Using Artificial Neural Network to Predict Mortality of Radical Cystectomy for Bladder Cancer

Kin-Man Lam
Department of Surgery, Tseung Kwan O Hospital
New Territories, Hong Kong, China
Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
justin.lam@connect.polyu.hk

Xue-Jian He, Kup-Sze Choi
Centre for Smart Health, School of Nursing
The Hong Kong Polytechnic University
Kowloon, Hong Kong, China
xjimhe@polyu.edu.hk, hskschoi@polyu.edu.hk

***Abstract*—Surgical removal of bladder, i.e. radical cystectomy, is a standard treatment option for muscle invasive bladder cancer. Unfortunately, the treatment is associated with significant morbidities and mortalities. Many studies have been conducted to predict the morbidities and mortalities of radical cystectomy based on statistical analysis. In this paper, an artificial neural network is employed to predict 5-year mortality of radical cystectomy. The clinico-pathological data from a urology unit of a district hospital in Hong Kong were used to train and test the model. The outcome of the surgery was computed by an artificial neural network based on the risk factors identified by a conventional statistical method. It was found that the best overall accuracy of the neural network model was 77.8% and the 5-year mortality predicted by the model was comparable to that achieved by conventional statistical methods. The results of this study reflect that artificial intelligence has great development potential in medicine.**

Keywords—Bladder cancer; Radical cystectomy; Artificial neural network; Outcome prediction; Health informatics

I. INTRODUCTION

Bladder cancer is a common cancer in genitourinary tract. In Hong Kong, it ranked as the 13th leading cause of death in 2011[1]. The crude incidence and mortality were 5.4 per 100,000 persons and 2.8 per 100,000 persons respectively. It mainly affects the elderly population. There are several histologic types of bladder cancer, namely urothelial carcinoma, adenocarcinoma, small cell carcinoma, lymphoma, other mixed cell type and cancer arising from another organ [1]. 90% of bladder cancer is transitional cell carcinoma. In this paper, bladder cancer refers to transitional cell carcinoma arising from bladder de novo. Treatment of bladder cancer, to a large extent, depends on staging. The stage of cancer is associated with the depth of cancer cell involved [2]. Higher stage usually implies poor prognosis in most cancers. Based on the American Joint Committee on Cancer, bladder cancer is staged as Tis, Ta, T1, T2, T3 and T4. Treatment of Tis and Ta bladder cancer is resection of bladder by endoscope, followed by the application of chemotherapeutic and immunotherapeutic agents into bladder. Treatment of bladder cancer at T2 stage or above, i.e. muscle invasive bladder cancer, is to remove the urinary bladder by surgical means. Surgeries for muscle invasive bladder cancer include excision of bladder, prostate in case of men, urethra, part of distal ureters and lymphatic tissue of pelvic. The whole procedure is known as radical cystectomy. Radical cystectomy can be carried out with conventional open surgery or minimal invasive approach, but minimal invasive surgery for muscle invasive bladder cancer is technical demanding and not widely adopted in many urology centres.

Despite a standard treatment option, radical cystectomy is associated with significant morbidities and mortalities. Many researchers have tried to identify risk factors in order to maximize operative outcomes. One of these outcomes is long-term survival after surgery. An urology team of the Prince of Wales Hospital in Hong Kong did a retrospective review on the 5-year survival of their patients who had undergone radical cystectomy for bladder cancers [3]. They used conventional statistical models to find out that age, tumour stage and preoperative serum albumin were significant predictors of 5-year survival. With the same data set, this study attempted to use one of the well-known artificial intelligence methods, artificial neural network (ANN), as an alternative approach to predict the 5-year survival after radical cystectomy. The risk factors identified by the urology team were used as inputs of the neural network. The output of the neural network was 5-year survival after radical cystectomy. The best overall accuracy achieved by ANN was 77.8% with 24 neurons in the hidden layer, compared to the best prediction accuracy 71.2% in the statistical model based on pre-operative serum albumin.

Prediction of outcomes of radical cystectomy involves building regression models and calculation of associated probabilities, which requires quite a lot of statistical and mathematical knowledge. Unfortunately, clinicians may not master the knowledge well. Hence, results generated from those studies may not be readily employed in clinical practice. In contrast, the output of ANN is comprehensible to clinicians and patients. Although this is a preliminary study and the sample size is small, it exemplifies a multidisciplinary effort through the collaboration of clinicians, statistician and computer scientists.

