

Soft Computing In Medicine: Nasopharyngeal Carcinoma Prognosis

O Baker, S Abdul-Kareem

Citation

O Baker, S Abdul-Kareem. *Soft Computing In Medicine: Nasopharyngeal Carcinoma Prognosis*. The Internet Journal of Medical Informatics. 2004 Volume 2 Number 1.

Abstract

Over the last twenty years, Soft Computing has developed rapidly as a discipline and method for the diagnosis and prognosis in Medical Informatics. This is a review of soft computing in the cancer subdomain of nasopharyngeal carcinoma cancer. An overview of the major categories of soft computing techniques is presented, and instances of these in NPC cancer research are discussed. An insight into the plausible makeup of NPC data, for current and future systems is presented. The review concludes by highlighting methods of employing unexploited areas of soft computing in NPC cancer. The research of Abdul Kareem group is the only research who applied artificial neural network and genetic algorithm to the NPC prognosis domain as well as considered a first step towards establishing a medical informatics research group and an attempt to increase the participation of Malaysians in the area of medical informatics by investigating the role of artificial Intelligence techniques as a cancer prognosticator.

INTRODUCTION

NASOPHARYNGEAL CARCINOMA

Cancer is a term for diseases in which abnormal cells divide uncontrollably, invading nearby tissues and spreading to other parts of the body via the bloodstream or lymphatic system (AJCC, 1998). Various types of cancer exist, including: lung, skin, breast, ovary, nasopharyngeal carcinoma, and cervix. In this study we are concerned with NPC because is one of the most common cancers in Malaysia, with incidence rates among Chinese males and females at 17.3 and 7.3 per 100,000 in the state of Selangor. The rates among the Malay males and females were 2.5 and 0.3 and among the Indian males, 1.1 while the incidence rate of Indian females are unknown (Prasad, 1992).

Nasopharyngeal cancer is a disease in which malignant (cancer) cells form in the tissues of the nasopharynx, the nasopharynx is the upper part of the pharynx (throat) behind the nose.

SOFT COMPUTING

Soft Computing comprises principally of genetic algorithms, artificial neural networks, and fuzzy logic.

ARTIFICIAL NEURAL NETWORK (ANN)

An artificial neural network (ANN) is an information processing system that tries to simulate biological neural

networks, ANN are distributed, adaptive, generally nonlinear learning machines built from many different processing elements (PE). Each PE receives connections from other PE and/or itself. The interconnectivity defines the topology. The signals flowing on the connections are scaled by adjustable parameters called weights.

Neural networks are typically arranged in layers. Each layer in a layered network is an array of processing elements or neurons. Information flows through each element in an input-output manner (Mizumoto and Shi, 1997)

GENETIC ALGORITHM (GA)

Genetic algorithms are global optimization algorithms based on the mechanics of natural selection and natural genetic and they employ a structured yet randomized parallel multipoint search strategy that is biased toward reinforcing search points of "high fitness". Genetic algorithms are similar to simulated annealing in that they employ random (probabilistic) search strategies. However, one of the apparent distinguishing features of genetic algorithms is their effective implementation of parallel multipoint search. (Yuan and Suarga, 1995)

FUZZY LOGIC (FL)

Fuzzy logic provides an approach to approximate reasoning in which the rules of inference are approximate rather than

exact. Fuzzy logic is useful in manipulating information that is incomplete, imprecise, or unreliable. Also called fuzzy set theory, fuzzy logic extends the simple Boolean operators, can express implication.

A FL system has a series of rules comprising of an antecedent and a consequent, combined as if –then semantics. An antecedent is a conjunction of input variables, each as an expressed and evaluatable degree of fuzzy set (membership function).a consequent is a single output variable as an expressed and evaluatable degree of some fuzzy set. A FL system consists of four stages in its operation (with various methods establishing a specific system): fuzzification, inference (Kosko, 1992), (Wang and Mendel, 1992) composition/ aggregation often combined with defuzzification process and defuzzification (Odusanya et al, 2000).

