A Machine Learning Prototype for the Identification of Patients who are Not Likely to Respond to Medical Treatment of BPH with α-Blockers

Barış ANDIRINLI[a] , Ali TEKIN[b], Yasin KAPUCU[c], I. Ilkay BODUROGLU[d]
[a,d] Informatics Institute, Istanbul Technical University, Istanbul, Turkey
[b,c] Haydarpaşa Numune Training and Research Hospital, Istanbul, Turkey

May 2006

Abstract: *This machine learning study aims to identify - prior to treatment - the patients who are not likely to respond to drug therapy - with α-blockers - of Benign Prostatic Hyperplasia (BPH). Medical records of 310 BPH patients are used. The data set is divided into two classes, namely positive and negative-response groups, with respect to the change in the patient's International Prostate Symptom Score (IPSS) after drug therapy. Age, IPSS, Prostate Volume, Post Void Residual Urine (PVR), Prostate Specific Antigen (PSA), free-PSA, and Creatinine are examined as baseline clinical factors, along with our novel synthetic features in order to propose a decision tree to classify a new patient. Our learning scheme, which uses synthetic variable generation and selection routines along with the C4.5 Decision Tree Algorithm correctly classifies 71.5% of patients, with sensitivity 71.3%, and specificity 71.6%. Synthetic variables do enhance its performance. Our prototypical classifier's success rate reveals that failure is predictable prior to medical treatment of BPH.*
Keywords: *Machine Learning, Benign Prostatic Hyperplasia, alpha-Blockers, Medical Treatment Decision, Decision Trees, Synthetic Variables.*

1. Introduction

BPH is defined as the non-cancerous enlargement of the prostate gland. It is the most frequent urological problem among men, which primarily results in lower urinary tract symptoms (LUTS). The prevalence of BPH among men over 60 is over 50%. LUTS are not classified as a disease; rather, they are characterized as bothersome voiding. The tolerance limit to these symptoms differs from one individual to another and from one community to another [2].

BPH has a significantly worsening effect on a patient's quality of life. If left untreated, BPH and associated LUTS may progress toward bladder infections, bladder calculi, and more serious complications such as acute urinary retention (AUR), which usually necessitates invasive treatment. 93% of urologists believe that BPH is a progressive disease [6].

The treatment choice of BPH would vary from patient to patient. The patient has to consider the severity of his symptoms and be aware of the possible benefits and risks of each treatment option. The urologist is responsible for guiding the patient to choose the most beneficial treatment. A survey conducted among practicing urologists revealed different preferences for the treatment of BPH [8]. Almost all urologists use Digital Rectal Examination (DRE) and PSA tests to diagnose BPH as recommended in the guidelines while preferences vary for optional tests. Beside emerging technologies in the treatment of BPH, there is an increasing tendency to use drugs. There are many articles in the literature comparing the efficacy of two classes of drugs, α-adrenergic blockers (a.k.a. α-blockers) and 5-α reductase inhibitors, while some researchers focus on the effect of combination drug therapy [9]. On the other hand, few studies examine the efficacy of every single drug for different categories of patients segmented according to baseline clinical factors. In one of the studies in identifying factors that predict the outcome of medical therapy, IPSS and prostate volume are reported to be useful predictors of medical treatment response [1]. Another study reveals that among men with large prostate, employing Finasteride is the most

[a] Informatics Institute, Istanbul Technical University, Maslak Campus, Sariyer, 34469 Istanbul, Turkey Tel: +90 (532) 486-7911 Fax: +90 (226) 462-9055 E-mail: andirinli@be.itu.edu.tr
[b,c] Haydarpaşa Numune Training and Research Hospital, Tibbiye Caddesi, Uskudar, Istanbul, Turkey Tel: +90 (216) 414-4502 Fax: +90 (216) 336-0565 E-mail: aaftekin@tr.net (Ali Tekin), dryasin@yahoo.com (Yasin Kapucu)
[d] Corresponding and presenting author: Informatics Institute, Istanbul Technical University, Maslak Campus, Sariyer, 34469 Istanbul, Turkey Tel: +90 (533) 575–1674 Fax: +90 (212) 285-7194 Ext: 9 E-mail: ilkay@be.itu.edu.tr

effective medical treatment [10]. Our aim is to identify the baseline factors for predicting the medical treatment failure of BPH and propose an easily interpretable decision tree to help urologists predict whether a patient will respond to α-blocking therapy, or not. This way, correctly classified negative-response group patients would be saved from facing the unnecessary trouble of treatment with α-blockers, by proceeding directly with surgical treatment, or choosing another medical treatment option.