The rest of the paper is as follows. The related work is reviewed in Section II. The ANN model constructed is described in detail in Section III. In Section IV, the results of

[1] www3.ha.org.hk/cancereg/statistics.html#cancerfacts

978-1-4799-5711-8/14/$31.00 ©2014 IEEE

the neural network modelling and a comparison with the results of statistical analysis are presented. In Section V, the findings from the ANN approach in this study are discussed, together with the drawbacks and limitations. Conclusions and improvement measures are given in the last section.

II. Related Work

In the fields of urology, many articles have been published on the application of ANN to prostate cancer related issues, including the analysis of prostate cancer screening, the prediction of pathological staging and the long-term treatment outcomes.

In prostate cancer screening, Prostate Specific Antigen (PSA) is a blood test and used as a tumour marker for prostate cancers. When PSA is above a certain level, biopsy of prostate is required to confirm prostate cancers. Djavan et al. [4] developed two neural networks for early detection of prostate cancers in men with total PSA from 2.5 to 4.0 ng/mL and from 4 to 10 ng/ML respectively. The predictive accuracy of the neural networks was compared with that of conventional statistical methods. The Area Under Curves (AUC) of the Receiver Operating Characteristics (ROC) curve were respectively 87.6% and 91.3% in the two ANN models, which were significantly higher than that of conventional statistical models. Finne et al. [5] constructed an ANN by a multilayer perceptron (MLP) to predict prostate biopsy results. The data were obtained from a randomized population-based prostate cancer screening study in Finland. A goal of the study was to compare the efficacy of ANN with that of logistic regression in eliminating false-positive PSA results. It was reported that 33% and 24% of false-positive PSA results could be eliminated by MLP and logistic regression respectively. It was claimed that the neural network performed better than the statistical model. Another study of employing ANN to predict prostate biopsy results was conducted by Babaian et al. [6]. At a sensitivity of 92%, the specificity of neural network to predict prostate cancer was 62% and the specificity of free PSA was 11% only. When the neural network was employed, it effectively saved 49% of all biopsies. Remzi et al. [7] developed an ANN based on the data of 820 men from the Vienna-based multicenter European referral database. The repeat biopsy detection rate was 10%. The specificity of ANN was 68% at 95% sensitivity. It outperformed other PSA derivatives where the best was 54% in free to total PSA. It was found that ANN reduced un-necessary biopsies by 68% in their study.

Another application of ANN is the prediction of prostate cancer staging. Lymph node staging is essential in planning of treatment for prostate cancers. Imaging modality such as computerized tomography or magnetic resonance imaging may not be sensitive for small lymph node spread in prostate cancers. Batuello et al. [8] developed an ANN model to predict lymph node spread in men with clinically localized prostate cancers. The AUC of ROC curve for one of the dataset was 0.81. The positive predictive and negative predictive values were 13.6% and 98.0% respectively. The risk curve generated by the neural network gave a concrete prevalence of lymph node spread instead of a dichotomous result of presence or absence of lymph node spread. Han et al. [9] also used a neural network to predict pathological stages in prostate cancers. The results were compared with Partin's nomogram based on a logistic regression model. It was reported that the neural network outperformed nomogram in predicting pathological stages.

ANN is also used for the prediction of the outcomes of long-term prostate cancer treatment. Recurrence may happen after surgical treatment of a prostate cancer. The first sign is an increase in blood PSA level, i.e. biochemical failure. Ziada et al. [10] developed an ANN model using MLP to predict the pathological staging and the time to biochemical failure after surgical treatment of prostate cancer. For the prediction of pathological staging, the sensitivity and specificity of the neural network were 79% and 81% respectively. The overall accuracy was 80% while it was 67% when multivariate regression analysis was applied. The sensitivity and specificity of the neural network for time to biochemical failure were 67% and 85%. It was found that the overall accuracy was similar to that of obtained by multivariate analysis. Tewari et al. [11] constructed a genetic adaptive neural network to predict the biochemical failure after surgical treatment of prostate cancers. The sensitivity and specificity of the neural network were 85% and 74% respectively. The overall accuracy was 76%. Han et al. [12] reported another neural work in predicting biochemical failure. It specifically investigated the outcome using Gleason score 7 after surgical treatments. In their study, neural network, logistic regression and proportion hazard models all showed that the Gleason score 7 was significant in predicting the outcomes. The performance of neural network was slightly better than the statistical model in their study. Potter et al. [13] developed a genetically engineered neural network to predict biochemical failure after surgical treatment of prostate cancers. The study included 214 patients' pathological data and 3 networks were developed based on different combinations of the data. Apart from pathological data, molecular makers were also used as input. The accuracies of the neural network in then prediction were respectively 74.3%, 80.0% and 78.1% in the pathological data group, the molecular maker group and in all results.

