paper as the input vector (rt: right side, lt: left side).

Training parameters

Three distinct training methods are considered separately as follows:

1) Leave-one-out ($Train_{-loo}$): This method selects one set of the patient's data at a time to be the validation vector, and the other sets (N_p – 1) are used to train the network until N_p iterations have been done. It is suitable for use when the amount of data is small and high accuracy is needed.
2) Random selection ($Train_{-random}$): Here, four kinds of arrangement are made. The first kind selects 60% of the patient's data (N_{60}) to be the training data and the remaining 40% are used as the test vector. In the second arrangement, 67% of data are chosen to be the training data and the remainder is the test vector; in the third arrangement, the training data and test vector comprise 75 and 25% of the data, respectively, and in the fourth, 80 and 20%, respectively.
3) User-defined ($Train_{-user}$): This method is similar to $Train_{-random}$, but differs in that the data are selected manually. We prefer to select typical data to train the network, because it is easier to describe the border of data when the population of data is small.

The cross-validation method is adopted to prevent ANNs from over-training. If an ANN model is over-trained, non-generalized problems may occur as a result; for example, the training time may be too long or the accuracy may fall below an acceptable level, and so on.

Artificial neural networks

ANN is a brain-like network that possesses self-learning and memory abilities. The signal routes of a network are similar to the axons that carry the electrical signal out to other layers (Gulliford et al., 2004). Three types of ANNs are instructed by three different training algorithms to evaluate the quality of the plans. MATLAB (v 7.9, The MathWorks, Natick, Massachusetts) is adopted to construct the network and to evaluate the treatment plans.

Feed-forward network

Many kinds of structures can be produced by combining learning modes and nodes. A basic one is the feed-forward network (ANN_{-FF}), it has three main layers (Isaksson et al., 2005; Deasy et al., 2007; Bortfeld and Webb, 2009; Luo et al., 2005; Zhang et al., 2010):

1) Input layer: This layer actually consists of input components. Generally, it has two types: one whose input components are weighted in neurons with a bias value, and another which just connects the input components to neurons directly without operations. Here, the first type is adopted.
2) Hidden layer: The intermediate layer between input and output layers. This layer receives input signals and processes them with a defined transfer function. The number of layers could be zero or multiple. Here, there is one hidden layer. It should be noted that the number of neurons will decide the speed of convergence.
3) Output layer: A layer whose output is the output of the network. An error between the output and the actual value could feed backward to the weighting matrix. This feedback procedure is done until the output converges.

Elman network

The Elman network structure (ANN_{-E}) (Cheng et al., 2002) is a simple recursive network (SRN). In an Elman network, some outputs of the hidden layer feed information back to the input layer. Those components are called context units and their weightings are fixed. This mechanism produces a signal which returns to where it came from, and thus the Elman network acts like a dynamic memory that can remember some previous information temporarily. The weighting is adjusted by an error back-propagation algorithm. The linear transfer function is adopted in the input layer and the output layer, but a hyperbolic tangent function is used in the hidden layer. A factor δ called self-feedback gain denotes the state of the previous inputs contained in a context component (Liu et al., 2006; Qi et al., 2008; Su et al., 2007; Yuan-Chu et al., 2008). When δ approaches 1, this means that a context unit possesses more previous states. Otherwise, the network returns to the standard Elman network as δ is equal to zero. The value δ is chosen to be zero in our study.

Pattern recognition network

A pattern recognition network (ANN_{-PR}) (Duin et al., 2007) can be trained to classify inputs according to target classes. The target data for pattern recognition networks should consist of vectors of all zero values except for a 1 in element i, where i is the class they are to represent. The input patterns are sensed and transformed into measurements by an ANN_{-PR}. The features will be extracted from those measurements during the preprocessing procedure and features extraction. Therefore, the ANN_{-PR} can recognize the patterns with respect to the features (Klopf and Gose, 1969; Lee and Bezdek, 1988; Maruno et al., 1993; Ulug, 1996; Weaver, 1975). In this study, the transfer function of the input layer is linear. Besides, there is only one stage in the hidden layer, whose transfer function is a hyperbolic tangent. The arrangement of the output layer is the same as that of the hidden layer here.

Three networks, named Elman, feed-forward, and pattern recognition, are adopted in this study. During training, the data are selected by the loo, random selected, and user defined methods separately. The over-training problem is also discussed in the following along with the inputted data and the number of neurons, and the accuracies of the three algorithms are evaluated by using ROC curves.

Statistical analysis

Statistical tests of differences between the models were performed using a two-tail matched-pair exact Student t-test. Differences were considered statistically significant for *p*-values ≤ 0.05. All data presented in the text, tables and figures refer to the mean and standard deviation. The Statistical Package for Social Sciences (SPSS)-16.0 software was used for data processing (SPSS, Inc., Chicago, IL, USA).

RESULTS

Neurons and training data

The cross-validation method is used to prevent the over-training situation from occurring for evaluating the number of neurons and the amount of training data; the results on two issues are as follow:

1) Training data: The accuracies of the system and training time of networks versus the size of training data is shown in Table 1, the simulated results for ANN_{-E} and ANN_{FF} show that the optimal size of training data is 84% of total inputs. In the case of ANN_{-PR}, the optimal size of training data is 76% of total inputs.
2) Number of neurons: Relations between the number of neurons and neural networks are as shown in Figure 3.