NPC AND SOFT COMPUTING RESEARCH

Evidently, there are many literatures on the use of Soft Computing in Cancer prognosis, especially from international publications (Odusanya et al, 2000, 2001, 2002) unfortunately there are very few write-ups on the applications of Soft Computing to NPC. Most of the studies are exploratory studies on neural network and its application on Medical informatics. There has been no reported instance of genetic algorithms or fuzzy logic in NPC cancer prognosis research.

ANN METHODS

Recently experiments of applying ANN on the prediction of survival outcome it has been gaining attractiveness. The researches on survival analysis using neural network are mostly in the domain of breast cancer, ovarian cancer, bladder cancer and prostate cancer only the research of Abdul-kareem has investigated the use of neural network technique to the NPC prognosis (Sameem et al, 1999, 2000a, 2000b , 2001a,2001b).

Abdul Kareem's group created and trained several back propagation neural networks using the same number of neurons, layers, epochs, goal and various training algorithms. Initially several training algorithms were used with respect to the NPC data model in order to determine the ideal training algorithm as is described in . Subsequently the resilient back propagation training algorithm was chosen based on the performance of the algorithm. The results obtained using different algorithms are shown in Table 1

Figure 1

Table 1: The AUC of Neural Network Models Using Different Training Algorithms

| Year | Trainoss | | Traincgp | | Traincgb | | Traincgr | | Trainrp | | Trainscg | |
|------|----------|-------|----------|-------|----------|-------|----------|-------|---------|-------|----------|-------|
| | ROC | SE | ROC | SE | ROC | SE | ROC | SE | ROC | SE | ROC | SE |
| 1 | 0.682 | 0.063 | 0.577 | 0.096 | 0.555 | 0.093 | 0.590 | 0.097 | 0.778 | 0.054 | 0.653 | 0.066 |
| 2 | 0.684 | 0.073 | 0.649 | 0.065 | 0.557 | 0.068 | 0.680 | 0.064 | 0.793 | 0.053 | 0.738 | 0.078 |
| 3 | 0.524 | 0.080 | 0.615 | 0.083 | 0.614 | 0.073 | 0.656 | 0.076 | 0.781 | 0.057 | 0.683 | 0.072 |
| 4 | 0.572 | 0.090 | 0.473 | 0.086 | 0.627 | 0.096 | 0.690 | 0.073 | 0.845 | 0.045 | 0.723 | 0.081 |
| 5 | 0.677 | 0.081 | 0.475 | 0.092 | 0.518 | 0.101 | 0.566 | 0.086 | 0.890 | 0.040 | 0.798 | 0.090 |
| 6 | 0.689 | 0.105 | 0.543 | 0.105 | 0.543 | 0.105 | 0.696 | 0.098 | 0.891 | 0.037 | 0.880 | 0.081 |
| 7 | 0.671 | 0.125 | 0.637 | 0.117 | 0.528 | 0.122 | 0.443 | 0.127 | 0.920 | 0.032 | 0.850 | 0.131 |
| 8 | 0.691 | 0.118 | 0.420 | 0.140 | 0.845 | 0.132 | 0.825 | 0.073 | 0.932 | 0.030 | 0.871 | 0.161 |
| 9 | 0.867 | 0.068 | 0.571 | 0.158 | 0.624 | 0.174 | 0.915 | 0.041 | 0.948 | 0.026 | 0.943 | 0.162 |
| 10 | 0.838 | 0.092 | 0.724 | 0.117 | 0.6 | 0.213 | 0.937 | 0.031 | 0.986 | 0.020 | 0.980 | 0.065 |

A comparison was also made between the performance of the back propagation neural network and the recurrent neural network for the prediction of the survival of nasopharyngeal carcinoma. The results are shown in Table 2.