On the other hand, researches on the application of ANN to bladder cancers are relatively less common. For non-muscle invasive bladder cancers, it does not require excision of the whole bladder. However, non-muscle invasive bladder cancers tend to recur and some of them may progress to more malignant forms, called *progression*. Clinicians try to look for factors to predict recurrence and progression, which include clinicopathological factors and molecular markers. Qureshi et al. [14] investigated the prediction of Ta and T1 bladder cancer recurrence within 6 months of diagnosis and 12-month cancer-specific survival in T2-T4 bladder cancers by using neural network. Instead of comparing with statistical models, the outcomes of ANN were compared to the diagnosis made by 4 consultant urologists. It was found that the prediction results of ANN and clinicians on stage progression and tumour recurrence in Ta/T1 tumours and 12-month cancer–specific survival had no significant difference, whereas the performance of neural network in predicting stage progression in T1G3 tumour was significantly better than that of clinicians. Abbod et al. [15] compared the predictive accuracies of neuro-fuzzy modelling, ANN and traditional statistical methods in bladder cancer behaviours. In all analyses, artificial intelligence was significantly better than the conventional statistical model, i.e. logistic regression. It was also found that the results of neuro-fuzzy model were significantly better than that of ANN in predicting tumour progression. Cai et al. [16] reported the use of ANN to predict recurrence-free survival of patients with high grade non-muscle invasive bladder cancer. The sensitivity and specificity of the neural network to predict recurrence-free status after endoscopic tumour resection at 12 months follow-up were 81.67% and 95.87% respectively. The overall prediction accuracy was 83.63%. ANN was also used to identify factors affecting the recurrence of bladder cancer. It was found that the number of tumour lesions, previous recurrence rate and the response to previous immunotherapy as the most influential variables affecting the recurrence risk of TaG3 bladder cancer.

III. METHODS

In this section, the proposed ANN model is first outlined, followed by the explanations of the model structure. In the last sub-section, the procedures used for determining the neuron numbers and data partitioning are presented.

A. Overview

In this study, ANN was employed to predict the outcomes of radical cystectomy for bladder cancers. The study was conducted with a reference to a previous study by Chan et al. who investigated the predictors of mortality after radical cystectomy for treatment of bladder cancers in people of Hong Kong [3]. The same clinical and pathological data were adopted here. The data of patients who had radical cystectomy for bladder cancers between May 2003 and March 2011 in the urology unit of Prince of Wales hospital in Hong Kong SAR were retrospective reviewed. The 30-day, 5-year cancer-specific, other-cause, and overall mortalities were calculated and statistically significant risk factors to predict the outcome were identified. The raw data re-analysed by the proposed ANN model and the results were then compared with that of the conventional statistical model presented by Chan et al. in [3].

B. Neural Network Model

The pattern recognition tool of MATLAB® was used to analyse the results of bladder cancer surgery. The standard network that was used for pattern recognition in the toolbox was a two-layer feed-forward network, with sigmoid transfer functions in both the hidden layer and the output layer. Fig. 1 shows the neural network structure, which includes an input layer, a hidden layer and an output layer. Table I lists the inputs and outputs of the model. The model contained 4 inputs, 1 output, and a hidden layer containing 24 neurons (to be discussed later). 117 samples were used. The two-layer feed-forward structure of this network is a typical configuration in pattern recognition. Given enough neurons in the hidden layer, the structure with a sigmoid transfer function can classify arbitrarily well [17]. In the toolbox, some configurations of the neural network (i.e. number of layers, transfer function types) are fixed, whereas the number of the hidden neurons can be defined by users. Details about the determination of the neuron number and the proportion of data used for training, validation and testing in the experiments are presented in the next sub-section.

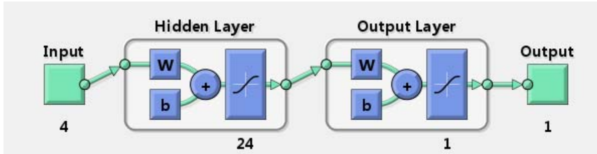

Fig. 1. *The neural network structure.*

TABLE I. INPUTS AND OUTPUT OF THE NEURAL NETWORK.

