Evolved Bayesian Networks as a Versatile Alternative to Partin Tables for Prostate Cancer Management

Ratiba Kabli
School of Computing
The Robert Gordon University
Aberdeen, United Kingdom
rk@comp.rgu.ac.uk

John McCall
School of Computing
The Robert Gordon University
Aberdeen, United Kingdom
jm@comp.rgu.ac.uk

Frank Herrmann
School of Computing
The Robert Gordon University
Aberdeen, United Kingdom
fh@comp.rgu.ac.uk

Eng Ong
Department of Urology
Aberdeen Royal Infirmary
Aberdeen, United Kingdom
engong@nhs.net

ABSTRACT

In this paper, we report on work done evolving Bayesian Networks with Genetic Algorithms. We use a Chain Model GA [19] to induce a Bayesian network model for the real world problem of Prostate Cancer management. Bayesian networks can and have been used in a wide range of complex domains, notably in medicine. In fact, they have shown powerful capabilities in representing and dealing with the uncertainties generally inherent in the clinical practice. In this study, we investigate those capabilities by testing the evolved model's predictive power and exploring its potential use as a more versatile alternative to the widely used Partin tables for prostate cancer pathology staging.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search - *heuristic methods*; I.2.6 [**Artificial Intelligence**]: Learning; I.2.1 [**Artificial Intelligence**]: Applications, Medicine and science

General Terms

Algorithms, Performance, Experimentation

Keywords

Genetic Algorithms, Bayesian Networks, Greedy Search, Medical Decision Support, Real-World Applications

1. INTRODUCTION

Prostate cancer is the second most common cause of cancer death in men, after lung cancer. It is the most common cancer among men in the United Kingdom [3]. Although little is known about its direct causes, a set of risk factors and associations have been empirically observed. These include patients' serum Prostate Specific Antigen (PSA) level, digital rectal examination (DRE) finding, age, race, family history, health status etc. Patients may or may not present with any symptoms. The symptoms vary from some lower urinary tract symptoms to back pain and anemia in advanced cases. Some of these may be associated with benign conditions e.g. benign prostatic hyperplasia or prostatitis instead of prostate cancer. The same applies to serum PSA level which may be elevated in these benign conditions. Lack of a specific test may result in unnecessary investigations including invasive procedures e.g. prostate biopsy associated with possible complications. Diagnosed patients usually undergo further staging investigations to decide on the most appropriate treatment based on current available evidence. All treatments are associated with possible side effects. For simplification, the four main types of treatment include watchful waiting, hormone therapy, radiotherapy and surgery. The clinicians have to partition and traverse a vast space of parameters e.g. PSA level, DRE finding, symptoms, age, general health status etc to decide on the most appropriate management plan for each individual patient. In some cases, this is unclear and may result in over treatment. It is this uncertainty surrounding the disease that makes the prostate cancer an attractive area for AI medical decision support.

