Expert Systems with Applications 40 (2013) 466–470

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

# Fuzzy expert system for predicting pathological stage of prostate cancer

M.J.P. Castanho [a,*], F. Hernandes [b], A.M. De Ré [b], S. Rautenberg [b], A. Billis [c]

[a] Department of Mathematics, Universidade Estadual do Centro-Oeste, Guarapuava, PR, Brazil
[b] Department of Computation Science, Universidade Estadual do Centro-Oeste, Guarapuava, PR, Brazil
[c] School of Medicine, Universidade Estadual de Campinas, Campinas, SP, Brazil

ARTICLE INFO

*Keywords:*
Fuzzy rule-based system
Genetic algorithm
Prostate cancer

ABSTRACT

Prostate cancer is the second most common cancer among men, responsible for the loss of half a million lives each year worldwide, according to the World Health Organization. In prostate cancer, definitive therapy such as radical prostatectomy, is more effective when the cancer is organ-confined. The aim of this study is to investigate the performance of some fuzzy expert systems in the classification of patients with confined or non-confined cancer. To deal with the intrinsic uncertainty about the variables utilized to predict cancer stage, the developed approach is based on Fuzzy Set Theory. A fuzzy expert system was developed with the fuzzy rules and membership functions tuned by a genetic algorithm. As a result, the utilized approach reached better precision taking into account some correlated studies.



## 1. Introduction

In Brazil as well as worldwide, prostate cancer is the second most common male malignancy (INCA, 2010). Accurate estimates of the pathologic stage are essential for physicians to decide the appropriate therapy. When the cancer is confined to the prostate, the cure rate after surgery is approximately 80%.

Clinical stage, serum prostate-specific antigen (PSA) concentration and Gleason score are among the most recognized factors. A combination of these three parameters leads to a score used to define prognostic groups that are routinely used in daily practice.

Lately, numerous tools to aid the medical decision-making have been developed. Particularly, for predicting final pathologic stage, there are statistical approaches and computer systems based on artificial intelligence.

In the literature we can find several papers about cancer using soft computing techniques, such as artificial neural networks (Han, Snow, Brandt, & Partin, 2001; Matsui et al., 2002; Saritas, Ozkan, & Sert, 2010; Snow, Smith, & Catalona, 1994), fuzzy logic (Castanho, Barros, Yamakami, & Vendite, 2008; Saritas, Allahverdi, & Sert, 2003; Seker, Odetayo, Petrovic, & Naguib, 2003), genetic algorithms (Baker & Abdul-Kareem, 2007; Ghosh, Mitchell, Tanyi, & Hung, 2010; Ludwig & Roos, 2010; Odusanya et al., 2002; Shah & Kusiak, 2007), neuro-fuzzy systems (Benecchi, 2006; Keles, Hasiloglu, Keles, & Aksoy, 2007; Papageorgiou et al., 2008) and genetic-fuzzy systems (Peña-Reyes & Sipper, 1999; Sedighiani & Hashemi-Khabir, 2009).

The aim of this study is to investigate the use of a genetic-fuzzy system to predict the pathological stage of prostate cancer, combining preoperative serum PSA, clinical stage, and biopsy Gleason score. In the next section, a literature review is done. Section 3 describes the development of the genetic-fuzzy system. Finally, Section 4 presents the best system found and compares its ability to discriminate prostate-confined cancer and probability tables.

## 2. Literature review

Predictive tools are essential for individualized, evidence-based medical decision making (Shariat, Karakiewicz, & Roehrborn, 2008a). Probability tables are statistical tools to make prostate cancer prognoses.

Also, some soft computing components are being used for this purpose.

### 2.1. Statistical tools

There are numerous published probability tables available to help clinicians in the task of predicting the pathologic stage of prostate cancer. However, just some of these were validated and their usefulness is not completely defined (Ross, Scardino, & Kattan, 2001).

