Application of artificial neural networks to establish a predictive mortality risk model in children admitted to a paediatric intensive care unit

Chan C H, Chan E Y, Ng D K, Chow P Y, Kwok K L

ABSTRACT

Introduction: Paediatric risk of mortality and paediatric index of mortality (PIM) are the commonly-used mortality prediction models (MPM) in children admitted to paediatric intensive care unit (PICU). The current study was undertaken to develop a better MPM using artificial neural network, a domain of artificial intelligence.

Methods: The purpose of this retrospective case series was to compare an artificial neural network (ANN) model and PIM with the observed mortality in a cohort of patients admitted to a five-bed PICU in a Hong Kong non-teaching general hospital. The patients were under the age of 17 years and admitted to our PICU from April 2001 to December 2004. Data were collected from each patient admitted to our PICU. All data were randomly allocated to either the training or validation set. The data from the training set were used to construct a series of ANN models. The data from the validation set were used to validate the ANN and PIM models. The accuracy of ANN models and PIM was assessed by area under the receiver operator characteristics (ROC) curve and calibration.

Results: All data were randomly allocated to either the training (n=274) or validation set (n=273). Three ANN models were developed using the data from the training set, namely ANN8 (trained with variables required for PIM), ANN9 (trained with variables required for PIM and pre-ICU intubation) and ANN23 (trained with variables required for ANN9 and 14 principal ICU diagnoses). Three ANN models and PIM were used to predict mortality in the validation set. We found that PIM and ANN9 had a high ROC curve (PIM: 0.808, 95 percent confidence interval 0.552 to 1.000, ANN9: 0.957, 95 percent confidence interval 0.915 to 1.000), whereas ANN8 and ANN23 gave a suboptimal area under the ROC curve. ANN8 required only five variables for the calculation of risk, compared with eight for PIM.

Conclusion: The current study demonstrated the process of predictive mortality risk model development using ANN. Further multicentre studies are required to produce a representative ANN-based mortality prediction model for use in different PICUs.

Keywords: artificial neural network, intensive care unit, mortality prediction models, paediatric intensive care unit, predictive mortality risk model

Singapore Med J 2006; 47(11):928-934

Department of Paediatrics
Kwong Wah Hospital
Waterloo Road
Hong Kong SAR
China

Chan C H, BSc
Research Assistant

Chan E Y, MRCPCH
Medical Officer

Ng D K, MD
Consultant

Chow P Y, FHKAM
Specialist Medical Officer

Kwok K L, FHKAM
Senior Medical Officer

Correspondence to:
Dr Daniel K Ng
Tel: (852) 3517 5055
Fax: (852) 3517 5261
Email: dkkng@ha.org.hk

INTRODUCTION

Mortality prediction models (MPMs) are used to describe severity of illness and probability of death. Paediatric risk of mortality (PRISM) and paediatric index of mortality (PIM) are the commonly-used MPMs in children admitted to the paediatric intensive care unit (PICU). The latest version of PRISM, PRISM III[1], was validated in PICUs in different countries[1-4]. However, PRISM III has been criticised for "laborious collection of information" and "fee required for the usage of its regression equations[3]". An alternative model, PIM[5], addressed these issues as the regression model is freely available and the number of required parameters are fewer than that required for PRISM. The performance of PIM was satisfactory in PICU cohorts in developed countries[2,3,6,7]. However, the performance of PIM in predicting mortality was reported to be poor in PICU, with significantly different settings from the derivative cohort, i.e. PICUs in developing countries[8].

PIM was developed using multivariate logistic regression technique. Artificial neural networks (ANN), a domain of artificial intelligence, are flexible nonlinear systems which were used previously to model the outcomes when the relationships between

the variables were complex, multidimensional and nonlinear as found in complex biological systems[9]. ANNs recognise the pattern in a set of data or training set using certain learning algorithm and they emerged as an integral part of paediatric clinical decision support system[10]. Use of ANN was reported in predicting mortality in ICU in adults[11-16] and preterm neonates[17]. To date, only one paediatric study to use ANN to predict mortality and length of stay in PICU has been published[18].