In this paper, we are interested in the use of evolutionary algorithms in order to evolve Bayesian Networks for assisting some parts of this medical decision making process for prostate cancer management. In particular, we investigate the potential of these evolved networks in improving on Partin tables or nomograms, currently used to assist urologists in predicting the extent to which the disease has progressed. In the next section we will give more details on the main indicators considered for prostate cancer management and explore the patient journey. In Section 3, we introduce the Bayesian networks formalism and our evolutionary algorithm technique for inducing these networks from the prostate patient data. We also describe how we make use of this technique for prostate cancer management. Section 4 describes our experiments with patient data and presents
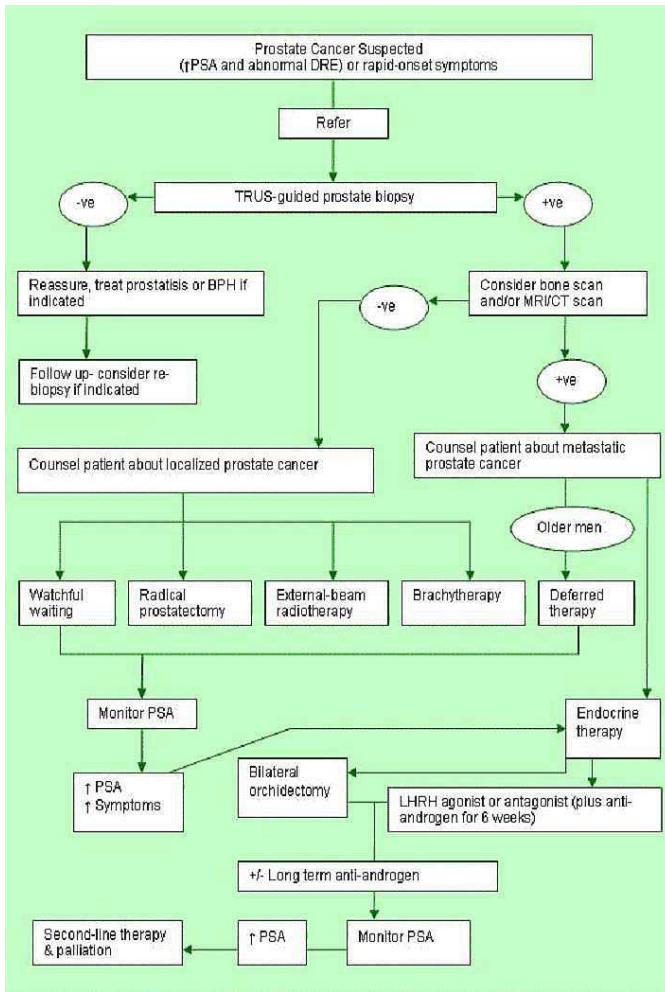
Figure 1: Prostate Cancer Patient Journey *(FAST FACTS: Prostate Cancer 3rd Ed.)*

our results. Implications of these results and possible applications are discussed in Section 5. In Section 6 we conclude with a brief outline of future work.

2. PROSTATE CANCER PATIENT JOURNEY

Generally a prostate cancer patient goes through a long disease journey filled with uncertainties. A general view of this journey is depicted in Figure 1. Recovery depends mainly on the stage of the disease and the health status of the patient. Prostate cancer is often slow growing and therefore patients may die of old age before the cancer has even spread. However, early diagnosis and an appropriate management plan taking account of the stage of the disease and other factors including the health status of the patient are paramount to both survival and good quality of life with the disease.

There may be no symptoms in the early stages of the disease, therefore regular check ups are crucial. An elevated serum PSA level and an abnormal DRE finding are suspicious of prostate cancer. With these findings, patients usually undergo ultrasound scan guided prostate biopsy. The biopsy confirms the diagnosis and tumour grading which provide the expected biology aggressiveness of the disease and this

was shown to be an important prognostic factor of the disease. Gleason score is the most commonly used grading system in prostate cancer [20]. Based on these results, the patient may undergo further staging investigation e.g. MRI scan or isotope bone scan to establish the extent of the disease. This extent of the disease is then categorized according to the TNM classification of malignant tumours [28]. The individual case is then discussed at a multidisciplinary meeting in which decision on recommended treatment options is made based on many parameters. The patient is then counseled and a treatment option is finalized. It is important to choose the most appropriate treatment plan with proper follow up measures for each individual patient based on individual circumstances.

3. BAYESIAN NETWORKS FOR PROSTATE CANCER MANAGEMENT

The use of Bayesian networks in medicine flourished in the age of expert systems in the 1970s. The probability based nature of medical decision making is inherent in these networks and therefore fits well with the way medical practitioners work. An example Bayesian Network based expert system is the Pathfinder/Intellipath system which was developed at Stanford to provide assistance with identifying disorders from lymph-node tissue sections [17]. Although, one can find various work on the use of Bayesian networks for the diagnosis and prognosis of various pathologies including various cancers, work on prostate cancer tends to be centered toward using traditional statistical techniques or using artificial neural networks. On the latter, recent work was carried out at the John Hopkins Hospital by Crawford et al [14]. The research resulted in what is called the *Prostate Calculator* which, given some inputs on patient disease information, outputs some diagnosis and staging probabilities for the patient. Other applications of ANNs for prostate cancer can also be found in [11, 32]. In this paper we explore the potential use of Bayesian networks for prostate cancer management. The white box nature of these networks assisted by their simple graphical representational aspect, should give us a more intuitive insight into the disease and its management. First, we briefly review key ideas in Bayesian networks.