Inputs	Outputs
Age Tumour stage Albumin level Surgical approach	5 year mortality

C. Neuron Number and Data Partitioning

Several experiments with different data sizes and neuron numbers had been conducted to select the optimal neuron number in the hidden layer and the proportion of data to be used for training, validation and testing. Here, the sample data were split into three partitions, denoted by *xx-yy-zz*, where *xx*, *yy* and *zz* represent the percentage of data partitioned for training, validation and testing respectively. Three typical ways of data partitioning were used, namely, (i) 60-20-20, (ii) 70-15-15 and (iii) 80-10-10. For each setting, the network was trained for 10 times with different initial conditions and samples. For the 60-20-20 data partition (Table II), the best mean overall accuracy was 72.48% for the network with 24 neurons in the hidden layer, whilst the network with 12 neurons had a relatively stable performance with a small standard deviation. The highest overall accuracy was 77.8% in the experiment #7, where 24 neurons were used. For the 70-15-15 data partition, the best mean overall accuracy was 72.64% for the network with 24 neurons (Table III). For the 80-10-10 data partition, the network with 15 neurons had a relatively high overall accuracy of 71.98% (Table IV). Comparing the results in the three tables, the best mean overall accuracy was the network with 24 neurons and the data partition was 70-15-15. This setting was therefore used in the following experiments.

It can be seen that the performance of the neural network varied with different neuron numbers and the ways of data partitioning. Even with the neuron number and data partition were fixed, the accuracy could still be different. This can be explained by the fact that the network performance was affected by the initial values of the weights and the data samples actually picked (by random) for training, validation and testing. Therefore, the mean overall accuracy was used to determine the neuron number to be adopted. Once the neuron number and data partition were determined, the network was trained several times to search for the network setting with the best overall accuracy. Then, this trained network would be used for predicting real cases.

TABLE II. OVERALL ACCURACY AND NEURON NUMBERS WITH DATA PARTITION OF 60-20-20.

Neuron Numbers	Overall Accuracy (%)											
	Experiment										Mean	S.D.
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10		
3	69.2	68.4	68.4	67.5	59.8	70.1	63.2	73.5	71.8	62.4	67.43	4.34
6	72.6	72.6	58.1	68.4	71.8	69.2	76.1	70.1	75.2	71.8	70.59	5.00
9	62.4	70.1	74.4	72.6	68.4	68.4	75.2	73.5	73.5	64.1	70.26	4.41
12	72.6	73.5	71.8	71.8	73.5	70.9	70.9	72.6	70.9	73.5	72.20	1.09
15	73.5	69.2	66.7	72.6	72.6	60.7	69.2	70.9	70.9	74.4	70.07	4.02
18	73.5	72.6	75.2	76.9	71.8	73.5	70.9	73.5	63.2	71.8	72.29	3.64
21	66.7	59	70.1	71.8	72.6	65.8	70.1	70.1	64.1	72.6	68.29	4.37
24	70.1	68.4	73.5	75.2	73.5	70.9	77.8	76.1	65.8	73.5	72.48	3.67
27	67.5	66.7	72.6	68.4	74.4	64.1	70.9	58.1	69.2	70.9	68.28	4.65
30	72.6	76.1	72.6	74.4	68.4	70.9	74.4	70.9	70.1	70.9	72.13	2.34

TABLE III. OVERALL ACCURACY AND NEURON NUMBERS WITH DATA PARTITION OF 70-15-15.

Neuron Numbers	Overall Accuracy (%)											
	Experiment										Mean	S.D.
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10		
3	71.8	70.9	74.4	74.4	71.8	56.4	53.0	70.9	70.1	73.5	68.72	7.58
6	66.7	70.1	72.6	72.6	71.8	69.2	77.8	71.8	68.4	58.1	69.91	5.13
9	73.5	70.9	73.5	72.6	73.5	72.6	70.9	70.1	74.4	57.3	70.93	4.99
12	74.4	74.4	76.1	54.7	75.2	68.4	74.4	73.5	76.9	72.6	72.06	6.52
15	75.2	75.2	73.5	73.5	65	61.5	55.6	73.5	74.4	53	68.04	8.60
18	65.8	71.8	76.8	74.4	68.4	73.5	75.2	74.4	55.6	70.9	70.68	6.25
21	62.4	72.6	69.2	70.1	55.6	76.1	75.2	73.5	70.9	70.9	69.65	6.23
24	76.1	68.4	75.2	70.9	71.8	72.6	73.5	72.6	75.2	70.1	72.64	2.45
27	74.4	76.1	70.1	70.9	71.8	70.1	70.9	70.9	73.5	70.1	71.88	2.09
30	73.5	74.4	76.1	65	76.9	70.9	69.2	70.9	75.2	70.1	72.22	3.67

TABLE IV. OVERALL ACCURACY AND NEURON NUMBERS IN WITH DATA PARTITION OF 80-10-10.