The 14 variables based-model was developed based on 8,081 patients records, and it could predict mortality in PICU with 85.7% accuracy. However, that model has not been validated externally with data other than the training set. ANN models are well-known for "overfitting" the training data set[19]. In order to guarantee the generalisability of ANN usage, the models derived from ANN should be externally validated using data other than the training data[19]. In this pilot study, we developed a series of ANN models trained with data set from a PICU in Hong Kong using a learning algorithm called "Group method of data handling" (GMDH)[20]. The objective of this study is to evaluate the applicability of ANN to model mortality in PICU. The potential of ANN in future MPM development was also studied.

METHODS

This study was conducted in a five-bed PICU in a Hong Kong non-teaching general hospital. Although this unit received referral from neonates to young adults of up to 18 years old from general paediatric wards, accident and emergency units, surgical wards and operating rooms, neonates were excluded from the current study to allow for a less heterogeneous population. This unit was staffed by one medical officer, one senior medical officer, one consultant and three registered nurses with 24 hours per day coverage.

Data were collected on all patients admitted to our PICU between April 2001 and December 2004. Age, gender, the principal diagnosis, pre-ICU or urgent intubation, length of stay, all eight items of PIM scores and PICU survival were tabulated (Table I). The quality requirement of PIM[5] was used in data collection. Data were collected by the attending medical officer on a standard data sheet and stored in a computer database. The current study only entailed re-analysis of existing routine surveillance dataset and it did not require any patients' contact or any extra intervention. Hence, informed consent was not obtained. Ethic approval by institutional review board was not regarded as necessary.

PIM-based mortality risk was computed by the regression equation provided by Shann et al[5]. An updated score, PIM2[21], has been available since 2003 but it was not used as the current study commenced at a much earlier date (April 2001). Some specific information for PIM2 was not collected, e.g. the details of underlying conditions. PIM was derived from patients aged less than or equal to 16 years, and patients older than 16 years were excluded from this study. From April 2001 to December 2004, 555 patients were admitted to our PICU. Eight of them were excluded because they were older than 16 years. The remaining 547 patients were randomly allocated to either the training or validation set on a 50-50 basis. The randomisation was done by "random sample of cases" function in Statistical Package for Social Science (SPSS) version 11.0 for Macintosh (SPSS, Chicago, IL, USA).

Table I. Data collected from each patient admitted to our unit.

Gender/Age	M/F
Principal diagnosis for PICU admission	
PIM: Elective admission	Yes/No
Information mandated by PIM	Yes/No
Tracheal intubation (Pre-ICU or urgent at anytime during first hour in ICU)	Yes/No
PICU survival	Death during PICU admission/survival upon PICU discharge

The data from the training set were transferred to KnowledgeMiner X 5.0 (Script Software, www.knowledgeminer.com) on an Apple G4 Macintosh computer (Apple Computer Inc, Cupertino, CA, USA) for the development of ANN-based MPM. The ANN architecture used was GMDH neural network[20]. The technical details of training mechanism and variables elimination mechanism are provided in Appendix 1. The formula of neural network model could be derived from this software and it was different from previous studies where the formulae of neural network were not disclosed[9,12-18].

Three neural network models were developed. The first model, ANN8, was trained with eight variables required for PIM scores. The second model, ANN9, was trained with eight variables required for PIM and one additional variable (Pre-ICU or urgent intubation). The third model, ANN23 was trained with nine variables of ANN9 and 14 principal ICU diagnosis variables. The additional variables to ANN9 and ANN23 were chosen by one of the authors (DKN) based on personal experience and previous report[22]. The cut-off point selections for each model

(PIM, ANN8, ANN9 and ANN23) were based on the following predefined principle: sensitivity equal to 100% with the highest specificity. The mortality of patients in the validation set was predicted by PIM, ANN8, ANN9 and ANN23.