Table 1. The result of the ANN_{-E}, ANN_{-FF}, and ANN_{-PR} (the optimal amount of training data can be selected for three ANNs).

The amount of training data (%)	ANN_{-E}		ANN_{-FF}		ANN_{-PR}	
	Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)
60	75.38 ± 1.22	0.99 ± 0.05	68.74 ± 1.21	4.60 ± 0.82	76.38 ± 1.27	0.56 ± 0.04
68	78.06 ± 1.27	0.99 ± 0.05	69.97 ± 1.97	4.55 ± 0.98	78.63 ± 1.22	0.56 ± 0.03
76	79.03 ± 1.00	0.99 ± 0.04	72.34 ± 1.72	4.85 ± 0.95	81.27 ± 1.35	0.57 ± 0.04
84	81.18 ± 1.94	0.97 ± 0.06	75.49 ± 2.43	5.36 ± 0.82	81.83 ± 2.33	0.59 ± 0.04

ANN_{-E}: The Elman network; ANN_{-FF}: the feed-forward network; ANN_{-PR}: a pattern recognition network; All data presented refer to the mean and standard deviation.

The three data selection methods are run with one model at a time for the optimal numbers of neurons selection. Table 2 shows that the training data are selected by the loo method and the optimal numbers of neurons for ANN_{-E}, ANN_{-FF}, and ANN_{-PR} are 6, 24, and 9, respectively. For the random-selection method, the optimal numbers of neurons are 26, 22, and 4, as shown in Table 2. Therefore, the optimal numbers with respect to the user-defined method are 12, 8, and 11, respectively. It should be mentioned that the condition of the training data used here is the same as in Table 1.

Accuracies and training time versus training data

The condition of the optimal neurons here is inherited from Table 2. Through this condition, the neural models are tested with varied training data to find out the most adaptive ones. Experimental results are as follows:

1) The three models with the highest accuracy are $ANN_{-PR\text{-}loo}$ (93.65 ± 3.60%), $ANN_{-FF\text{-}loo}$ (88.05 ± 5.84%), and $ANN_{-E\text{-}loo}$ (87.55 ± 5.86%).
2) The three models with the shortest training time are $ANN_{-PR\text{-}random}$ (0.55 ± 0.11 s), $ANN_{-PR\text{-}user}$ (0.59 ± 0.08 s), and $ANN_{-E\text{-}user}$ (1.07 ± 0.16 s).
3) The ANN_{-FF} model is used as a benchmark for statistical comparison. Statistical significance is deemed a p-value < 0.05, and the accuracy of $ANN_{-PR\text{-}loo}$ and $ANN_{-PR\text{-}random}$ models is found to be the statistical best ones.

ROC curves

The ROC curve is used to estimate the adaptability of the neural networks as depicted in Figure 4. The sensitivity specificity analyses are listed in Table 3. The best case is $ANN_{-PR\text{-}loo}$, which has 99% sensitivity and 100% specificity. The worst case is $ANN_{-E\text{-}user}$, which has 67% sensitivity and 64% specificity.

Error estimation

Through the classification of three algorithms, the highest missing rate occurred in the 97 and 98th dataset as shown in Figure 5. Each dataset missed six times in total.

Plan improvement

Plan improvement suggestions are given after an execution of the neural models in a pop-up window in Figure 6. One example before/after improvement is demonstrated in Figure 7, the DVHs and isodose curves are included. Plan improved from rank C (left hand side) to rank A (right hand side).

DISCUSSION

To find an optimal size of train data to avoid an over-training problem is a primary goal in our research. So, $Train_{-user}$ is adopted here instead of $Train_{-random}$, because we want the input data to be distributed uniformly, otherwise the amount of train data will increase dramatically and then the optimal population size will never be found. For ANN_{-E}, there is no significant difference for the system accuracy in the amounts of training data used (of 76 and 84%) among the three models as shown in Table 1. But the training time for 84% train data is shorter than the others, so the amount of training data (84%) is adopted in the case of ANN_{-E}. In the case of ANN_{-FF}, the amount of training data of 84% is taken, because the accuracy is highest. However, in the ANN_{-PR} model, the outcomes in the amount of training data of 76 and 84% used are almost the same. Based on the reason for the less training time, the 76% is taken.

The accuracies and training time of the ANN_{-E}, ANN_{-FF}, and ANN_{-PR} models can be estimated after the number of neurons and the size of train data have been decided. Table 2 shows that the $ANN_{-PR\text{-}loo}$ possesses the highest accuracy, but with much longer training time, and the size of train data for $Train_{-loo}$ is fixed. In the same loo method group the other two results for $ANN_{-E\text{-}loo}$ and $ANN_{-FF\text{-}loo}$ are similar. On average, the accuracy is highest in the group $Train_{-loo}$, followed by the group $Train_{-user}$ and lastly the group $Train_{-random}$. The advantage of $Train_{-loo}$ models is that all of their parameters have been trained to result in