Figure 2

Table 2: The AUC and Percent Accuracy of Feed Forward and Elman Network

| Year | FF | | | Elman | | |
|------|-------|-------|-------|-------|-------|-------|
| | ROC | SE | % | ROC | SE | % |
| 1 | 0.795 | 0.062 | 79.71 | 0.734 | 0.059 | 68.49 |
| 2 | 0.782 | 0.053 | 71.64 | 0.621 | 0.076 | 68.49 |
| 3 | 0.838 | 0.053 | 81.2 | 0.637 | 0.079 | 68.49 |
| 4 | 0.753 | 0.057 | 76 | 0.672 | 0.078 | 73.97 |
| 5 | 0.795 | 0.051 | 79 | 0.717 | 0.080 | 72.60 |
| 6 | 0.864 | 0.045 | 79.5 | 0.798 | 0.092 | 86.30 |
| 7 | 0.845 | 0.046 | 84 | 0.810 | 0.130 | 86.30 |
| 8 | 0.918 | 0.034 | 90 | 0.865 | 0.102 | 86.30 |
| 9 | 0.918 | 0.036 | 90 | 0.876 | 0.130 | 89.04 |
| 10 | 0.957 | 0.028 | 93 | 0.915 | 0.222 | 93.15 |

The main advantage of neural network technology is that the internal representation and distribution of data need not be known. Although neural network has not been tested extensively for modeling survival data, it is considered a good alternative for the prediction of survival of individual patients and it offers no obstacle to handling censored data (Prasad and Rampal, 1992).

GA AND FL

Soft computing researches into NPC cancer has been embarked on recently, and would continue to attract attention, as more successes are reported. Currently Baker and Abdul-Kareem are developing a new system based on the genetic algorithms for determining the prognosis of NPC.

For this purpose two models are created, namely:

1. A genetic algorithm that evolves algebraic rule-based classifiers.

2. A genetic algorithm with a hybrid function strategy that evolves algebraic rule-based classifiers.

The system in this instance would involve a decision as to the survival or non-survival of a patient within a time range of twenty years. When a patient is deemed to survive, this is specified for at least twenty years, when a patient is deemed not to survive the time of expected death is provided to the nearest one year, up to twenty years. To achieve the one year interval-based classifier, a series of sub-classifiers are generated to predict a specific time range each, and then chained together to operate in unison and resolve efficiently the prognostic outcome for a given patient (Baker and Abdul Kareem, 2004).

For the numeric NPC cancer data, a series of rules that have the following form are proposed:

$$C = C_{11}+C_{12}+C_{13}+C_{14}+C_{21}+C_{22}+C_{23}+C_{24} \quad (1)$$

Where C = prognostic value that is used to determine if the patient will live or die according to algorithm 1

$$C_{mn} = C_{24} \quad M = \text{rule no} \quad N = \text{chromosome no} \quad C = K_1F_1(X_1) + K_2F_2(X_2) + \dots + K_nF_n(X_n) \quad (2)$$

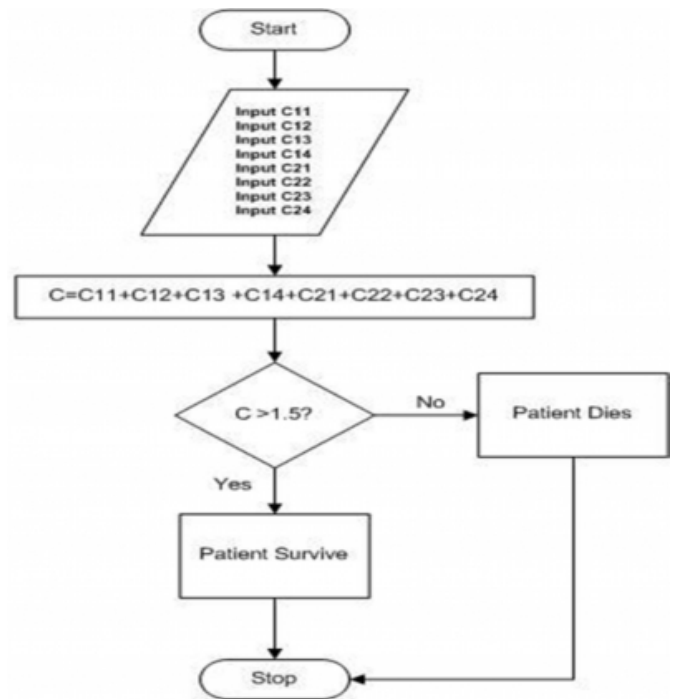
Where $X_n : n = 1$ to 22 represent one of the giving prognostic factor

$F_j (j = 1$ to 4) represent one of the following functions: $F(x) = x$, $F(x) = x^{0.02}$, $F(x) = \arctan(x)$, $F(x) = x^{0.45}$ $K_j (j = 1$ to 4) are coefficients to be determined

A threshold value of 1.5 is imposed for absolute numeric values evaluated for each classifier. The classifier value C is then compared with a threshold, as in Figure 1.