Descriptive statistics (Median with interquartile range for continuous variables and percentage for categorical variables) were used to report age, gender, principal ICU diagnoses, length of stay in PICU, and mortality. The above data were compared between the training and validation sets by either chi-square test with continuity correction or Mann-Whitney test. The discriminatory power, i.e. how well the mortality and survival in the data set were separated, of each model was assessed by the area under the receiver-operator characteristic (ROC) curve. Calibration of models, i.e. how close the model probability is to the true probability, across different risk groups was evaluated using the method similar to Ozer et al[8]. Standardised mortality ratio (SMR) was calculated for each model to compare the observed and expected PICU death in each risk group. Statistical analysis was carried out with SPSS 11.0 for Macintosh.

RESULTS

The descriptive statistics of patients in the training and validation sets are presented in Table II, and no significant difference was found between them. Three ANN models were successfully developed using 274 cases from training set. Number of variables required for ANN8, ANN9 and ANN23 were four, five and nine, respectively (Table III). The results of internal validation of PIM, ANN8, ANN9 and ANN23 are presented in Table IV. The ROC curves of four models are presented in Fig. 1. ANN8, ANN9 and ANN23 had a larger area under the curve than PIM although the 95% CI of area under the curve of four models crossed. All four models were then used to predict the mortality in the validation set. The ROC curves of four models are presented in Fig. 2 and Table V. ANN9 had the highest discriminatory power, even though the 95% CI of area under the curve crossed that of PIM. Both ANN8 and ANN23 give a suboptimal discriminatory power.

The calibration of the three models was evaluated (Table VI). SMR could not be computed in certain risk groups because there was no observed death. Goodness-of-fit test was not done because of insufficient number of observed death (<5) in all risk groups. ANN models had a similar but apparently poorer calibration than PIM, since the mortality of patients in the very low risk group was largely underestimated. However, three models also significantly overestimated the mortality in very high risk groups as the SMRs were significantly lower than one. The formula of ANN9 is listed in Fig. 3.

Table II. Comparison between training and validation sets.

Variables	Training set n=274	Validation set n=273	p-value
Median age (IQR)	1 year (5 months to 6 years)	1 year (6 months to 7 years)	0.634
Male (%)	175 (63.9%)	161 (59.0%)	0.277
Female (%)	99 (37.1%)	112 (41.0%)	
PICU death (%)	7 (2.6%)	5 (1.8%)	0.775
Urgent tracheal intubation (%)	43 (15.7%)	31 (11.4%)	0.174
Median length of stay (IQR)	3 days (1 day to 6 days)	3 days (1day to 6 days)	0.298
Principal PICU diagnosis (%)			
Neurological	51 (18.6%)	52 (19.0%)	0.524
Cardiovascular	10 (3.6%)	15 (5.5%)	
Respiratory	109 (39.8%)	106 (38.8%)	
GI/Liver	9 (0.3%)	13 (4.8%)	
Renal	7 (2.6%)	7 (2.6%)	
Postoperative	50 (18.2%)	44 (16.1%)	
Overdose	3 (1.1%)	2 (0.7%)	
Trauma	0 (0%)	3 (1.1%)	
Metabolic	7 (2.6%)	9 (3.2%)	
Sepsis	11 (4.0%)	3 (1.1%)	
Scald	2 (0.7%)	4 (1.5%)	
Haematological	2 (0.7%)	3 (1.1%)	
MOF	1 (0.4%)	0 (0.0%)	
Miscellaneous	12 (4.4%)	12 (4.4%)	

IQR: Interquartile range; GI: Gastrointestinal; MOF: Multi-organ failure

DISCUSSION

The discriminatory power of ANN9 was higher than PIM, although the difference was not significant. Nonetheless, ANN9 provided additional benefit to PIM because the number of variables required for computation of risk was fewer. ANN23 and ANN8 gave impressive discriminatory power in the training set but not in the validation set, indicating that they were misfitted models and the generalisability of these models were limited. The comparison of calibrations in the four models was inconclusive because of insufficient death cases in our cohort.

The application of ANN models in medical problems was evaluated extensively in the literature[23]. ANN models served as an attractive alternative to the

Table III. Variables required for each model.