Neuron Numbers	Overall Accuracy (%)											
	Experiment										Mean	S.D.
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10		
3	69.2	59.8	70.1	67.5	73.5	71.8	71.8	69.2	60.7	69.2	68.28	4.57
6	75.2	66.7	72.6	76.9	70.9	61.5	74.4	70.1	66.7	50.4	68.54	7.87
9	74.4	75.2	65.8	76.1	68.4	69.2	72.6	70.9	67.5	74.4	71.45	3.60
12	62.4	69.2	74.4	73.5	73.5	71.8	69.2	74.4	72.6	56.4	69.74	5.93
15	75.2	76.1	70.1	77.8	66.7	75.2	66.7	73.5	70.1	68.4	71.98	4.08
18	69.2	68.4	66.7	70.1	73.5	67.5	65.8	72.6	70.9	70.9	69.56	2.51
21	71.8	60.7	53	71.8	72.6	71.8	52.1	68.4	74.4	74.4	67.1	8.63
24	72.6	70.9	70.9	60.7	68.4	76.9	73.5	69.2	73.5	73.5	71.01	4.37
27	65.8	69.2	69.2	64.1	72.6	66.7	70.9	59	77.8	72.6	68.79	5.23
30	70.1	76.1	64.1	73.5	72.6	71.8	65	76.9	72.6	67.5	71.02	4.34

IV. EXPERIMENTAL RESULTS

In this section, we firstly present the performance of the proposed ANN model. Then, we compare the results with the statistical analyses.

A. Neural Network Performance

From statistical analysis, age, tumour stage and preoperative serum albumin and surgical approach were identified as significant factors in predicting 5-year survival after radical cystectomy. They were employed as inputs of the neural network for the prediction of 5-year survival rate. The structure of the neural work and the partitioning of data were discussed previously. All input data were divided into training, validation and testing data sets. The allocation of data was carried out in random manner such that 70% was used for training, 15% was used to validate the network and to stop training before over fitting and the last 15% was used as a completely independent test of network generalization. Table V shows the data size of the three data sets. The performance of the neural network model is given in terms of the mean squared error (MSE) and the percentage error (%E), which are also given in the table. The calculation was stopped after 30 epochs as shown in Fig. 2. The best validation performance was at epoch 24. The computation time was relative short, mainly attributed to the small sample size.

TABLE V. DATA SIZES AND PERFORMANCE (NEURON NUMBER: 24).

	Samples	MSE	%E
Training (70%)	83	0.163	21.7
Validation (15%)	17	0.133	23.5
Testing (15%)	17	0.213	23.5