2. Set Up

Medical records of 310 patients who visited Istanbul Haydarpaşa Numune Training and Research Hospital's Urology Clinic between 1999 and 2006 and were diagnosed with BPH were used in this study.

2.1. Data Set

BPH patients with LUTS, over 50 years of age, and who received at least two months of medical treatment with α-blockers were eligible for our study. It is widely accepted that two months are enough to determine whether a patient responds to this kind of treatment, or not. Exclusion criteria from the study were previous prostate surgery, history of AUR, prostate malignancy, urinary bladder malignancy, neurogenic bladder, PSA > 10ng/ml, patients with significant hepatitic, renal or cardiovascular dysfunction [1].

310 patients were appropriate for these criteria. IPSS, Prostate Volume, Post Void Residual Urine (PVR), PSA, f-PSA, and Creatinine data were evaluated. Prostate Volume and PVR were measured with a transrectal USG. Standard ellipsoid form [11] was used for the measurement of Prostate Volume (width x length x height x 0.5236), and chemo-luminescence, for PSA measurement [12].

All patients were prescribed with an α-blocker. The dosage of the α -blocker was Doxazosin (4 mg), Terazosin (5 mg), Tamsulosin (0.4 mg) and Alfuzosin (10 mg), once daily. IPSS was re-measured after 2 months of medical treatment. We define medical treatment failure (or negative response) as the less than 30% decrease in IPSS at the end of two months. We ended up with a set of 310 instances satisfying the above criteria. 167 of them were tagged as "positive-response", while 143 as "negative-response". The negative-response rate was 46.1%. Related studies have a smaller failure rate, such as 22.9% [1]. We believe that medical treatment failure rate of the patients who come to Haydarpaşa Numune Training and Research Hospital is above average due to the fact that it is a well-known state tertiary referral center in Istanbul, whose population is 12 million. Furthermore, Istanbul is easy to access from other cities of the enclosing Marmara Region.

2.2. Features

We included in our analysis the factors, or features, that predict the progression of BPH. Even though there is no consensus, The American Urological Association (AUA) Symptom Index, a.k.a. IPSS [4], Serum Creatinine [2], PVR [22], PSA [13], Urinary Flow Rate [9] (Q_{max} and Q_{min}) are considered to be among strong predictors. In light of these, all the features relevant to the progression of BPH are considered as relevant for the prediction of medical treatment failure and are included in this study. However, we had to exclude Q_{max} and Q_{min} values since urinary flow rate data are missing for most of our patients. Being a state hospital accommodating middle and lower-income-level people with or without health insurance, Haydarpaşa urologists do not prescribe the Urinary Flow Rate measurement, which is an optional test.

3. Method

Machine Learning is the process of discovering patterns in data. Useful patterns allow us to make nontrivial predictions on new instances of data. In our specific case, we try to discover rules that would classify patients who would benefit from medical treatment of BPH, prior to medical treatment (with α -blockers.)

3.1. The C4.5 Decision Tree Algorithm

Decision trees have been popular in machine learning applications due to their simplicity, short decision evaluation time, and high interpretability. We employed the C4.5 Decision Tree Induction Algorithm [14], which is a successor to ID3, another well-known algorithm in the same field. Note that C4.5 can handle numerical attributes as well as discrete valued ones, whereas all attributes have to be discretized prior to training with ID3.

3.2. Performance Metrics

The usual confusion matrix is a 2x2 matrix consisting of the numbers of testing data instances classified by the classifier as either true positive (TP), or false negative (FN), or false positive (FP), or true negative (TN). The overall classification success rate is simply given by $\frac{TP+TN}{TP+TN+FP+FN}$. Other well-established performance metrics are sensitivity, $\frac{TP}{TP+FN}$, and specificity, $\frac{TN}{TN+FP}$. Obviously, the classifier's performance has to be measured on test data independent from the data with which the classifier is trained.