Model	Relevant variables
ANN8	Elective admission, underlying condition, mechanical ventilation, pupils response
ANN9	Elective admission, underlying condition, mechanical ventilation, urgent tracheal intubation, pupils response
ANN23	Respiratory diseases, cardiovascular diseases, neurological diseases, urgent tracheal intubation, pupils response, FiO_2, elective admission, underlying condition, mechanical ventilation
PIM	FiO_2, PaO_2, base excess, systolic BP, elective admission, underlying condition, mechanical ventilation, pupils response

Table IV. Area under ROC curve in predicting mortality in training set, cut-off probability and specificity of four models.

Model	Area under ROC curve (95% CI)	Cut-off probability	Specificity
PIM	0.872 (0.748 to 0.996)	1.26%	50.2%
ANN8	0.909 (0.756 to 1.000)	0.11%	8.7%
ANN9	0.956 (0.903 to 1.000)	0.46%	81.6%
ANN23	0.994 (0.982 to 1.000)	9.75%	95.9%

CI: Confidence interval

Table V. Area under ROC curve, sensitivity and specificity in predicting mortality in validation set by four models.

Model	Area under ROC curve (95% CI)	Sensitivity	Specificity
PIM	0.808 (0.552 to 1.000)	80.0%	55.2%
ANN8	0.683 (0.391 to 0.975)	100.0%	0.7%
ANN9	0.957 (0.915 to 1.000)	100.0%	85.4%
ANN23	0.393 (0.009 to 0.795)	40.0%	94.4%

CI: Confidence interval

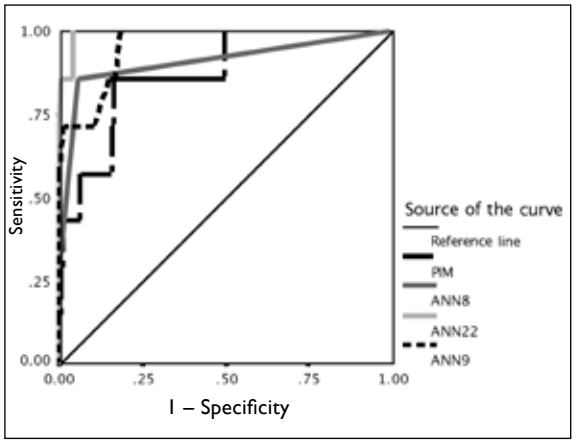


Fig. 1 ROC curves of four models in predicting mortality in the training set.

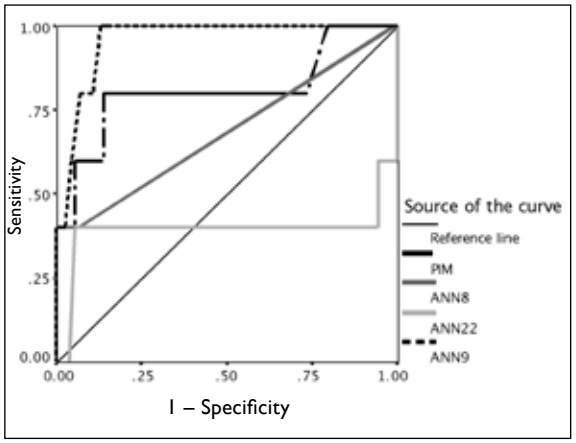


Fig. 2 ROC curves of four models in predicting mortality in the validation set.