3.1 The Bayesian Network Formalism

Bayesian Networks (BN) are probabilistic models useful for reasoning with, or representing knowledge under uncertainty. Essentially a BN can be defined as a pair (G, P). Here, G is a Directed Acyclic Graph (DAG) $G = (V, E)$ with the vertices V as the nodes in the network. Each node represents a random variable X_i relevant to the problem domain and $P = P(X)$ is the joint probability distribution of those variables. The dependencies among these variables are represented by the the set of edges E in the underlying DAG providing the following useful factorization of $P(X)$:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \qquad (1)$$

with $Pa(X_i)$ as the set of parent nodes for node X_i. To exploit the power of Bayesian Networks in knowledge representation and inference, they first have to be constructed for the given domain. There are two parts to fully specifying a Bayesian network. We have to define both the

Directed Acyclic Graph (DAG) structure representing the network and the underlying conditional probability distribution. This can be done manually given an extensive knowledge or study of the problem domain, or automatically from problem domain data. Where the former approach can be very time consuming and not always possible especially for domains where a great deal of uncertainty exists, numerous algorithms have been developed to empirically induce the networks from data instead. These algorithms generally fall into two categories: Search and Score methods and Conditional Independence Testing methods. The latter is a constraint based approach which relies on a number of statistical tests to determine whether two variables are independent or dependent given a set of conditioned variables. Tests such as Pearson's Chi-Square and mutual information are often used. Work by de Campos [10] and Spirtes and Glymour's PC algorithm [29] illustrate this. This approach tends to give good results with sparse networks and small samples of data however it does not scale very well for large datasets and dense networks.

Learning Bayesian Networks is an NP-hard problem [4]. The number of possible Bayesian network structures for a given problem grows super exponentially given the number of variables in that problem. Robinson [27] quantifies the number as $O(n!2^{\binom{n}{2}})$ for a problem of size n. So where a 3 variable problem would have 25 possible networks, a 5 variable problem would have 29,281 and a 6 variable problem would have 3,781,503 possible networks. This makes exact methods for structure discovery impractical and seldom used without imposing a great deal of restrictions [21]. The search and score approach relies on approximate methods and involves searching through the space of possible network structures for one that best describes the data. An information criterion to measure the goodness and differentiate between candidate structures met while traversing the search space is employed. The goal is to maximize this information measure or score by moving from one structure to another by means of some local variation such as a deletion or an addition of a link between two nodes and then evaluating the overall effect of the move. After a number of iterations, an optimal score is found and the associated network is then chosen to represent and explain the data. Examples of work done in both score functions and search algorithms used for this purpose can be found in [1, 2, 6–8, 13, 18].

3.2 Genetic Algorithms for Learning the Bayesian Network model

Staying within the scope of the search and score approach and in order to avoid getting stuck in local optima and explore the networks' search space better, work has been carried out looking into methods which consider a group of network structures at a time rather than a single structure. A scoring function is used to evaluate the current group (or population of structures) from which a new and better population is created and then evaluated and so on until a set stopping criteria is reached. A range of such algorithms have been proposed. Mainly they are based on evolutionary algorithms such as genetic programming [31] and genetic algorithms [16, 22, 30].