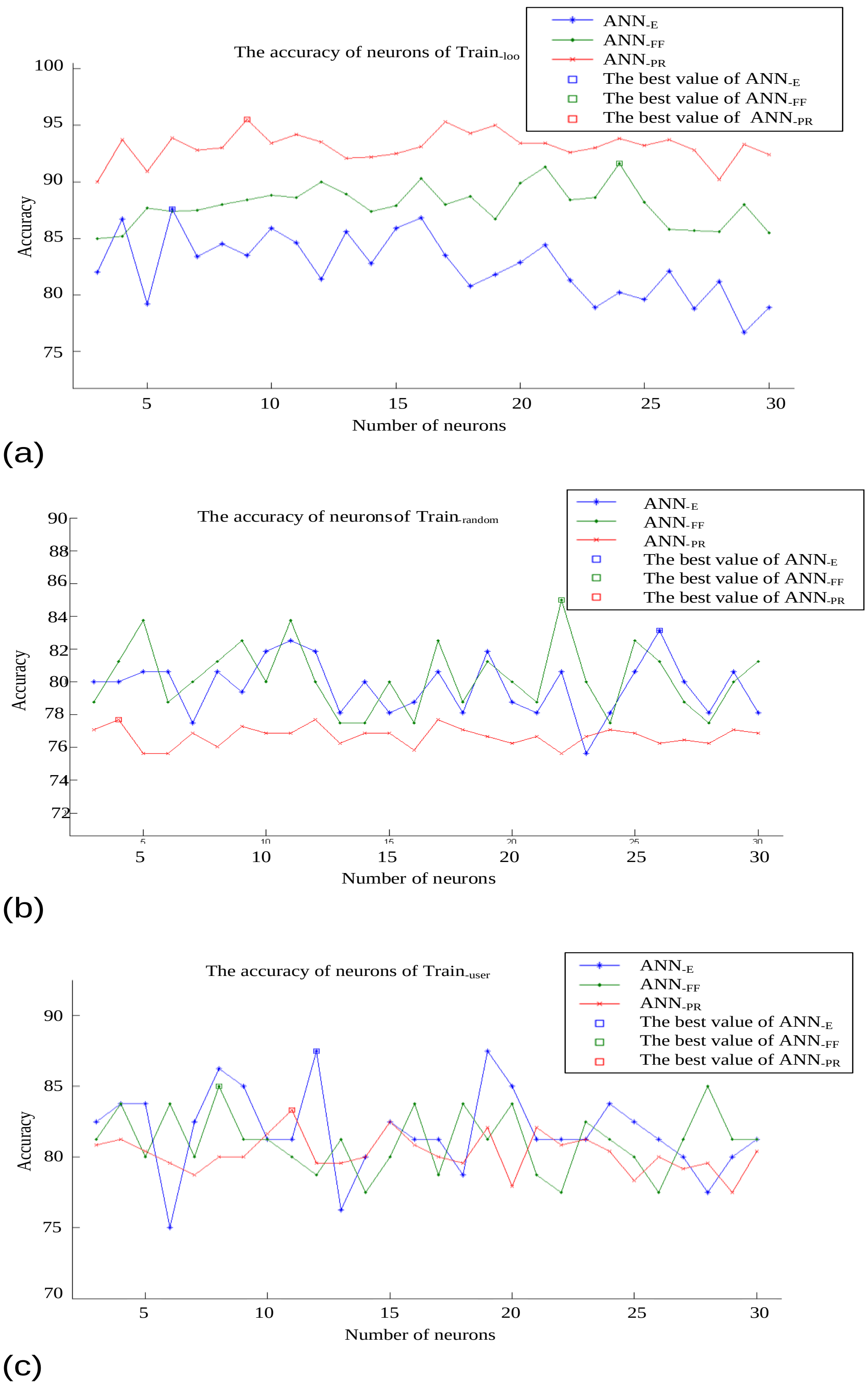


Figure 3. Relations between the number of neurons and neural networks; (a) for Train$_{-loo}$ (b) for Train$_{-random}$, and (c) for Train$_{-user}$. (Square marked with the best solutions). ANN$_{-FF}$: The feed-forward network; ANN$_{-E}$: the Elman network; ANN$_{-PR}$: A pattern recognition network; Train$_{-loo}$: ANN with leave-one-out method for training data selection; Train$_{-random}$: ANN with random selection method for training data selection; Train$_{-user}$: ANN with user-defined method for training data selection; Statistical tests of differences between the models were performed using a two-tail matched-pair exact Student t-test. ANN$_{-FF}$ was used as a benchmark. Differences were considered statistically significant for p-values ≤ 0.05; NS: not statistical significance.

Table 2. The result of the ANN_{-E}, ANN_{-FF}, and ANN_{-PR} for the three different data selection methods (the best number of neurons are selected).

Method	Input data (%)	Number of neurons	Accuracy (%)	p value	Training time (s)	p value
Leave-one-out method ($Train_{-loo}$)						
ANN_{-E}	100	6	87.55 ± 5.86	0.782^{NS}	43.62 ± 1.44	0.186^{NS}
ANN_{-FF}	100	24	88.05 ± 5.84	-	42.38 ± 4.38	-
ANN_{-PR}	100	9	93.65 ± 3.60	<0.005	37.34 ± 0.58	<0.005
Random selection method ($Train_{-random}$)						
ANN_{-E}	84	26	83.13 ± 7.34	0.039^{NS}	1.08 ± 0.13	<0.005
ANN_{-FF}	84	22	85.00 ± 5.27	-	4.83 ± 2.23	-
ANN_{-PR}	76	4	77.71 ± 2.80	<0.005	0.55 ± 0.11	<0.005
User-defined method ($Train_{-user}$)						
ANN_{-E}	84	12	87.50 ± 11.79	0.642^{NS}	1.07 ± 0.16	0.007
ANN_{-FF}	84	8	85.00 ± 9.86	-	4.33 ± 3.00	-
ANN_{-PR}	76	11	84.38 ± 6.21	0.648^{NS}	0.59 ± 0.08	<0.005