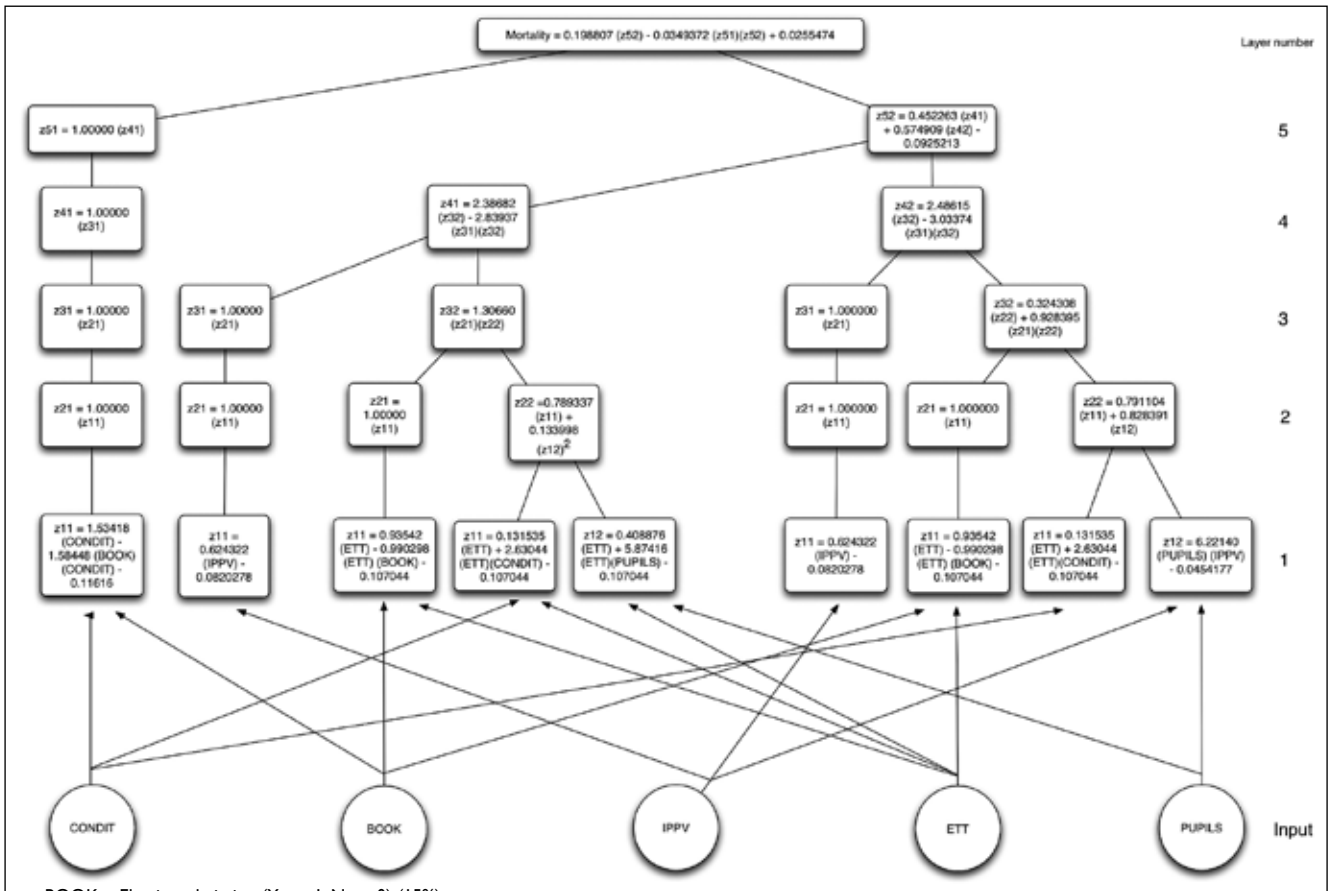
conventional logistic regression models because of zero human intervention in training, theoretical higher accuracy, and availability of software for development[9]. Dreiseitl et al[19] surveyed 72 studies comparing the medical prognostic accuracy of logistic regression and ANN classification models. 51% of them concluded that ANN models had a higher discriminatory power than logistic regression models, while 42% of them showed no difference. The remaining 7% favoured logistic regression models. Dreiseitl et al concluded that the higher percentage of studies favouring the ANN models may be explained by publication bias in favour of the new learning methods, incorrect model building and validation mechanism, and incorrect statistical analysis to compare the performance of models. Dreiseitl et al also recommended that future studies of this kind should provide details of modelling, estimate of the generalisation error and compare the calibration and discriminatory power of models.