General practice is to reserve 1/3 of the original data as test data and the remainder as training data. However, it is not very efficient to use only 2/3 of the data in the training phase. Therefore, we employ the Leave-One-Out Validation Method, which is simply an *n*-fold cross-validation method, where *n* is the number of instances in the original data set [15]. This way, *(n-1)* data instances are used for training and only one is left out for testing. Of course, this procedure is repeated *n* times, until each data instance gets tested exactly once using the classification method at hand. Note that the Leave-One-Out Method validates a given *classification method* and *not* a given *classifier* produced by this method since there are always *n* (possibly distinct) classifiers produced in this validation method. However, since validating the classification method using testing data independent from the training data is enough for all practical purposes, we believe that this validation method is the most efficient one.

4. The Machine Learning Process

4.1. Motivation

While the general practice is to evaluate the overall symptom score, IPSS, where IPSS = **I + F + H + U + A + S + N** (See Table 1 for the definitions of these seven IPSS subscores), a BPH study conducted among Japanese men reports that the "**N**octuria [sub]score of IPSS behaves differently in the symptom complex." [16] This study, along with our initial findings listed in Table 1, encouraged us to handle the components of IPSS separately.

In order to put together Table 1, we performed an independent two-sample t-test and generated a confidence interval for all of the features included in this study. (Only 202 data instances were used in this step since these 202 instances did not have any missing data fields in them.) Notice that even though the means of IPSS for the two groups are significantly different, there is enough reason to suspect that adding the seven subscores to form the overall IPSS would result in a loss of useful information. This is because only three subscores pass the significance test. Note that up till here, what we have done is to discuss what motivated us to do research in this direction. In other words, the training stage of machine learning has not started as of now. Since training has not started yet, we need not to worry about separating testing data from training data.

4.2. Synthetic Feature Generation & Selection Among Many

Note that we use the Leave-One-Out Validation Scheme to validate the synthetic feature generation and selection step and at every step after this one. Since training starts at this point in the game, we are fine. That is, since we do not use the results from Table 1 in this step or in any one of the remainder of the training process, our testing instances are independent from our training instances.

Feature generation and selection is a very important step of the representation problem in machine learning. This is because useless attributes deteriorate the performance of learning schemes. For example, adding a perfectly arbitrary binary attribute (by tossing an unbiased coin) to a standard data set deteriorates the performance of C4.5 by 5 to 10% [15]. This seems meaningless at the first glance since at each step, C4.5 chooses the "best" attribute to split with respect to minimizing "entropy" and thus maximizing information value. However, when proceeding down the tree, as less data are available to choose the best attribute, an irrelevant attribute may accidentally seem the best to split the tree further.

Synthetic feature generation explicitly reveals feature dependencies that may not have been recognized by the learning scheme using existing features [17]. We employed a genetic-algorithm-like method to generate synthetic features [17] using only the discrete valued variables, namely, the IPSS subscores and QoL. We represented each IPSS subscore and QoL as a

Table 1

Comparison of the Baseline Clinical Factors for Positive & Negative Response Groups

μ = sample mean, σ = sample standard deviation, CI = Confidence Interval

Factors	All Patients		Positive		Negative		p-value	95% CI	
	μ	σ	μ	σ	μ	σ			
IPSS Subscores									
I ncomplete emptying	2.8	1.8	2.7	1.8	2.9	1.8	0.350	-0.262	0.738
F requency	3.0	1.6	2.9	1.7	3.2	1.4	0.177	-0.135	0.729
H esitancy	2.3	1.7	2.2	1.7	2.5	1.7	0.180	-0.148	0.781
U rgency	0.7	1.1	0.4	0.8	1.1	1.3	**0.000**	0.380	0.986
A weak Stream	3.2	1.6	3.0	1.7	3.4	1.4	0.082	-0.050	0.822
S training to initiate urination	2.5	1.6	2.1	1.5	2.9	1.6	**0.000**	0.432	1.310
N octuria	3.2	1.2	3.0	1.2	3.4	1.3	**0.017**	0.074	0.758
IPSS	17.8	6.1	16.2	6.2	19.4	5.6	**0.000**	1.571	4.844
Quality of Life (QoL)	3.7	1.0	3.4	1.0	4.0	0.9	**0.000**	0.330	0.838
Creatinine	1.0	0.3	1.0	0.2	1.0	0.3	0.131	-0.124	0.016
PSA	3.4	3.0	3.2	2.6	3.7	3.3	0.184	-0.265	1.373
f-PSA	0.8	0.8	0.7	0.6	0.9	0.9	0.175	-0.064	0.350
Prostate Volume	57.5	41.7	51.8	24.5	63.3	53.2	0.051	-0.035	23.025
PVR	76.9	67.1	62.0	53.8	91.8	75.6	**0.002**	11.549	47.996
Age	65.0	7.6	64.6	7.8	65.4	7.5	0.435	-1.280	2.963