$Train_{-loo}$: ANN with leave-one-out method for training data selection; $Train_{-random}$: ANN with random selection method for training data selection; $Train_{-user}$: ANN with user-defined method for training data selection; ANN_{-E}; the Elman network; ANN_{-FF}; the feed-forward network; ANN_{-PR} : a pattern recognition network; All data presented refer to the mean and standard deviation. Statistical tests of differences between the models were performed using a two-tail matched-pair exact Student t-test. ANN_{-FF} was used as a benchmark. Differences were considered statistically significant for p-values ≤ 0.05; NS: not statistical significance.

higher accuracy; however, the disadvantage is that the training time is much longer. The common factor in the groups $Train_{-user}$ and $Train_{-random}$ is that the sizes of the training populations adopted are the same. But there is a slight difference in accuracy between them, because of the data-selection methods. The user-selection model possesses higher average accuracy than the random-selection model, but the average training time is almost the same.

$ANN_{-PR-loo}$ is the most precise model and is also faster than ANN_{-E-loo} and $ANN_{-FF-loo}$. So, $ANN_{-PR-loo}$ is suitable for the cases where high accuracy is required. The training time for $ANN_{-PR-user}$ remains within one second with 84.38% accuracy rate, which is most suitable for some real-time applications.

As shown in Figure 5, classification errors occurred six times in dataset 97 and 98th. Dataset reviewed showed that the dataset 97th belongs to class D based on the RTOG 0225 criteria and the department standard; meanwhile, the PTV margin overlaps BS's and parotid's margins, which belong to class B. This increases the probability of error in classification. An enlarged margin was used to deal with some specific marginal situations which caused a marginal overlapping problem. This also happens on dataset 98, whose margin overlaps SC's, BS's, parotids', and lens' margins, which are classified into class C. However, the overlapping situation can be effectively improved with a larger size of training data population, as occurred with $Train_{-loo}$ on dataset 98th.

In terms of medical examination, the number of false negatives (1-sensitivity) should be as low as possible, because this means the sensitivity is greater. Therefore, the ROC results show that $ANN_{-PR-loo}$ is the optimal solution among the three in this study.

The ANN_{-PR} model has not only the shortest training time, as stated by Lampariello and Sciandrone, (2001), but also the highest accuracy. The multilayer perceptron that was introduced by Kolasa et al. (2009) possesses five neurons in the hidden layer and is used to predict the survival rate of patients with bladder cancer. The accuracy of Kolasa's model is 90%, but the accuracy of $ANN_{-PR-loo}$ presented in this study is 99%. Bassi et al. (2007) proposed a method that adopts 93 sets of data to train neurons. The ratio of training data to validation data is 1:1 (model A) and the ratio in model B is 2:1. The results show that the areas under the ROC curves are 0.89 and 0.88. However, with our $Train_{-loo}$ algorithm, the area under any ROC curve is 0.96 or above. Besides, the ratio of training data to validation data is 19:6 (approximates to 3:1) in $ANN_{-PR-user}$ and the area under the ROC is 0.89. Mathieu (2005a) proposed the idea of adopting a neural network to optimize the dose distributions. Therefore, we adopted this idea in our ANN_{-FF} model to improve the speed of the calculation. The accuracy of our model is the same as that of the conventional and self-adaptive filter methods, both of which are proposed by Isaksson et al. (2005).

The neural networks use dosimetric indices and parameters from DVHs to classify the quality of treatment plans, in this study. Three kinds of ANN are run along