A number of studies evaluated the accuracy of ANN models and logistic regression models for prediction of mortality in adult ICU[11-16]. Most of them compared the ANN models with logistic models derived from the study cohort[12-14]. Thus, the logistic regression equations developed by these authors were trained on relatively-small data sets and were not validated elsewhere. Three adult studies[11,15,16] compared ANN models with a currently accepted and already validated MPM called APACHE II. All studies[11,15,16] showed that ANN models required few number of variables in calculation of risk than APACHE II. The current study is the first study compared the accuracy of ANN models with a currently accepted and already validated MPM used in PIM.

Table VI. Calibration of three models in predicting the mortality in the validation set.

Model		Probability of mortality				
		≤1%	1-5%	5-15%	15-30%	>30%
PIM	Number of patients	107	121	31	5	9
	Mean risk	0.0055	0.0176	0.0863	0.2277	0.6579
	Observed death	1	0	2	0	2
	Expected death	0.59	2.13	2.68	1.14	5.92
	SMR	1.69 (0.00 to 5.60)	NA	0.74 (0.00 to 1.77)	NA	0.34 (0.00 to 0.81)
ANN9	Number of patients	264	0	0	4	5
	Mean risk	0.0012	-	-	0.300	0.984
	Observed death	3	-	-	0	2
	Expected death	0.3168	-	-	1.2	4.92
	SMR	9.47 (0.00 to 20.55)	NA	NA	NA	0.41 (0.00 to 0.97)
ANN23	Number of patients	247	0	16	1	9
	Mean Risk	0.0036	-	0.0988	0.28	0.791
	Observed death	3	-	2	0	0
	Expected death	0.8892	-	1.5808	0.28	7.119
	SMR	3.37 (0.00 to 7.18)	NA	1.27 (0.00 to 3.03)	NA	NA

* SMR: Standardised mortality ratio; NA: Not available



BOOK = Elective admission (Yes = 1, No = 0) (15%)
CONDIT = Underlying condition (Yes = 1, No = 0) (18%)
IPPV = Mechanical ventilation (Yes = 1, No = 0) (16%)
ETT = Urgent tracheal intubation (Yes = 1, No = 0) (25%)
PUPILS = Pupils response (26%)

The percentage in parentheses reflect the contribution of an input to reducing the model error.

Fig. 3 Formula of ANN9.

Our study was similar to adult studies done in the UK and Canada[11,15] that found no significant difference in discriminatory power between ANN models and validated MPM. However, the current study was different from an Indian study by Nimgaonkar et al[16] who found a significantly higher discriminatory power in ANN compared with APACHE II. Nimgaonkar et al commented that the significantly superior discriminatory power in ANN models in the Indian cohort might be explained by the fact that the ANN models in Nimgaonkar et al's study were developed using the data from an Indian ICU while other ANN models[11,15] were developed in ICUs in developed countries. The casemix, admission criteria, standards of care and availability of human and material resource in Indian ICU were different from ICUs in developed countries where the APACHE II model was developed.

Walczak and Scorpio[18] reported the only study evaluating the ANN models in predicting length of stay and acuity of care in PICU using the data available in the first ten minutes of admission. ANN derived by Walczak et al could accurately predict 85.7% of mortality in a PICU. However, the high accuracy of ANN model reported by Walczak et al may not be generalised, as the figures reported was derived from the training set only. No external validation using data other than the training set was attempted by Walczak et al. The ANN23 model in the current study served as an excellent example to illustrate the importance of external validation. ANN23 model derived in the current study provided the highest discriminatory power among four models. However, the discriminatory power was significantly dropped in the validation set. This drop in discriminatory power can be explained by misfitting. The higher complexity of ANN models compared to logistic regression made ANN models more susceptible to overfitting. An overfitted model was not useful in predicting the outcome in data other than the training set. The problem of overfitting cannot be diagnosed by internal validation only[19].

Both PIM and ANN9 gave a similar discriminatory power in both the training set and validation set, indicating the low possibility of overfitting. We evaluated three ANN models because model selection is a very important strategy to prevent misfitting in the ANN model. The addition of the variable "Pre-ICU or urgent intubation" in ANN9 on top of the variables required for PIM was based on personal experience of one of the authors (DKN) and the study by Earle et al[22], who found that tracheal intubation was associated with an increased mortality rate of patients in PICU.