"gene" and applied the (+) and the (^2 + ^2) operators to (some, but not all of the) "gene combinations" having at most four genes. The motivation behind this choice of operators is as follows: The sum of IPSS subscores of 3 + 3 + 3 for a positive-response patient would be considered identical to that of a patient belonging to the negative-response group with subscore values 1 + 3 + 5. However, the (^2 + ^2) operator identifies the distinctness of the two patterns observed with (9 + 9 + 9) and (1 + 9 + 25). In this study, we have not done anything near an exhaustive search of operators, or of the optimal number of operands, because of the obvious computational burden. That is why we call this study "a machine learning prototype". Therefore, the next step in our line of research will be to make use of the high performance computing facility at the Informatics Institute of Istanbul Technical University for synthetic feature generation (and selection) by searching over a larger space of operators and operands.

At the end of synthetic feature generation, we employed a "wrapper approach" to select the relevant features, starting with all the features - the original fifteen along with all the synthetic ones. The wrapper approach often outperforms the "filter approach", because only the former optimizes feature selection for the particular learning algorithm at hand [18].

Table 2 shows the results when we did a greedy reduction in the number of features included in the tree, initially containing all the features. In every iteration, a randomly selected feature that was close to the lowest level of the tree was put aside, and the tree was created from scratch. If and only if the new tree had a higher success value, the aforementioned feature was removed from the model forever. As seen in the bottom row of Table 3, we acquired the highest performance using only the following three features, (**S + U + A + N**), ($\mathbf{S}^2 + \mathbf{U}^2 + (\text{QoL})^2$), and PVR. We call this set of features the "Relevant Feature Set". Up till here, we have been using 202 instances, none of which had any missing data fields. As we have just decided that the four IPSS components, **S**, **U**, **N**, and **A**, along with QoL, are among the most relevant discrete valued fields, we also excluded fro the study the 12 instances, out of the initial 310, because they had missing values for at least one of these five fields. Furthermore, we removed the features Age, I, F, H, PSA, f-PSA, and IPSS from the model. At this point, we filled in missing values for the fields of PVR, Volume, and Creatinine using Amelia, a software package that uses the Expectation Maximization Importance Sampling (EMis) algorithm [19, 20], ending up with 298 instances, 96 of whose one or more missing data fields have been imputed.

Table 2
The performance of the classifier with various feature sets

Number of Candidate Features	Performance (%)*
39	57.4
29	61.4
19	63.3
9	65.9
3	70.8
* measured as percentage of correctly classified instances out of 202.	