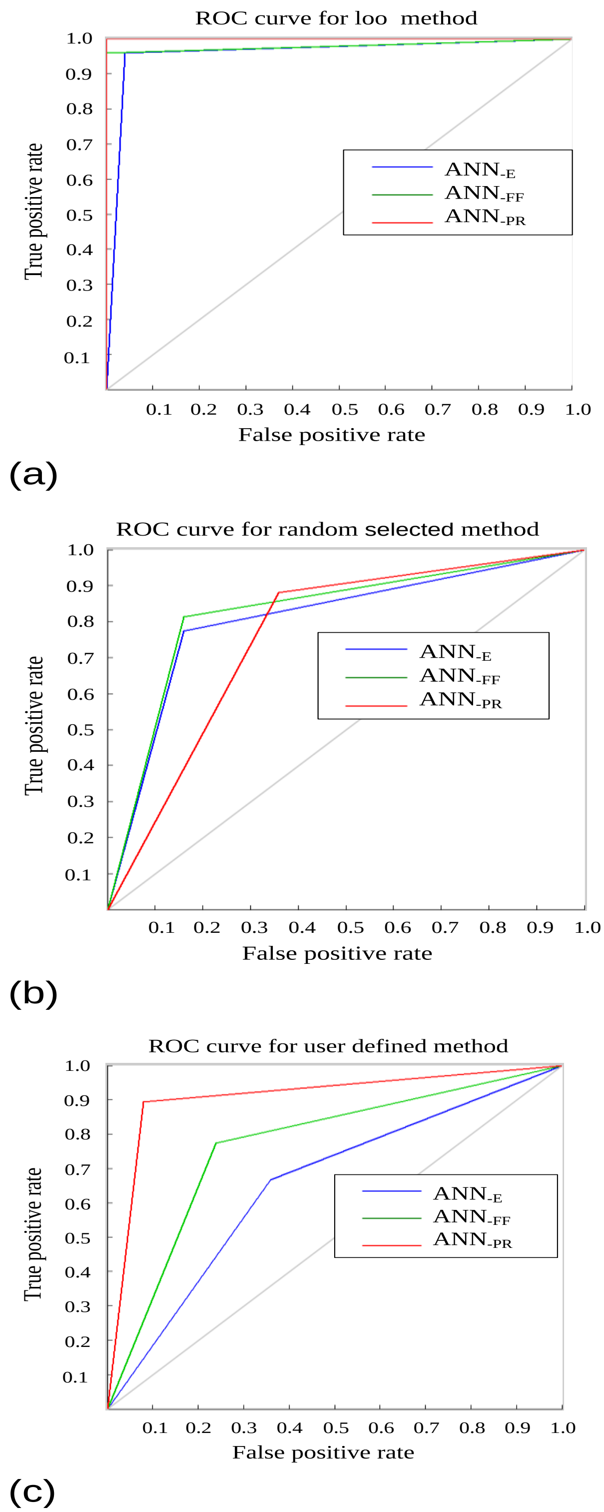


Figure 4. Three different ANNs ROC curves from (a) $Train_{-loo}$ (b) $Train_{-random}$ (c) $Train_{-user}$. ANN: artificial neural networks; ANN_{-E}: the Elman network; ANN_{-FF}; the feed-forward network; ANN_{-PR}: a pattern recognition network; ROC: receiver operating characteristic; $Train_{-loo}$: ANN with leave-one-out method for training data selection; $Train_{-random}$: ANN with random selection method for training data selection; $Train_{-user}$: ANN with user-defined method for training data selection.

Table 3. Sensitivity and specificity analysis for the models.

Method	Sensitivity (%)	Specificity (%)
Leave-one-out method		
ANN_{-E}	96	4
ANN_{-FF}	96	0
ANN_{-PR}	99	0
Random selection method		
ANN_{-E}	77	16
ANN_{-FF}	81	16
ANN_{-PR}	88	36
User-defined method		
ANN_{-E}	67	36
ANN_{-FF}	77	24
ANN_{-PR}	89	8

ANN_{-FF}: The feed-forward network; ANN_{-E}: the Elman network; ANN_{-PR}: a pattern recognition network.

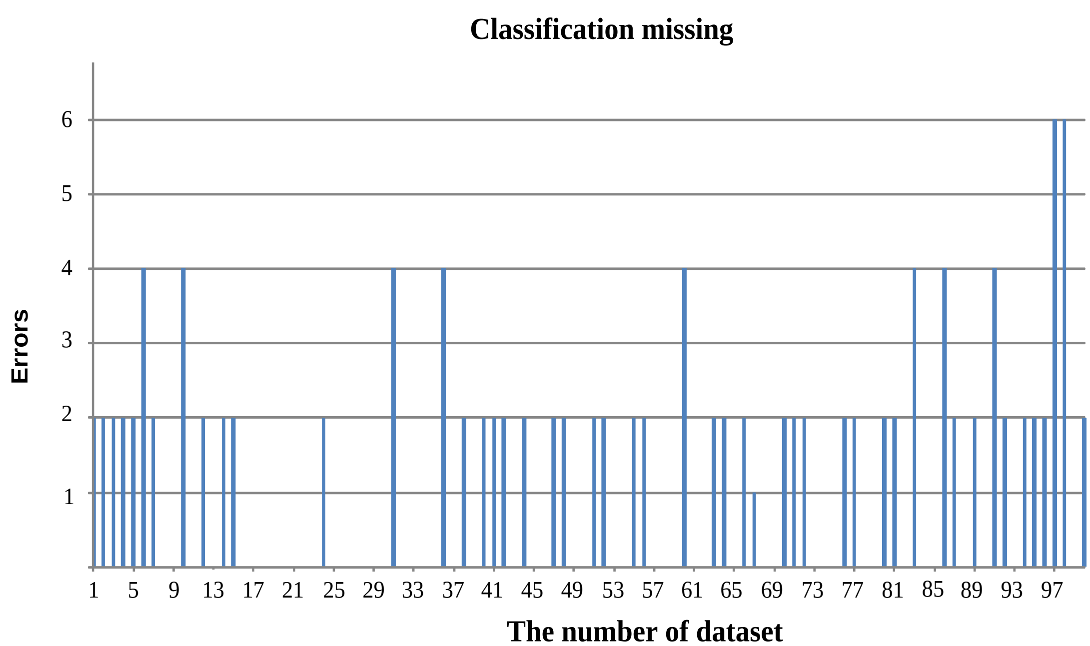
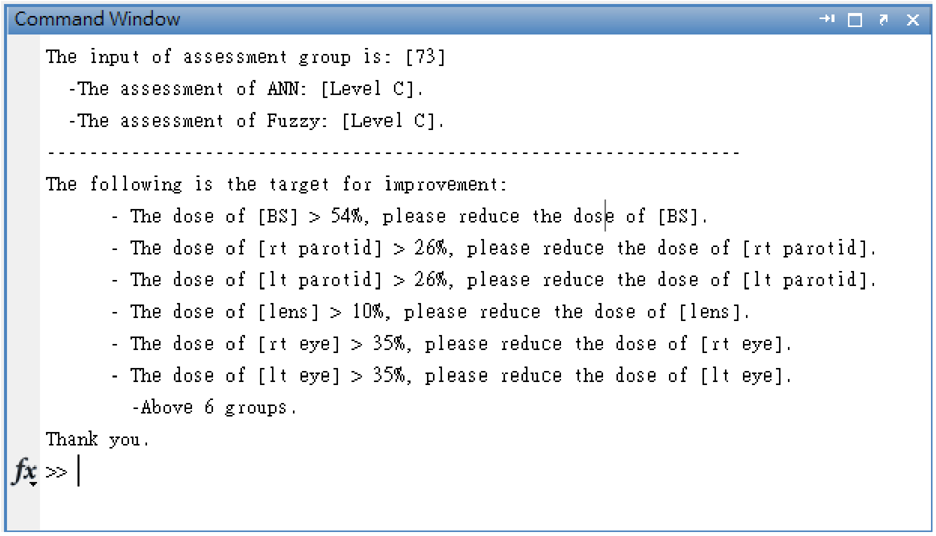