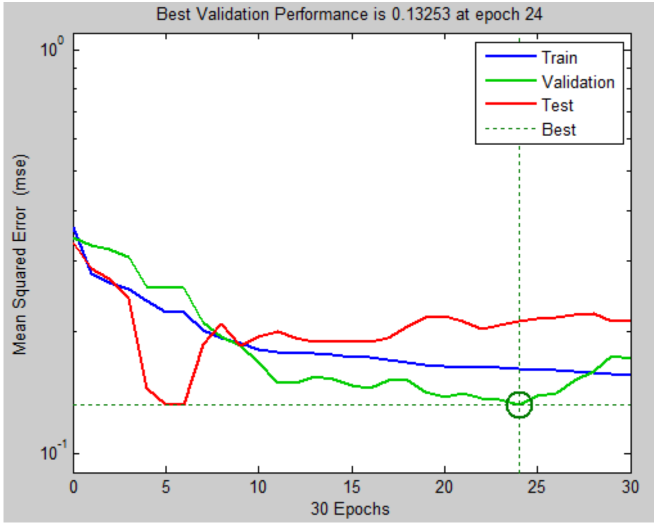

Fig. 2. *Neural network performance.*

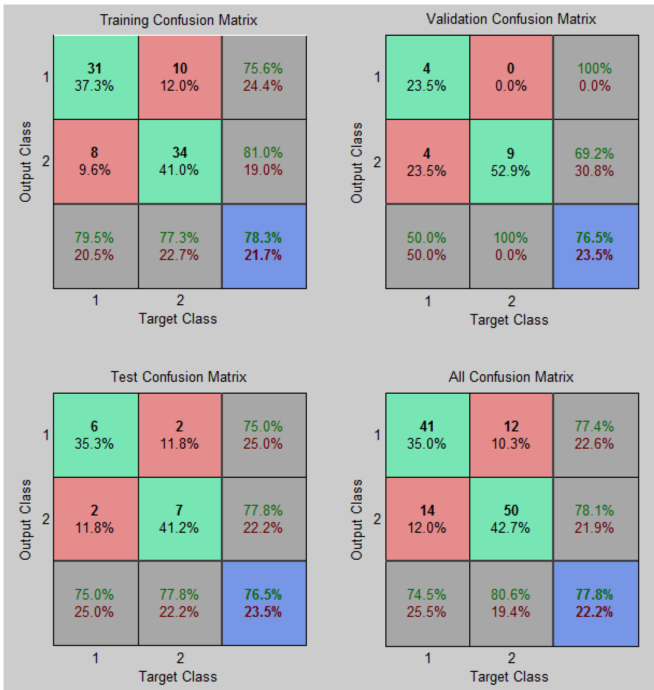

Fig. 3. *Confusion matrix.*

More detailed analysis on the accuracy of the neural network model was presented in the confusion matrix as shown in Fig. 3. High accuracy was indicated by the numbers of correct responses in the green boxes (true positives or true negatives) and the low numbers of incorrect responses in the red boxes (false positives or false negatives). The lower right blue boxes give the overall accuracy of 77.8% in the experiment. The accuracy in the training dataset was 78.3% and it was a little bit larger than that of the overall accuracy. In model testing, there were 2 false-positive cases and 2 false-negative cases among the 17 samples. In model validation, there were 4 false-negative cases among the 17 samples.

B. Comparison with Statistical Analysis

The quality of the classifier was evaluated by ROC as shown in Fig. 4. The ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (specificity) as the threshold varies. A perfect result is indicated by having the points located at the upper-left corner, meaning 100% sensitivity and 100% specificity. The AUC of "All ROC" of the neural network is 0.829 (lower right of Fig. 4), which is calculated by taking all the four ANN inputs into account, i.e., age, tumour stage, surgical approach and serum albumin. As a reference, the ROC curve obtained based on statistical method for serum albumin only is provided in Fig. 5, and the corresponding AUC is 0.712.

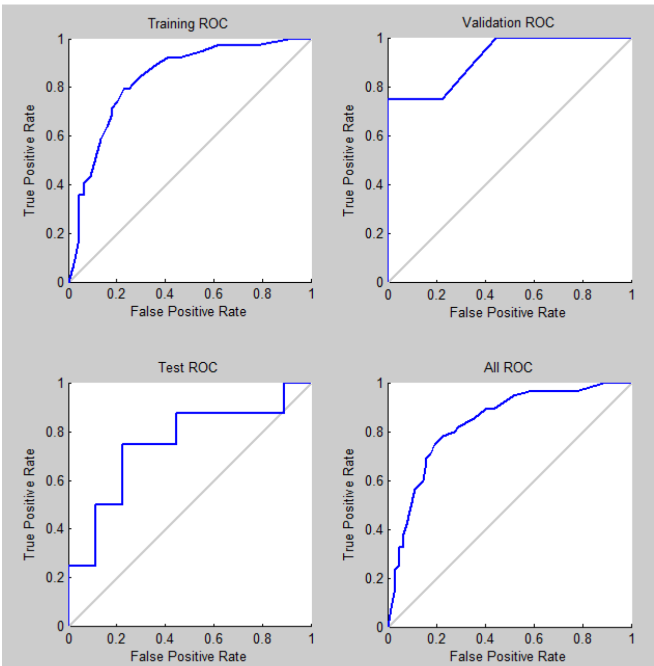

Fig. 4. *ROC analysis of the neural network.*

V. Discussions

The drawback of this study is the small sample size and the relatively short follow-up time of the cases. The small sample size may result in type II errors, while short follow-up time may introduce information bias. The small sample size also leads to inadequacy of data for the calculation in the neural network model. It can be seen from Fig. 4 that the curves in the validation and testing (with lesser samples allocated in data partitioning) are coarser than that in the training. This could be explained by the issue that the numbers of samples in validation and testing were relatively small, with only 17 cases in both processes.

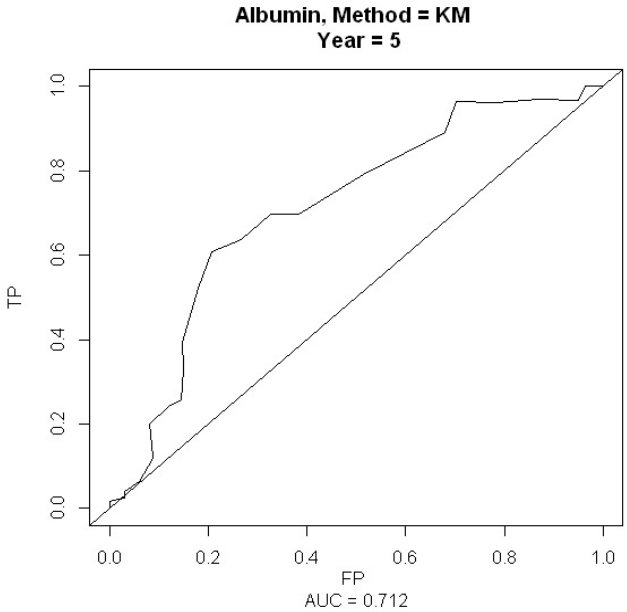

Fig. 5. *ROC analysis of serum albumin.*

The "black box" paradigm is another concern in using ANN in data analysis and interpretation. Unlike statistical methods that have clear assumption and calculation relying on statistical distribution, ANN methods consists of adjustment of the weighted information and activation of nodes by mathematical formulae that are not well known to end users. In other words, the way neural network generating outputs is difficult to comprehend for non-specialists.