5. Results

Using the new data set with $n = 298$ instances, the best performance out of C4.5 decision tree algorithm was acquired using only the Relevant Feature Set while restricting the minimum number of instances in a leaf to 5 and allowing pruning of the tree at the end. (See Table 3.) WEKA-3-4 was used for all classification tasks [21]. The confusion matrix elements are: TP = 112, TN = 101, FP = 40, and FN = 45. As mentioned before, these are results from the Leave-One-Out Validation Method. We noticed that adding features to the Relevant Feature Set deteriorates the performance, in accordance with our findings summarized in Table 3. The decision tree structure that we propose is shown in Figure 1. It can be employed as an easy-to-interpret guideline by urologists to determine if a BPH patient would benefit from medical treatment with α-blockers, or not. Each path to the leaves is interpreted as a rule. For example, the rightmost path means: If $(\mathbf{S})^2 + (\mathbf{U})^2 + (\mathrm{QoL})^2 > 25$ and PVR > 76, then the patient will not benefit from medical treatment, with a certain probability. This probability is *not* the success rate of the classifier because the success rate, 71.5%, is just a prior probability. The posterior probability that the patient will not benefit from medical treatment, given the classifier predicted a negative response, is $\frac{TN}{TN+FN}$, which equals 69%. On the other hand, following the leftmost path, If $(\mathbf{S})^2 + (\mathbf{U})^2 + (\mathrm{QoL})^2 \leq 25$ and $(\mathbf{S} + \mathbf{U} + \mathbf{A} + \mathbf{N}) \leq 4$, then the patient the patient will benefit from medical treatment, with a certain probability. In this case, the posterior probability that medical treatment would be beneficial is 73.7%. Obviously, it is not preferred practice to give point estimates of probabilities as they exclude the effect of the size of the testing data set. Therefore, it is more appropriate to give a window for the future performance of the classifier. The upper and lower boundaries for the probability of a certain future performance are calculated by the equation in [15], $p = \left(f + \frac{z^2}{2n} \pm z\sqrt{\frac{f}{n} - \frac{f^2}{n} + \frac{z^2}{4n^2}} \right) \Big/ \left(1 + \frac{z^2}{n} \right)$.

Note that f is our point estimate of the aforementioned probability, which is also the observed rate of the aforementioned performance, n is the number of instances in the testing data set, and z is the value satisfying the equation $probability[-z \leq X \leq z] = c$ for a normally distributed random variable X, where c is the specified confidence level. This approach is based on the assumption that we have n independent and identically distributed Bernoulli experiments each of which having an *observed* success rate of f. This is a sound assumption for our case. Employing this equation, along with the Leave-One-Out validation scheme results ($n = 298$), we get the window of [67.6%, 75.1%] for success probability, which means that we can be 85% sure ($z = 1.44$) that our proposed decision tree will correctly classify a new patient with the success ratio falling into this interval.

6. How to Assess Our Proposed Method

We need a benchmark to measure our proposed method against. We decided not to use the success statistics related to medical treatment decisions that were previously made by urologists at Haydarpaşa, for two reasons:

(a) There is no way to find out whether the patients who have undergone surgery would actually have benefited from medical treatment with alpha-blockers.

(b) There is a lot subjectivity involved in making this decision: Urologists at Haydarpaşa first use a combination of the IPSS and the QoL index to choose between applying medical treatment with alpha-blockers and proceeding with invasive surgery. These are subjective scores by construction. Second, they consult with the patient and get the patient's subjective assessment of the situation before finally making a decision.

Table 3
The performance of C4.5 with various feature sets and different minimum number of instances allowed in a leaf. The maximum element in each row is written in bold font.

Included Features	**Minimum Number of Instances Allowed in a Leaf**								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Relevant Feature Set (RFS)	64.1	64.1	64.1	**71.5**	70.1	69.5	64.8	64.8	64.8
RFS + Volume	60.8	60.8	60.8	60.8	61.1	61.4	**62.4**	**62.4**	61.4
RFS + Volume + QoL	**66.1**	65.8	60.1	63.1	59.7	60.7	62.4	62.8	60.1
RFS + Volume + QoL + Creatinine	**64.8**	64.1	58.7	57.1	54.7	56.0	57.7	57.7	55.7
RFS (Without Pruning)	57.1	57.1	57.1	65.4	**70.1**	69.5	64.8	64.8	64.8