Figure 5. Classification missed evaluation in dataset.

with three data selection methods, named the loo, random-selection, and user-selection methods. The answers we found on how the training data population and training data quality affect the operation of each model and the ROC is also an important factor in quantifying the model's accuracy. Based on the results, $Train_{-loo}$ possesses high accuracy, but its training time is too long to be applied in some real-time applications. If we need a model that is capable of an instant response, then $Train_{-random}$ or $Train_{-user}$ is a suitable choice. However, $Train_{-user}$ needs to execute one more operation to avoid biased distribution of the training data and to keep the accuracy within an acceptable range. Summarizing all the findings and considerations mentioned earlier, the best neural network for application to the evaluation of the quality of treatment plans is $ANN_{-PR-loo}$. The overlapping problem in this model can be solved by importing more training samples.


Command Window

```
The input of assessment group is: [73]
  -The assessment of ANN: [Level C].
  -The assessment of Fuzzy: [Level C].
-----------------------------------------------------------------
The following is the target for improvement:
      - The dose of [BS] > 54%, please reduce the dose of [BS].
      - The dose of [rt parotid] > 26%, please reduce the dose of [rt parotid].
      - The dose of [lt parotid] > 26%, please reduce the dose of [lt parotid].
      - The dose of [lens] > 10%, please reduce the dose of [lens].
      - The dose of [rt eye] > 35%, please reduce the dose of [rt eye].
      - The dose of [lt eye] > 35%, please reduce the dose of [lt eye].
        -Above 6 groups.
Thank you.
fx >>
```

Figure 6. Plan improvement suggestions.