The most important finding in the current study is that ANN9 could provide a similar discriminatory power to PIM by using less data, i.e. only five binary variables, compared with eight variables required for PIM. Thus, the ANN9 model required lesser effort in data collection than PIM with similar accuracy. It is our experience that the collection of the four continuous variables in PIM is cumbersome and gives rise to inaccuracy. The main disadvantage was the complex calculation of risk by ANN9 that required a programmable calculator or computer. The authors had written a Python script for calculation of risk score using ANN9 which can be executed in either desktop computer or personal data assistant.

One of the main problems of the current study was the limited sample size which hindered the measurement of calibrations in all models. Our PICU was a low risk unit with only 12 mortality cases within four years. Upon randomisation to either the training or validation set, mortality cases in each set further reduced to seven and five, respectively. This is less than the number of ten death cases for each group to provide a more reliable ROC analysis[24]. Moreover, the study by Clermont et al[14] assessed the robustness of ANN-based MPM modelling using small sample size and suggested that a training set of less than 800 cases was generally inadequate. However, the same study also demonstrated that decreased sample size from 1,200 to 200 cases in a training set did not significantly change the misclassification rates of ANN-derived MPM.

Another problem with our study was the long period of data collection, i.e. four years, and the annual advances in our PICU may have decalibrated all models in this study. Similar to another study[16], we randomly stratified all cases to training and validation sets in order to minimise the influence of technological advancement and increased resources. We conducted our study at a single institution, meaning that the external validity of ANN9 is not proven. One of the important features of the ANN model is the retraining potential[17]. Retraining can be done in case of advances in medical care, or simply in case of additional data becoming available. Theoretically, the ANN model can achieve a higher discriminatory and calibration power by repeated retraining with new data.

This pilot study can be easily extended by incorporating future data from continuous surveillance of our PICU. However, it is advisable to develop a MPM based on the data from different centres because a multicentre study allows for collection of a large amount of data in a short period of time. This pilot study showed that ANN could be regarded as

an alternative to conventional logistic regression for modelling PICU mortality. Due to inadequacies in the current study (Table VI), it is premature to draw any conclusions from the model's performance. However, ANN is likely to produce a MPM with a similar discriminatory power to those developed by logistic regression (PIM) but require fewer number of variables. In order to produce a representative ANN-based MPM, we recommended MPM developers to incorporate the ICUs from different regions around the world[8,16]. In conclusion, our study demonstrated the process of MPM development using ANN. Further multicentre studies are required to produce a representative ANN-based mortality prediction model for use in different PICUs.