The first one and the most widely used is the Partin tables. In 1997, Partin et al. (1997), presented probability tables combining preoperative serum prostate-specific antigen (PSA), TNM clinical stage and Gleason grade from the prostate biopsy. Multinomial log-linear regression was performed to provide the likelihood of final pathological stages as follows: organ-confined, capsular penetration, positive seminal vesicle involvement and lymph node

* Corresponding author. Address: Caixa Postal 3010, CEP 85015-430, Guarapuava, PR, Brazil. Tel.: +55 04236298348; fax: +55 04236211090.
*E-mail address:* zeza@unicentro.br (M.J.P. Castanho).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.07.046

involvement. These tables were updated in 2001 (Partin et al., 2001) and 2007 (Makarov et al., 2007). Augustin, Sun, Isbarn, Pummer, and Karakiewicz (2012) compared the three versions and did not identify the best one.

The Partin tables represent the most widely used guide towards the selection of definitive therapies in Europe and North America (Bhojani et al., 2009). The European Association of Urology (EAU) guidelines recommend the use of Partin tables and Kattan nomogram for prostate cancer staging (Briganti, Karakiewicz, Joniau, & Van Poppel, 2009).

The Partin tables were validated in several studies around the world. Some results are described in Table 1.

It has been demonstrated that predictions made by probability tables are more accurate than those by clinicians, regardless of their level of expertise (Chun et al., 2006; Shariat et al., 2008a).

### 2.2. Soft computing methods

Over the last twenty years, soft computing techniques have been applied to develop computational systems for diagnosis and prognosis in Medicine. Soft computing comprises principally artificial neural networks, fuzzy logic and genetic algorithms (Baker, Hassani, & Kareem, 2008).

An artificial neural network (ANN) is a system based on the operation of biological neural networks. The use of ANNs in prostate cancer was pioneered by Snow et al. (1994). Some years later, Han et al. (2001) evaluated an ANN for the prediction of the pathologic stage in prostate carcinoma and found results slightly superior to the probability tables. To organ-confined disease the AUC was 0.77 using ANN and 0.72 with the probability tables. In 2001, Sargent (2001) carried out a literature review of 28 studies comparing ANN with standard statistical techniques (regression) and concluded that regression approaches have several desirable features in terms of ease of use and the ability to draw inferences based on their output. In addition, neither method supersedes the other in predictive performance. Matsui et al. (2002) compared ANN for predicting pathological stage of clinically localized prostate cancer in the Japanese population with the Partin tables. The AUC of the ANN with the same three parameters of the tables was 0.825 while with Partin tables it was 0.756 but did not attain statistical significance. Shariat, Karakiewicz, Suardi, and Kattan (2008b) carried out an analysis of the methods available in the literature: nomograms, risk groupings, ANN, probability tables and classification and regression tree analysis. They concluded that the probability tables have the highest accuracy and the best discriminating characteristics for predicting outcomes in prostate cancer patients. Saritas et al. (2010) developed an ANN to predict whether patients have prostate cancer or not before biopsy. The AUC was 0.94. In 2003 they developed a fuzzy expert system to diagnose prostate cancer but the results overestimate the literature (Saritas et al., 2003). Chen, Zhang, Xu, Chen, and Zhang (2012) compare the diagnostic performances of ANN and multivariable logistic regression analysis for differentiating between malignant and benign lung nodules on computed tomography scans.

Another soft computing technique is based on fuzzy logic (FL). Fuzzy rule-based system is a mathematical tool for dealing with the uncertainty and the imprecision typical in medical field. The reasoning is based on compositional rule of fuzzy inference and the knowledge of specialists is important to determine the parameters. Seker et al. (2003) investigated the Fuzzy K-Nearest Neighbor (FK-NN) algorithm to estimate the accuracy of breast and prostate cancer prognoses and determined the significance of prognostic markers. Castanho et al. (2008) developed a fuzzy rule-based system to predict the pathologic stage of prostate cancer. For organ-confined disease the AUC was 0.76.