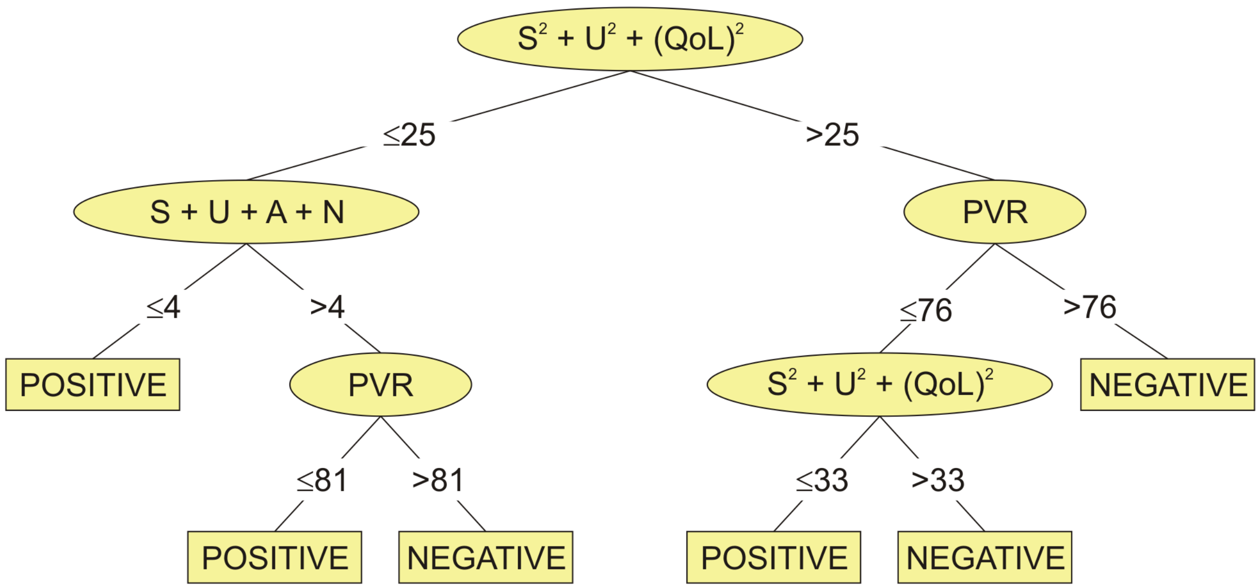

Figure 1. Our proposed decision tree classifier

Therefore, in order to make a fair comparison between the success of our proposed decision tree method and that of current practice at Haydarpaşa Numune Hospital, we compared our own method to two well-known Machine Learning methods, Fisher's Linear Discriminant Analysis (LDA) and Logistic Regression (LR) [15] each of which uses, as input, only the IPSS and/or the QoL scores. We chose these two variables since these are the only variables whose values have been clearly logged and used by the Haydarpaşa urologists when they make their decisions. In other words, patients' subjective assessments of their own situation were not logged even though they were part of the decision making process.

The machine learning results for LDA and LR came out to be almost identical. Using the Leave-One-Out validation method, we found the following success ratios from LDA. QoL alone: 63.2%, IPSS and QoL together: 62.4%, IPSS alone: 61.7%. Since these results come from the Leave-One-Out validation scheme (n = 298), we get the window of [59.1%, 67.1%], which means that we can be 85% sure (z = 1.44) that QoL alone, when used with LDA, will correctly classify a new patient with the success ratio falling into this interval. It also means that at the confidence level of 85%, our proposed method is superior to the benchmark since the lower end of the window of success rate of our decision tree is above the higher end of that of the benchmark.

7. Conclusions

Synthetic variables do enhance the performance of our classifier in this problem. Our decision tree, which uses PVR along with synthetic variables that were created using the Quality of Life index along with the subscores that comprise the IPSS, is statistically superior, at the 85% confidence level, in predicting the failure to respond to the medical treatment of BPH with α-blockers than when one uses either the IPSS, or the Quality of Life index, or both in well-known machine learning techniques such as Discriminant Analysis and Logistic Regression. This comparison is meaningful because urologists generally use IPSS and/or the Quality of Life index in making such decisions. Further research needs to be done in order to enhance the performance of the classifier that we propose here. That means, more efficient algorithms for generating and selecting synthetic variables will need to be utilized. This research may end up needing a high performance computing platform due to the computational complexity of the search for useful synthetic variables. Finally, the performance of our prototypical classifier indicates that failure to respond to medical treatment of BPH using α-blockers is predictable, to a certain degree, for new patients, prior to treatment.

8. References

[1] Hong SJ, Ko WJ, Kim SI, Chung BH. Identification of baseline clinical factors which predict medical treatment failure of benign prostatic hyperplasia: an observational cohort study. European Urology 2003;44(1):94-9.