Figure 3

Figure 1: Evaluating the classifier threshold



Finally hybrid functions namely (fminsearch, patternsearch and fminunc) were examined with the genetic algorithm system and comparisons are made between the different methods. It is worthy of note that there has been no reported instance of the successful use of FL in the NPC domain.

NPC DATA STRUCTURE

The information to be analysed in NPC cancer could either be images or variables, parameters or markers.

Abdul-kareem has pointed out key markers in order of finding the ideal soft computing technique for NPC survival, where she developed a list of twenty-two variables that are considered important to NPC prognosis and are described in Table 3.

Figure 4

Table 3: Independent Variables

| Name | Description |
|-----------|------------------------------------|
| AGE | Age of patient |
| SEX | Sex |
| RACE | Ethnicity |
| NATURE | Nature of primary tumour |
| NERVE | Cranial nerve involvement |
| NODE | Node involvement |
| UPNECK | Node involvement in the upper neck |
| LONECK | Node involvement in the lower neck |
| DM | Distant metastasis |
| DURATION | Duration of symptoms |
| GLAND | Glandular presenting symptoms |
| EXT | Extension of primary tumour |
| SYMP | Sympathetic nerve involvement |
| BONE | Bone involvement |
| SUPRA | Supraclavicle node involvement |
| DIAM | Diameter of node |
| DIAGNOSIS | WHO Type |
| TREATMENT | Treatment Type |
| UICCT | T Type |
| UICCN | N Type |
| UICCM | M Type |
| UICCS | Stage |

The choice of the above variables is supported by numbers of specialist and the experiment result that conducted by Abdul-Kareem (Abdul-Kareem et al, 2000a).

CONCLUSION

While soft computing has been widely employed in breast cancer and ovarian cancer, there are very few to the area of nasopharyngeal carcinoma cancers, most specially the soft computing fields of the GA and FL systems.

Medical prognosis is a prediction of the future course and outcome of a disease and an indication of the likelihood of recovery from that disease. The researches on survival

analysis using neural network are mostly in the domain of breast cancer, ovarian cancer, bladder cancer and prostate cancer only the research of Abdul-Kareem and Baker has investigated the use of neural network technique and Genetic algorithm in the NPC prognosis. This shows that while there are a number of medical informatics researchers who are keen on investigating the potential application of neural networks in the various types of cancer, many previous studies have omitted the use of genetic algorithm in the field of NPC prognostic.

References

r-0. Abdul Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, Mohd Ibrahim A Wahid, "A Neural Network Approach to Modelling Cancer Survival". Radiotherapy Oncology, Journal of the European Society for Therapeutic Radiology and Oncology, vol. 58, Supp 1, Elsevier Science Inc, New York. Jan 2001.

r-1. Abdul-Kareem, S, Sapiyan Baba, Mohd Ibrahim A Wahid, 2000, "Research In Medical Informatics", Health Informatics Journal, Issue 6.2, Sheffield Academic Press Ltd., Sheffield England. June.

r-2. Abdul-Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, Mohd Ibrahim A Wahid, 2001, "A Neural Network Approach to Modelling Cancer Survival". Proceeding of the ICRO2001 Conference in Melbourne, 31st January - 3rd February 2001.

r-3. Abdul-Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, Mohd Ibrahim A Wahid, 2000 "ANN As A Tool For Medical Prognosis" Health Informatics Journal, Issue 6.3, Sheffield Academic Press Ltd., Sheffield England. September.