The third soft computing component is genetic algorithm (GA), which is a methodology inspired by natural evolution. This component was used by Shah and Kusiak (2007) to identify and classify ovarian, prostate and lung cancer; Ludwig and Roos (2010) and Odusanya et al. (2002) to investigate the prognosis of breast cancer; Baker and Abdul-Kareem (2007) to introduce a new method for the prognosis of nasopharyngeal carcinoma and Ghosh et al. (2010) for prostate cancer treatment planning.

Since these three methods are complementary rather than competitive, many researchers have hybridized ANNs, FL and GAs to develop a better performance system. Neuro-fuzzy systems use fuzzy systems to represent and process the knowledge in a clear way with easy interpretation and utilize the learning capacity of ANN. They were used to classify prostate cancer (Benecchi, 2006; Keles et al., 2007) and to diagnose breast cancer (Keles, Keles, & Yavuz, 2011). Karabatak and Ince (2009) presented an automatic diagnosis system for detecting breast cancer based on association rules and neural network.

A genetic-fuzzy system is a rule-based fuzzy system where a genetic algorithm is used to optimize the parameters. They were used in breast cancer diagnoses (Peña-Reyes & Sipper, 1999; Sedighiani & HashemiKhabir, 2009). The importance of genetic-fuzzy systems was emphasized in Cordón, Gomide, Herrera, Hoffmann, and Magdalena (2004) and Herrera (2008). Peña-Reyes and Sipper (1999) combined fuzzy-genetic approach and attained high classification performance in breast cancer diagnosis also simplified the resulting system. In this context and believing in improving the literature results, in this study, a genetic-fuzzy system for predicting pathological stage of prostate cancer is proposed.

## 3. Proposed system

Between January 1997 and June 2008, 331 patients ranging from 43 to 76 years old (median 64) were treated with radical prostatectomy for clinically localized prostate cancer at Clinics Hospital located at the Universidade Estadual de Campinas, Brazil. A database description is in Table 2. The surgical specimens (prostate and adjacent structures) from each patient were examined by the same pathologist and the pathologic stage established: 289 patients had organ-confined cancer (TNM, pT2) and 42 patients had extraprostatic cancer (TNM, >pT2). The pathological grade

**Table 1**
Performance of Partin tables validated in different population samples. Area under Receiver Operating Characteristic curve (AUC) values for organ confined disease (OC), extraprostatic extension (EPE), seminal vesicle involvement (SVI) and lymph node involvement (LN).

| Year | Authors | Samples | OC | EPE | SVI | LN |
|---|---|---|---|---|---|---|
| 2005 | Eskicorapci et al. | Turkish | 0.66 | – | 0.73 | 0.76 |
| 2005 | Gorziza | Brazilian | 0.65 | 0.54 | 0.63 | 0.77 |
| 2006 | Ayyathurai et al. | Welsh | 0.73 | – | 0.74 | 0.78 |
| 2008 | Naito et al. | Japanese | 0.70 | – | – | – |
| 2008 | Heath el al. | African American | 0.73 | 0.62 | 0.77 | – |
| 2010 | Yu et al. | American | 0.68 | 0.62 | 0.77 | 0.74 |
| 2010 | Fanning et al. | Irish | 0.58 | 0.54 | 0.66 | 0.55 |

**Table 2**
Classification of patients database.

| PSA | Patients | Gleason score | Patients | Clinical stage | Patients |
|---|---|---|---|---|---|
| 0–2.5 | 15 | 4 | 3 | T1a | 9 |
| 2.6–4 | 23 | 5–6 | 211 | T1b | 4 |
| 4.1–6 | 72 | $7 = 3 + 4$ | 84 | T1c | 149 |
| 6.1–8 | 63 | $7 = 4 + 3$ | 17 | T2a | 118 |
| 8.1–10 | 49 | 8–10 | 16 | T2b | 41 |
| >10 | 109 | | | T2c | 10 |