[2] Thorpe A, Neal D. Benign Prostatic Hyperplasia. Lancet 2003; 361 (9366):1359-1367.

[3] Berry SJ, Coffey DS, Walsh PC, Ewing LL. The development of human benign prostatic hyperplasia with age. J. Urol 1984; 132(3):474–9.

[4] Garraway WM, Collins GN, Lee RJ. High prevalence of benign prostatic hypertrophy in the community. Lancet 1991; 338 (8765): 469–71.

[5] Hammarsten J, Hogstedt B, Holthuis N, Mellstrom D. Components of the metabolic syndrome-risk factors for the development of benign prostatic hyperplasia. Prostate Cancer Prostatic Dis 1998; 1:157–62.

[6] Djavan B, Nickel JC, Rosette J, Abrams P. The Urologist View of BPH Progression: Results of an International Survey. European Urology 2002; 41 (5): 490-496.

[7] Blute ML, Larson T. Minimally Invasive Therapies for Benign Prostatic Hyperplasia. Urology 2001; 58(6): 33-40.

[8] Barry MJ, Fowler FJ Jr, Bin L, Oesterling JE. A nationwide survey of practicing urologists: Current management of benign prostatic hyperplasia and clinically localized prostate cancer. J Urol. 1997; 158:488-492.

[9] Logan YT, Belgeri MT. Monotherapy versus combination drug therapy for the treatment of benign prostatic hyperplasia. The American Journal of Geriatric Pharmacotherapy 2005; 3(2): 103-114.

[10] Boyle P, Gould AL, Roehrborn CG. Prostate volume predicts outcome of treatment of benign prostatic hyperplasia with finasteride: Meta-analysis of randomized clinical trials. Urology 1996; 48(3):398-405.

[11] Mochtar CA, Kiemeney L.A.L.M., M.M. van Riemsdijk MM, Laguna MP, Debruyne FMJ, de la Rosette JJMCH. Post-Void Residual Urine Volume is Not a Good Predictor of the Need for Invasive Therapy Among Patients With Benign Prostatic Hyperplasia. The Journal of Urology 2006; 175(1):213-216.

[12] Roehrborn CG, McConnell J, Bonilla J, Rosenblatt S, Hudson, PB, Malek GH, Schellhammer PF, Bruskewitz R, Matsumoto AM, Harrison LH. Serum Prostate Specific Antigen is a strong predictor of future prostate growth in men with Benign Prostatic Hyperplasia. The Journal of Urology 2000; 163(1): 13-20.

[13] Fernandes AW, Amonkar MM, Madhavan SS, Nseyo UO. Assessment of urologists' practice preferences for the management of benign prostatic hyperplasia in the veterans administration program. Current Therapeutic Research 2001; 62 (2): 129-141.

[14] Quinlan, JR. C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.

[15] Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. 2nd Ed., Morgan Kaufmann, San Francisco, 2005.

[16] Homma Y, Yamaguchi T, Kondo Y, Horie S, Takahashi S, Kitamura T. Significance of Nocturia in the International Prostate Symptom Score for Benign Prostatic Hyperplasia. Journal of Urology; 167(1):172.

[17] Ritthoff O, Klinkenberg R, Fischer S, Mierswa I. A Hybrid Approach to Feature Selection and Generation Usingan Evolutionary Algorithm. No. CI-127/02, Collaborative Research Center 531, University of Dortmund, Dortmund, Germany, 2002.

[18] Hall, MA, Correlation-based feature selection for machine learning. Ph.D. Thesis. Department of Computer Science, University ofWaikato, 1999.

[19] King G, Honaker J, Joseph A, Scheve K. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation, 2001. American Political Science Review; 95(1):49-69.

[20] Honaker J, Joseph A, King G, Scheve K, Singh N. Amelia: A Program for Missing Data (Windows version), Harvard University, Cambridge, MA, 2001. Available at http://GKing.Harvard.edu/.

[21] WEKA 3-2-3, 2000. http://www.cs.waikato.ac.nz/ml/weka/.

[22] Roehrborn CG, Definition of at-risk patients: baseline variables, April 2006; BJU International, 97 (s2):7.