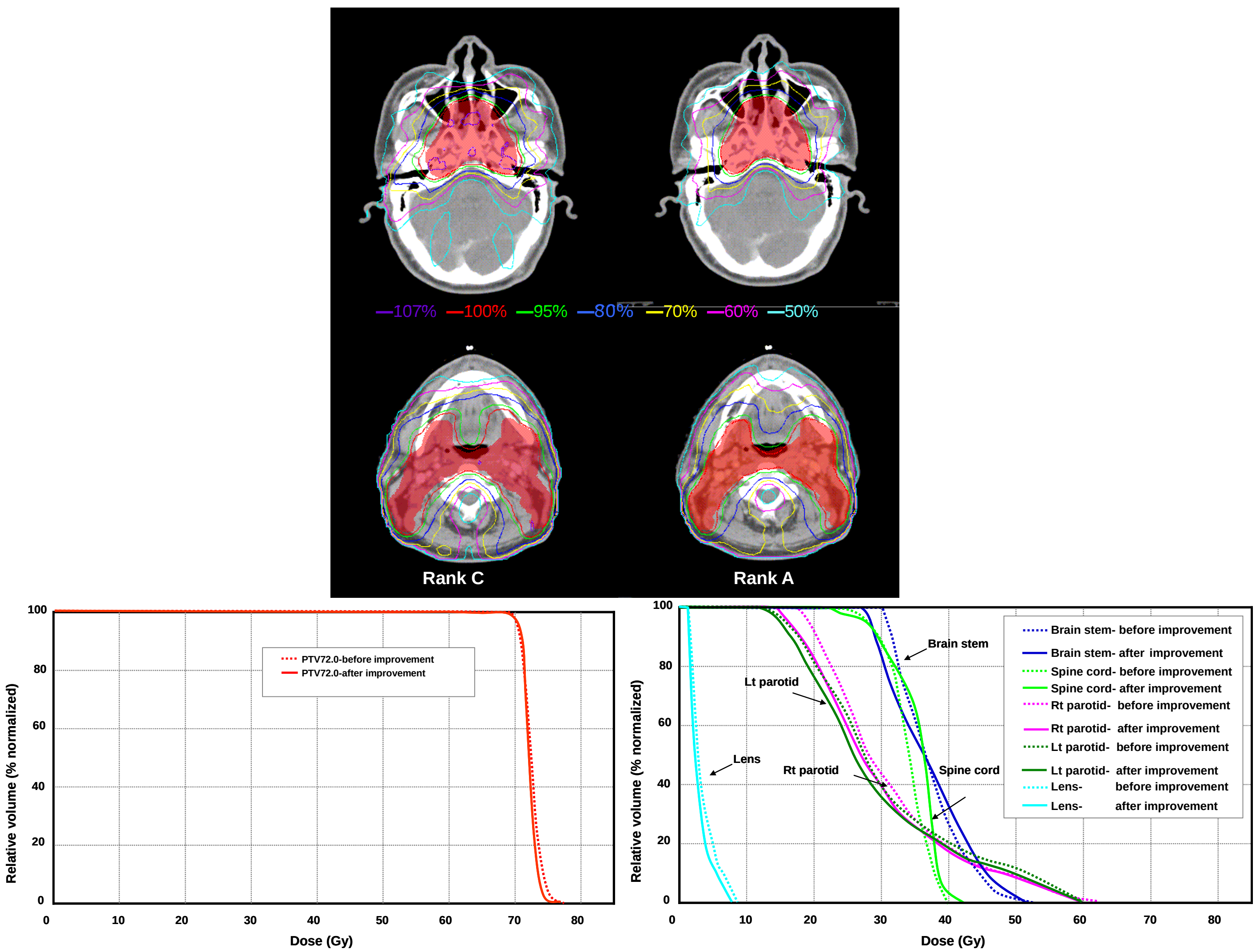


Figure 7. A planning sample before (left)/after (right) improvement (improved from rank C to rank A).

Conclusions

$ANN_{-PR-loo}$ reduces the amount of trail-and-error during the iterative process of generating inverse treatment plans. It is concluded that the $ANN_{-PR-loo}$ is an excellent model among the three for classifying the quality of treatment plans for NPC. This system is able to classify the calculated result and offer suggestions to planners that reduce the amount of interaction between planner and TPS during the iterative process of generating inverse treatment plans. It is a convenient and effective way to evaluate the quality of treatment plans.

ACKNOWLEDGEMENTS

This study was supported financially, in part, by grants from Kaohsiung Chang Gung Memorial Hospital (CGMH) and the National Science Council (NSC) of the Executive Yuan of the Republic of China. “CMRPG890062”, “NSC 100-2221-E-151-003.