Since the sample size in the present study is rather small (117 cases), heuristic identification of risk factors as input of the ANN may not be a suitable approach. It is therefore resorted to the adoption of the findings obtained previously from statistical analysis on the same data set [3], which suggest that age, tumour stage, surgical approach and serum albumin age are the risk factors. These four parameters are then employed as the inputs of the ANN. Indeed, other parameters such as gender, length of stay, tumour grade may also be related to the mortality rate. When more data become available in the future, the risk factors and thus the inputs of ANN can be identified heuristically.

ANN may be also subject to the "over-fit" problem. When too many samples (which may involve noisy data) are used to train the model in supervised learning, the resulting model may describe the noise instead of the real underlying relationships. In addition, while the error in training set is usually reduced to a very small value, when new data is presented to the network the error may be large. An over-fit model will have poor predictive performance. There are several ways to overcome over-fit and improve the generalizability of the network. The size of the network may be reduced so that the network has not enough power to over-fit the data. However, it is difficult to know exactly the required size of a network for a specific application. Another way to overcome over-fitting is to re-train neural networks. A large dataset can be used to train several networks in order to result in a network with good generalization ability. When a dataset is relatively small with noisy data, it can be trained with multiple neural networks and the outputs are averaged.

Improvement of this study necessitates a large sample size and long follow-up period. Nevertheless, the incidence of bladder cancers is static. There is actually a decreasing trend in recent years. Caseload in individual hospital is limited so a multi-centre trial would be required to increase the sample size. To facilitate data collection for clinical research, a cancer database is needed for different centres to collect and aggregate information for research purpose. A standardized treatment and follow-up protocols are also required.

In this study, the architectural design of the neural network is based on commercial software. Toolboxes of standard settings are provided to facilitate non-specialists to set up and make use of neural networks. For example, the popular sigmoid function is built-in in some of the toolboxes. Further research may be conducted to explore other possible transfer functions. In addition, the network structure, the number of hidden nodes and the proportion of data for training, testing and validation sets may be further evaluated to increase the accuracy of the model.

Despite the potential of ANN models in medical applications, non-specialist medical researchers and clinicians could hardly understand the mechanism associated with the nodes, hidden layers, weights and transfer functions of neural networks. Popularity and implementation of neural-network-based clinical applications thus remain low and slow. Another issue is that medical researchers and clinicians are often not familiar with computer programming. Commercial neural network tools usually require certain knowledge of programming, which is a barrier for them to adopt this technology. Therefore, neural network tools shall be developed so that they are friendly to be used by clinicians and non-specialist health care professionals.

VI. CONCLUSIONS

In this study, an ANN model was constructed to predict 5-year mortality of cystectomy. A typical two-layer feed-forward ANN model was adopted, with 4 inputs in the input layer, 1 hidden layer and 1 output layer. Age, tumour stage, albumin level and surgical approach were the 4 inputs of the model. The hidden layer had 24 neurons, and sigmoid transfer functions were used in both the hidden layer and the output layer. 117 samples were used in the experiments. The data were partitioned such that 70% of them were used for training, 15% for validation and testing respectively. The best overall accuracy achieved by this artificial neural network model was 77.8%, with the AUC equal to 0.829. The prediction of neural network is better than that obtained by regression model – the AUC obtained with conventional statistical method is 0.712 only [3]. The prediction even appears to be more accurate than the studies employing the Surveillance, Epidemiology, and End Results (SEER) data [18]. This result is encouraging and it reflects the potential of artificial intelligence to augment conventional statistical methods in the prediction of bladder cancer mortality after operation.

Interpretation of clinical data is traditionally based on statistics. This study has shown that ANN can be used to predict the survival of bladder cancer patients after radical cystectomy. The prognostic accuracy of the ANN model is not only comparable and likely to be better than conventional statistical models. The validity of this model may be challenged by the small sample size, short follow-up period, black box features and possible over-fitting of neural networks. This study can be revisited with more cases accumulated.