REFERENCES

1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. Crit Care Med 1996; 24:743-52.
2. Slater A, Shann F. ANZICS Paediatric Study Group. The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. Pediatr Crit Care Med 2004; 5:447-54.
3. Gemke RJ, van Vught J. Scoring systems in pediatric intensive care: PRISM III versus PIM. Intensive Care Med 2002; 28:204-7.Comment in: Intensive Care Med 2002; 28:105-7.
4. Wang JN, Wu JM, Chen YJ. Validity of the updated pediatric risk of mortality score (PRISM III) in predicting the probability of mortality in a pediatric intensive care unit. Acta Paediatr Taiwan 2001; 42:333-7.
5. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. Intensive Care Med 1997; 23:201-7. Comment in: Intensive Care Med 1997; 23:141-2, Intensive Care Med 2000; 26:145.
6. Fraser J, Maskrey C, Taylor H. Evaluation of the Paediatric Index of Mortality in children managed on adult intensive care units. Arch Dis Child 2004; 89:974-6.
7. Tibby SM, Taylor D, Festa M, et al. A comparison of three scoring systems for mortality risk among retrieved intensive care patients. Arch Dis Child 2002; 87:421-5.
8. Ozer EA, Kizilgunesler A, Sarioglu B, et al. The comparison of PRISM and PIM scoring systems for mortality risk in infantile intensive care. J Trop Pediatr 2004; 50:334-8.
9. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak 2005; 5:3.
10. Ramnarayan P, Britto J. Paediatric clinical decision support systems. Arch Dis Child 2002; 87:361-2.
11. Frize M, Ennett CM, Stevenson M, Trigg HC. Clinical decision support systems for intensive care units: using artificial neural networks. Med Eng Phys 2001; 23:217-25.
12. Hanson CW 3rd, Marshall BE. Artificial intelligence applications in the intensive care unit. Crit Care Med 2001; 29:427-35.
13. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. Lancet 1996; 347:1146-50.
14. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. Crit Care Med 2001; 29:291-6. Comment in: Crit Care Med 2002; 30:724.
15. Wong LS, Young JD. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks. Anaesthesia 1999; 54:1048-54.
16. Nimgaonkar A, Karnad DR, Sudarshan S, Ohno-Machado L, Kohane I. Prediction of mortality in an Indian intensive care unit. Comparison between APACHE II and artificial neural networks. Intensive Care Med 2004; 30:248-53.
17. Zernikow B, Holtmannspoetter, Michel E, et al. Artificial neural network for risk assessment in preterm neonates. Arch Dis Child Fetal Neonatal Ed 1998; 79:129-34.
18. Walczak S, Scorpio RJ. Predicting pediatric length of stay and acuity of care in the first ten minutes with artificial neural networks. Pediatr Crit Care Med 2000; 1:42-7.
19. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002; 35:352-9.
20. Ivakhnenko AG. Polynomial theory of complex systems. IEEE Trans Syst Man Cybern 1971; 1: 364-78.
21. Slater A, Shann F, Pearson G. PIM2: a revised version of the Pediatric Index of Mortality. Intensive Care Med 2003; 29:278-85.
22. Earle M Jr, Martinez Natera O, Zaslavsky A, et al. Outcome of pediatric intensive care at six centers in Mexico and Ecuador. Crit Care Med 1997; 25:1462-7. Comment in: Crit Care Med 1997; 25:1445-6.
23. Greenwood D. An overview of neural network. Behav Sci 1991; 36:1-33.
24. Obuchowski NA, Lieber ML, Wians FH Jr. ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin Chem 2004; 50:1118-25. Comment in: Clin Chem 2005; 51:471; author reply 471-2.

Appendix I: Technical details of GMDH.

The software used for ANN development was Knowledge Miner X. GMDH input-output model was used for modelling of ANN8, ANN9 and ANN 22. The paper by Ivakhnenko[20] provided the details of the mathematical derivative process of this algorithm. The best-fit model derived by GMDH was presented as a Kolmogorov-Gabor polynomial (Fig. 2).

Variables are the inputs of ANN model. The mathematical combination of two inputs is called a "neuron". Thus, a basic neuron (z1) is actually equal to:

$z1 = w0 + w1(u_1) + w2(u_2)$

where w0, w1, w2 are polynomial coefficients and u_1 and u_2 are inputs.

Neurons can also be mathematically combined to form another neuron. In Fig. 2, znx are all neurons. z22 is a neuron produced by combination of z11 and z12. Thus, a neuron (zn+1) can also be:

$zn+1 = w0 + w1\ (zn_1) + w2\ (zn_2)$

where w0, w1, w2 are polynomial coefficients and zn_1 and zn_2 are neurons.

The output (outcome, y) of an ANN model is a function of neurons too.

$y = w0 + w1\ (zn_1) + w2\ (zn_2)$

A hierarchical multilayer network of neurons is generated. "n" in znx is actually the layer number of that particular neuron.

Training a GMDH network involves the following processes:

1. Finding out the method of mathematical combination of inputs and neurons based on the outcome variable.
2. Selecting the relevant neurons which is useful for our prediction.

Training in a GMDH network is based on evolution principle. Computers generated all possible inputs and the best-fitted method of combination of inputs with respect to the outcome. The selection of relevant inputs required for an ANN model was based on a user-defined criterion. This criterion is governed by the percentage of performance (in terms of accuracy) which increases if a certain input is added to that model. In this study, we defined the inputs with less than 10% increase in performance to be eliminated from the ANN model. We restricted the maximum number of layers in all of our models to five.