Expert Systems with Applications 41 (2014) 8092–8100

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome

Arso M. Vukicevic [a], Gordana R. Jovicic [a,*], Miroslav M. Stojadinovic [b], Rade I. Prelevic [c], Nenad D. Filipovic [a]

[a] Faculty of Engineering, University of Kragujevac, Serbia
[b] Clinic of Urology and Nephrology, Kragujevac, Serbia
[c] Military Medical Academy, Belgrade, Serbia

ARTICLE INFO

Article history:
Available online 11 July 2014

Keywords:
Bladder cancer
Expert systems
Artificial Neural Network
Genetic Algorithms
Regret Theory

ABSTRACT

Development of reliable medical decision support systems has been the subject of many studies among which Artificial Neural Networks (ANNs) gained increasing popularity and gave promising results. However, wider application of ANNs in clinical practice remains limited due to the lack of a standard and intuitive procedure for their configuration and evaluation which is traditionally a slow process depending on human experts. The principal contribution of this study is a novel procedure for obtaining ANN predictive models with high performances. In order to reach those considerations with minimal user effort, optimal configuration of ANN was performed automatically by Genetic Algorithms (GA). The only two user dependent tasks were selecting data (input and output variables) and evaluation of ANN threshold probability with respect to the Regret Theory (RT). The goal of the GA optimization was reaching the best prognostic performances relevant for clinicians: correctness, discrimination and calibration. After optimally configuring ANNs with respect to these criteria, the clinical usefulness was evaluated by the RT Decision Curve Analysis. The method is initially proposed for the prediction of advanced bladder cancer (BC) in patients undergoing radical cystectomy, due to the fact that it is clinically relevant problem with profound influence on health care. Testing on the data of the ten years cohort study, which included 183 evaluable patients, showed that soft max activation functions and good calibration were the most important for obtaining reliable BC predictive models for the given dataset. Extensive analysis and comparison with the solutions commonly used in literature showed that better prognostic performances were achieved while user-dependency was significantly reduced. It is concluded that presented procedure represents a suitable, robust and user-friendly framework with potential to have wide applications and influence in further development of health care decision support systems.

1. Introduction

According to the clinical reports, bladder cancer (BC) represents the most common urologic cancer in men and the fifth most common malignancy worldwide (Jemal et al., 2013). Since recently reported bladder tumours are superficial, the possibility of their progression and their muscle invasive (MI) nature make BC treatment a very challenging task with a profound influence on health care. The most important factors for predicting pathological stages are clinical staging based on physical examination, transurethral resection (TUR) pathology and imaging (Compérat & Van der Kwast, 2013; Ramy & Yair, 2011). However, predictions in BC remain nontrivial both in staging of primary tumour as well as in nodal staging (Bostrom et al., 2010). Since cancer classification commonly relies on clinical and histopathological information, clinicans' decision may turn out to be incomplete or misleading. As a result, clinical prediction has evolved from physician judgment alone to the use of various medical decision support systems and predictive models (Lughezzani et al., 2010). In literature, two most frequently used predictive models for prediction of BC outcome are Artificial Neural Networks (ANNs) and Logistic Regression (LR). After extensive work performed on comparing and analysing their performances, ANNs have gained increasing popularity over LR in research community (Bassi, Sacco, De Marco, Aragona, & Volpe, 2007; Dreiseitla & Ohno-Machadob, 2002; Hu et al., 2013). ANNs are data mining technique developed on the

* Corresponding author. Address: Sestre Janjic 6, Kragujevac 34000, Serbia. Tel.: +381 34334379; fax: +381 34333192.

E-mail addresses: arso_kg@yahoo.com (A.M. Vukicevic), gjovicic.kg.ac.rs@gmail.com (G.R. Jovicic), midinac@EUnet.rs (M.M. Stojadinovic), drrprelevic@hotmail.com (R.I. Prelevic), fica@kg.ac.rs (N.D. Filipovic).

http://dx.doi.org/10.1016/j.eswa.2014.07.006
0957-4174/

basis of biological process of learning (Looney, 1993). In cancer research, they have been used due to their ability to learn and recognize complex data patterns and identify nonlinear interactions between input (dependent) and output variables (Lisboa & Taktak, 2006). In oncological urology ANNs have been applied with promising results for the diagnostic (Çınar, Engin, Engin, & Ateşçi, 2009), staging (Anagnostou, Remzi, & Djavan, 2003) and prognostic (Saritas, Ozkan, & Sert, 2010) problems of prostate cancer. In BC patients ANNs have been used to predict outcome following screening (Finnea et al., 2000), prostate biopsy (Meijer et al., 2009) and radical cystectomy (El-Mekresh et al., 2009), as well as disease progression and tumour recurrence of non-invasive transitional cell carcinoma (Remzi, Anagnostou, & Ravery, 2003).