It is planned to line up with other cluster hospitals to establish a database in bladder cancer research. With more samples, a more reliable ANN model may be constructed, which could be used to identify the risk factors and to predict outcomes. With a reliable and convincing model, individual hospitals can establish their own neural network to identify locally the risk factors and to perform the prediction. The model can be presented as user-friendly software for use by urologists during patient counselling to provide second opinions and help preventing bias.

ACKNOWLEDGMENT

This work was supported in part by the Research Grants Council of Hong Kong SAR (Project No. PolyU 5134/12E). The authors would like to thank Dr. Eddie Chan, consultant urologist of the Prince of Wales Hospital, for sharing the data and giving advice on the study.

REFERENCES

1. Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA (2007) Campbell-Walsh Urology e-dition. JAMA 298 (18):2201-2202

2. Rubin E, Reisner HM (2009) Essentials of Rubin's pathology. Lippincott Williams & Wilkins.

3. Chan ES, Yip SK, Hou S, Cheung H, Lee W, Ng C (2013) Age, tumour stage, and preoperative serum albumin level are independent predictors of mortality after radical cystectomy for treatment of bladder cancer in Hong Kong Chinese. Hong Kong medical journal 19 (5):400-406

4. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M (2002) Novel artificial neural network for early detection of prostate cancer. Journal of Clinical Oncology 20 (4):921-929

5. Finne P, Finne R, Auvinen A, Juusela H, Aro J, Määttänen L, Hakama M, Rannikko S, Tammela TL, Stenman U-H (2000) Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology 56 (3):418-422

6. Babaian RJ, Fritsche H, Ayala A, Bhadkamkar V, Johnston DA, Naccarato W, Zhang Z (2000) Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. Urology 56 (6):1000-1006

7. Remzi M, Anagnostou T, Ravery V, Zlotta A, Stephan C, Marberger M, Djavan B (2003) An artificial neural network to predict the outcome of repeat prostate biopsies. Urology 62 (3):456-460

8. Batuello JT, Gamito EJ, Crawford ED, Han M, Partin AW, McLeod DG, O'Donnell C (2001) Artificial neural network model for the assessment of lymph node spread in patients with clinically localized prostate cancer. Urology 57 (3):481-485

9. Han M, Snow PB, Brandt JM, Partin AW (2001) Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma. Cancer 91 (S8):1661-1666

10. Ziada AM, Lisle TC, Snow PB, Levine RF, Miller G, Crawford ED (2001) Impact of different variables on the outcome of patients with clinically confined prostate carcinoma: prediction of pathologic stage and biochemical failure using an artificial neural network. Cancer 91 (8 Suppl):1653-1660

11. Tewari A, Issa M, El-Galley R, Stricker H, Peabody J, Pow-Sang J, Shukla A, Wajsman Z, Rubin M, Wei J (2001) Genetic adaptive neural network to predict biochemical failure after radical prostatectomy: a multi-institutional study. Molecular urology 5 (4):163-169

12. Han M, Snow PB, Epstein JI, Chan TY, Jones KA, Walsh PC, Partin AW (2000) A neural network predicts progression for men with Gleason score 3+ 4 versus 4+ 3 tumors after radical prostatectomy. Urology 56 (6):994-999

13. Potter SR, Miller MC, Mangold LA, Jones KA, Epstein JI, Veltri RW, Partin AW (1999) Genetically engineered neural networks for predicting prostate cancer progression after radical prostatectomy. Urology 54 (5):791-795

14. QURESHI KN, NAGUIB RN, HAMDY FC, NEAL DE, MELLON JK (2000) Neural network analysis of clinicopathological and molecular markers in bladder cancer. The Journal of urology 163 (2):630-633

15. Abbod MF, Catto JW, Chen M-Y, Linkens DA, Hamdy FC Artificial intelligence techniques for the prediction of bladder cancer progression. In: ESANN, 2005. pp 109-114

16. Cai T, Conti G, Nesi G, Lorenzini M, Mondaini N, Bartoletti R (2007) Artificial intelligence for predicting recurrence-free probability of non-invasive high-grade urothelial bladder cell carcinoma. Oncology reports 18 (4):959-964

17. Jiang X, Harvey Kam Siew Wah A (2003) Constructing and training feed-forward neural networks for pattern classification. Pattern recognition 36 (4):853-867

18. Isbarn H, Jeldres C, Zini L, Perrotte P, Baillargeon-Gagne S, Capitanio U, Shariat SF, Arjane P, Saad F, McCormack M (2009) A population based assessment of perioperative mortality after cystectomy for bladder cancer. The Journal of urology 182 (1):70-77