r-4. Abdul-Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, Mohd Ibrahim A Wahid, 1999, "Artificial Neural Network as a Prognostic Tool for Patients with Nasopharyngeal Carcinoma" proceeding of the XV Asia Pacific Cancer Conference, APCC, Chennai, India (12th-15th December 1999).

r-5. Abdul-Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, U Prasad, 2001, "Back Propagation Neural Network for Medical Prognosis: A Comparison of Different Training Algorithms", EJSAT, Electronic Journal of the School of Advanced Technologies, Vol.3 Issue 1, April 2001. Asian Institute of Technologies, Thailand, 2001. (<http://www.sat.ait.ac.th/ej-sat/>).

r-6. Abdul-Kareem, S, Sapiyan Baba, Yong Zulina Zubairi, U. Prasad, Mohd Ibrahim A Wahid, 2000, "Non Parametric Survival Models of Cancer", Proceedings of the International Mathematics Conference: Mathematics in the New Millennium Putra Jaya July 18-21st.

r-7. American Joint Committee on Cancer (AJCC) 1998, AJCC Cancer Staging Manual, ed. I.D. Fleming, J.S. Cooper, D.E. Henson, R.V.P. Hutter, pp. 1-41. Pennsylvania: Lippincott-Raven.

r-8. Cox, E. The fuzzy systems handbook. A practitioner's guide to building, using, and maintaining fuzzy systems. Academic Press, Inc., 1994.

r-9. Kosko, B. Neural networks and fuzzy systems. A Dynamical systems approach tomachine intelligence. 1st ed. Prentice-Hall International Editions, 1992.

r-10. Mizumoto, M., Shi, Y. A new approach of neurofuzzy learning algorithm. IN: Ruan, D. ed. Intelligent Hybrid Systems, Fuzzy Logic, Neural Networks, Genetic Algorithms, Kluwer Academic Publishers, 1997, pp.

109-129.

r-11. Odusanya, A.A., Odetayo, M.O., Naguib, R.N.G., and Petrovic, D. A review of soft computing and gynaecological cancer. IN: Sincak, P., and Vascak, J. eds. Quo Vadis Computational Intelligence?, Springer-Verlag Company, 2000, pp. 485-490.

r-12. Odusanya, A.A., Odetayo, M.O., Petrovic, D., and Naguib, R.N.G. The use of evolutionary and fuzzy models in oncological prognosis. IN: John, R., and Birkenhead, R. eds. Developments in Soft Computing, Springer-Verlag, 2001, pp. 207-215.

r-13. Odusanya, A.A., Odetayo, M.O., Petrovic, D., Naguib, R.N.G., Lakshmi, M.S., and Sherbet, G.V. A genetic algorithm-based model for breast cancer prognosis. IN: Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI2002), 2002. pp. 394-397.

r-14. Ohno-Machado L, Rowland TR. Neural Network Applications in Physical Medicine and Rehabilitation. American Journal of Physical Medicine and Rehabilitation 1999;78(4):392-8.

r-15. Oras F. Baker, Sameem Abdul Kareem, "Assessment of the use of soft computing models into survival analysis", International Medical Journal (IMJ). Volume 3 Number 2, December 2004.

r-16. Prasad, U. & Rampal, L., 1992, "Descriptive Epidemiology of Nasopharyngeal Carcinoma in Peninsula Malaysia", Cancer Causes Control, vol. 3, pp. 179-182.

r-17. Takagi, H. Introduction to fuzzy systems, neural networks, and genetic algorithms. IN: Ruan, D. ed. Intelligent hybrid systems. Fuzzy logic, neural networks, and genetic algorithms. 1st ed. Kluwer Academic Publishers, 1997. pp. 3-33.

r-18. Wang, L.-X., and Mendel, J.M. Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man, and Cybernetics, 1992, 22(6), November/December, 1414-1427.

r-19. Yuan, Y., and Suarga, S. On the integration of neural networks and fuzzy logic systems. IN: IEEE International Conference on Systems, Man and Cybernetics, 1995. pp. 452-7.

Author Information

Oras F. Baker

Faculty of Computer Science & Information Technology, University of Malaya

Sameem Abdul-Kareem

Faculty of Computer Science & Information Technology, University of Malaya