REFERENCES

Bahi J, Contassot-Vivier S, Makovicka L, Martin E, Sauget M (2006). Neural network based algorithm for radiation dose evaluation in heterogeneous environments. Artif. Neural Networks–ICANN, pp. 777-787.
Bassi PF, Sacco E, De Marco V, Aragona M, Volpe A (2007). Prognostic Accuracy of an Artificial Neural Network in Patients Undergoing Radical Cystectomy for Bladder Cancer: a Comparison with Logistic Regression Analysis. BJU int., 99(5): 1007-1012.
Bortfeld T, Webb S (2009). Single-Arc IMRT? Phys. Med. Biol., 54(1): N9-20.
Chao PJ, Lee TF, Huang WL, Chen LC, Su TJ, Chen WP (2010). Applications of Artificial Neural Network to Building Statistical Models for Qualifying and Indexing Radiation Treatment Plans. World Acad. Sci. Eng. Technol., 70: 37-42.
Chen S, Zhou S, Zhang J, Yin FF, Marks LB, Das SK (2007). A Neural Network Model to Predict Lung Radiation-Induced Pneumonitis. Med. Phys., 34(9): 3420-7.
Cheng Y, Qi WM, Cai W (2002). Dynamic properties of Elman and modified Elman neural network. in IEEE, 2: 637-640.
Deasy JO, Alaly JR, Zakaryan K (2007). Obstacles and Advances in Intensity-Modulated Radiation Therapy Treatment Planning. Front. Radiat. Ther. Oncol., 40: 42-58.
Duin R, Juszczak P, Paclik P, Pekalska E, De Ridder D, Tax D, Verzakov S (2007). A matlab toolbox for pattern recognition. Delft University of Technology, PRTools4. 1.
Fang F, Tsai W, Lee T, Liao K, Chen H, Hsu H (2010). Multivariate analysis of quality of life outcome for nasopharyngeal carcinoma patients after treatment. Radiother. Oncol., 97(2): 263-269.
Feuvret L, Noel G, Mazeron JJ, Bey P (2006). Conformity index: A review. Int. J. Radiat. Oncol. Biol. Phys., 64(2): 333-342.
Gulliford SL, Webb S, Rowbottom CG, Corne DW, Dearnaley DP (2004). Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. Radiother. Oncol., 71(1): 3-12.
Isaksson M, Jalden J, Murphy MJ (2005). On Using an Adaptive Neural Network to Predict Lung Tumor Motion During Respiration for Radiotherapy Applications. Med. Phys., 32(12): 3801-9.
Klopf AH, Gose E (1969). An Evolutionary Pattern Recognition Network. Systems Science and Cybernetics, IEEE Trans., 5(3): 247-250.
Kolasa M, Wojtyna R, D ugosz R and Jo wicki W (2009). Application of Artificial Neural Network to Predict Survival Time for Patients with Bladder Cancer. Comput. Med. Act., pp. 113-122.
Lampariello F and Sciandrone M (2001). Efficient Training of RBF Neural Networks for Pattern Recognition. IEEE Trans Neural Netw., 12(5): 1235-1242.
Lee JSJ, Bezdek JC (1988). A Feature Projection Based Adaptive Pattern Recognition Network. in Neural Networks, 1988. IEEE Int. Conf., 1: 497-505.
Lee N, Kramer A, Xia P (2003) 'RTOG 0225: a phase II study of intensity modulated radiation therapy (IMRT)+/-chemotherapy for nasopharyngeal cancer'.
Lee TF, Chao PJ, Fang FM, Su TJ, Leung SW, Hsu HC (2010). Helical Tomotherapy for Single and Multiple Liver Tumours. Radiather. Oncol., 5: 58.
Lee TF, Chao PJ, Wang CY, Lan JH, Huang YJ, Hsu HC, Sung CC, Su TJ, Lian SL and Fang FM (2011). Dosimetric Comparison of Helical Tomotherapy and Dynamic Conformal arc Therapy in Stereotactic Radiosurgery for Vestibular Schwannomas. Med. Dosim., 36(1): 62-70.
Lee TF, Fang FM, Chao PJ, Su TJ, Wang LK, Leung SW (2008). Dosimetric Comparisons of Helical Tomotherapy and Step-And-Shoot Intensity-Modulated Radiotherapy in Nasopharyngeal Carcinoma. Radiother. Oncol., 89(1): 89-96.
Leung LHT, Kan MWK, Cheng ACK, Wong WKH, Yau CC (2007). A new dose–volume-based Plan Quality Index for IMRT plan comparison. Radiother. Oncol., 85(3): 407-417.
Liu H, Wang S, Ouyang P (2006). Fault Diagnosis Based on Improved Elman Neural Network for a Hydraulic Servo System. in Robotics, Automation and Mechatronics, 2006 IEEE Conf., pp. 1-6.
Luo J, Yu G, Yang H, Zhao H, Ai B (2005). Bus Protection ANN Model with Function Approximation Ability. in Power Engineering Conference, 2005. IPEC 2005. The 7th Int., pp. 1-478.
Maruno S, Imagawa T, Kohda T, Shimeki Y (1993). Object Recognition System Using Temporal Pattern Recognition Networks with Quantizer Neuron Chip. in Neural Networks, 1993. IJCNN '93-Nagoya. Proceed. 1993 Int. Joint Conf., 2: 1285-1288.
Mathieu R, Martin E, Gschwind R, Makovicka L, Contassot-Vivier S, Bahi J (2005a). Calculations of Dose Distributions Using a Neural Network Model. Phys. Med. Biol., 50(5): 1019-28.
Mathieu R, Martin E, Gschwind R, Makovicka L, Contassot-Vivier S, Bahi J (2005b). Calculations of dose distributions using a neural network model. Phys. Med. Biol., 50: 1019.
Oldham M, Sakhalkar H, Guo P, Adamovics J (2008). An investigation of the accuracy of an IMRT dose distribution using two-and three-dimensional dosimetry techniques. Med. Phys., 35: 2072.
Qi Z, Liu M, Wang H(2008). Process Modeling Method Based on an Improved Elman Neural Network. in IEEE, pp. 8188-8192.
Stieler F, Yan H, Lohr F, Wenz F, Yin F-F (2009). Development of a neuro-fuzzy technique for automated parameter optimization of inverse treatment planning. Radiather. Oncol., 4(1): 39.
Su G, Wang L, Xu Y, Wang X (2007). Date Gas Load Forecasting With OIF-Elman Network. in Control Conf. 2007. CCC 2007. Chinese, pp.167-170.
Su M, Miften M, Whiddon C, Sun X, Light K and Marks L (2005). An Artificial Neural Network for Predicting the Incidence of Radiation Pneumonitis. Med. Phys., 32(2): 318-25.
Ulug ME (1996). Combining Fuzzy Pattern Recognition and Fuzzy Control in an AI Driven Neural Network. in Fuzzy Systems, 1996. Proc. Fifth IEEE Int. Conf., 3: 2071-2077.
Vasseur A, Makovicka L, Martin E, Sauget M, Contassot-Vivier S, Bahi J (2008). Dose calculations using artificial neural networks: A feasibility study for photon beams. Nuclear Instruments and Methods in Physics Research Section B: Beam Interact. Mater. Atoms., 266(7): 1085-1093.
Weaver CS (1975). Some Properties of Threshold Logic Unit Pattern Recognition Networks. IEEE Trans. Comp., 24(3): 290-298.
Webb S (2004). Contemporary IMRT: Developing Physics and Clinical Implementation. ed., Taylor & Francis.
Widesott L, Pierelli A, Fiorino C, Dell'Oca I, Broggi S, Cattaneo GM, Di Muzio N, Fazio F, Calandrino R, Schwarz M (2008). Intensity-modulated proton therapy versus helical tomotherapy in nasopharynx cancer: planning comparison and NTCP evaluation. Int. J. Radiat. Oncol. Biol. Phys., 72(2): 589-596.
Wilkens JJ, Alaly JR, Zakarian K, Thorstad WL, Deasy JO (2007). IMRT

Treatment Planning Based on Prioritizing Prescription Goals. Phys. Med. Biol., 52(6): 1675-92.
Wu HY, Hsu CY, Lee TF, Fang FM (2009). Improved SVM and ANN in incipient fault diagnosis of power transformers using clonal selection algorithms. Int. J. Innov. Comput. Inf. Control, 5(7): 1959-1974.
Yuan-Chu C, Wei-Min Q, Jie Z (2008). A new Elman Neural Network and its Dynamic Properties. Cybern. Intell. Syst. 2008 IEEE Conf., pp. 971-975.
Zhang LD, Jia L, Zhu WX (2010). Overview of Traffic Flow Hybrid ANN Forecasting Algorithm Study. Comput. Appl. Syst. Model. (ICCASM), 2010 Int. Conf., 1: 615-619.