However, the application of ANNs in clinical practice remains limited due to a few reasons. It is assumed that users are able to choose an optimal combination of various ANN configuration parameters (such as: available activation functions, number of neurons and layers, learning algorithms, momentum, how much data to use for training, validation, testing, which objective-function to use for measuring quality of training and other settings). It may be noticed that for the efficient usage of ANNs clinicians should be familiar with a complex evolving foundation of ANN framework. As a result of such limitations, a wider usage of ANNs among clinicians remains unpopular despite promising results obtained by research community. In order to reduce user-dependency of ANNs based expert systems, evolutionary ANNs (EANNs) were proposed (Castellani, 2013; Yao & Liu, 1998). The basic idea of EANNs is to use an optimization procedure for increasing ANN performances iteratively with respect to some criteria (Castellani, 2013; Rivero, Dorado, Rabuñal, & Pazos, 2010; Tallón-Ballesteros & Hervás-Martínez, 2011). Independently from the progress in the field of neurocomputing, authors focused on the evaluation of medical decision support systems found that measuring clinical usefulness of decisions in medicine is not equal to measuring accuracy (according to which traditional ANN-based expert systems were configured) (Baker, 2009; Baker & Kramer, 2012; Holmberg & Vickers, 2013). It is explained that traditional way of measuring predictive accuracy does not capture important characteristics of intelligent behavior necessary for making rational decisions in medicine (for example taking into account clinical implications such as harms and benefits of making wrong and correct decisions – see Section 2.3 later) (Mallett, Haligan, Thompson, Collins, & Altman, 2012). To our best knowledge, obtaining reliable predictive models with respect to clinical needs remains an open question the answer to which could advance knowledge discovery in medicine (Esfandiari, Babavalian, Moghadam, & Tabar, 2014). Taking this into account, the aim of this study was to develop a robust procedure for obtaining ANN predictive models with high prognostic performances for the prediction of advanced BC in patients undergoing RC.

2. Materials and methods

For simplicity, this section is divided into three parts. First, the traditional expert-dependent configuration of ANNs is briefly introduced and the problems which this study aims to solve are higlighted. Next, in order to reach performances of expert ANN users with minimal user-system interaction, an evolutionary approach for configuring of ANN was presented. Finally, a regret approach was applied for evaluating clinical usefulness of ANN predictive models in order to estimate best ANN model for a particular problem.

2.1. Common ANN configuration and training procedure for the purposes of classification and prediction of advanced BC

ANN (Wallis, 1999) may be described as a mathematical model which on a much smaller scale mimics the way a biological neural network works (Fig. 1(a)). Transmission of electrical signals over neuron connections (axon and dendrites) is mathematically modeled as a sum of n weighted scalar inputs p_i and constant b (called bias): $s = \sum_i^n p_i w_i + b$. The result is then used as an argument of an activation function f, which produces the output $t = f(s)$.

The central idea of the ANN framework is that by adjusting the scalar parameters b_i and w_i an artificial neuron can exhibit a desired intelligent behavior (such as classification, prediction or estimation). The most frequently used types of transfer functions are the hard-limit (or step), linear, sigmoid (or logistic) and soft max, to name just a few (Karlik & Olgac, 2010). A common ANN architecture consists of many neurons organized in layers. The ANN architecture considered in this paper is a feed-forward multi-layer perception (FFN) with a single hidden layer as the most suitable for the purposes of binary classification and survival prediction (Zhang, 2000) (Fig. 1(b)).

Regarding the ANN training, there are two main principles of learning: supervised and unsupervised. The subject of this paper was supervised learning with backpropagation learning algorithm (Yu & Chen, 1997), when the network is provided with a set of examples (pair of inputs and known correct outputs). For the purpose of learning, the input data are commonly divided into three sets: training, validation and testing. A training data set was used only for learning (adjusting weights and biases). Validation set was used to decide when to stop the training process – to avoid overfitting (a situation when the ANN memorizes the training data rather than learning the rules that govern them). A testing set was used for independently measuring performance of the trained network. In the beginning of the training, the network was initialized with randomly chosen weights. After the inputs were applied to the ANN, for every qth iteration (learning epoch), the prediction was compared with the true category by calculation of classification error E^q. Since E^q is a continuous and differentiable function of l weights between neurons, it could be minimized by using an iterative gradient descent procedure. After calculating $\nabla E^q = \left(\frac{\partial E^q}{\partial w_1^q}, \frac{\partial E^q}{\partial w_2^q}, \ldots, \frac{\partial E^q}{\partial w_l^q}\right)$, each weight in the next $(q+1)$ epoch was updated using the increment Δw_i^q: $w_i^{q+1} = w_i^q - \Delta w_i^q = w_i^q - \gamma \frac{\partial E^q}{\partial w_i^q}$ for $i = 1 \ldots l$, where γ represents the learning constant which defines the step length (in weights space). Therefore, the whole process of ANN learning is in the end reduced to iteratively adjusting the neurons' interconnections ("weights") until the classification error converges ($\nabla E = 0$) with respect to some criteria and configuration parameters. For the purposes of minimizing error-function, different algorithms may be used: Levenberg–Marquardt, Gradient Descent, Polak–Ribiére Conjugate Gradient and BFGS Quasi-Newton, to name just a few (Ghaffari et al., 2006; Lethaus, Baumann, Köster, & Lemmer, 2013; Mukherjee & Routroy, 2012).

To sum up, training of the ANN to do a particular task may be intuitively described as choosing various above mentioned parameters. Since a significant number of parameters is required to be set correctly, a deeper understanding of ANN framework remains necessary which represents an obstacle for the wider application of ANN predictive models among clinicians (or nonexperts in general).

2.2. Managing the process of ANN training by using Genetic Algorithms

Genetic Algorithm (GA) is an iterative method for solving both constrained and unconstrained optimization problems (Yang, (2014, chap. 5 – Genetic algorithms)). The process of optimization starts from an initial guess of parameters (called population), which are the subject of optimization. At each iteration (called generation), the GA selects some portion of best individuals from the current population and uses them as parents to produce the