In this work, we propose using the Chain-Model Genetic Algorithm (chainGA) described in [19]. This algorithm has shown promising results in terms of speed and efficiency as well as suitability for decision making diagnostic networks.
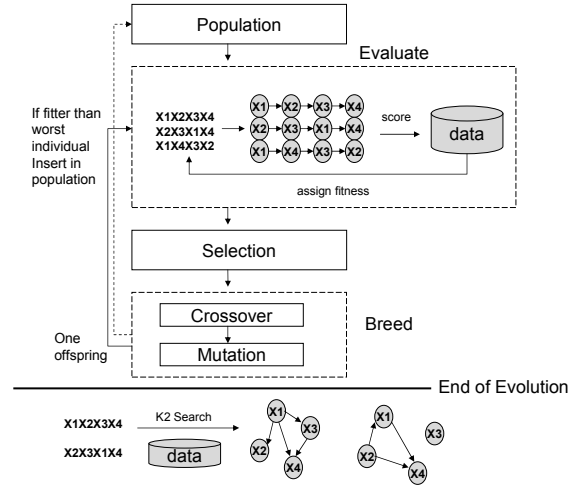

Figure 2: Chain Model GA for Learning Bayesian Networks

chainGA works by evolving a population of Bayesian network topological node orderings. At each evaluation step, a chain structure[1] of the given ordering is constructed and evaluated using the Cooper and and Herskovitz metric [7]. The orderings then assume the fitness returned by their associated chains and are evolved, crossed over and mutated for a predefined number of generations. At the end of each run, the K2 greedy search algorithm [7] is run on a percentage of the best orderings found in order to search for the best network structure. It can be seen that the low resolution evaluation of chain structures acts as a pre-selection phase where orderings with inferior scores are rejected and those with promising scores are preserved for breeding.

The greedy search K2 algorithm used here was proposed by Cooper and Herskovitz [7]. The algorithm assumes that a priori, all structures are equally likely and that cases in the data occur independently and are complete. Moreover, it assumes the presence of a node ordering and imposes a maximum number of parents a node can have (inbound edges). With these conditions satisfied, K2 starts with an empty ancestor set for each node and incrementally adds links that maximize the score of the resulting structure. The algorithm stops when no more ancestor node additions improve the score. K2 was originally used along the CH score which captures the probability of a candidate network structure B_s given a set of data D. Formally the discrete probability $P(B_s, D)$ is given by

$$P(B_s, D) = P(B_s) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (2)$$

Where q_i denotes the number of possible different instances the parent of variable X_i can take. r_i is the number of values

---

[1]The chain structure here is defined by the simple network with an outgoing edge from each node to its immediate successor in the ordering.

X_i has, N_{ijk} denotes the number of cases in the dataset D in which X_i takes value k of its x_i instance when its parent Pa_i has its jth value. N_{ij} is the sum of all N_{ijk} for all values x_i can take. The use of the chain structures reduces computation time since the number of links to evaluate is fixed in contrast to K2. The best scoring Bayesian network is chosen to be the optimal model to represent the data in hand. The chainGA framework is illustrated in Figure 2.

3.3 The Data

For this work, we build our model from data which was collected in collaboration with the Aberdeen Royal Infirmary (ARI), Scotland. A cohort of 320 patient cases was assembled. The collection consists of retrospective data of patients diagnosed and treated for prostate cancer in the ARI over a period of two and a half years 2002-2004. The patient records include data that depict the different stages of the disease management process. It varies in its nature and is both qualitative and quantitative. The qualitative data includes most of the patient personal information such as their occupation, location, as well as the additional information in form of free text, such as further comments on past medical history, on the results of the scans and of the patient's diet. It also includes Yes and No data such as presence of family history of the disease, a history of respiratory disease, etc. The quantitative data available is present in different formats: discrete values, such as age and the International Prostate Symptoms Score (IPSS); continuous values such PSA level and prostate volume; percentages such as percentage cancer cells present in biopsied tissues, categorical and ordinal data such as the DRE result and treatment options. Table 1 illustrates some examples. In this study, we choose to learn our network model over the discretized domain and therefore the continuous variables utilized are descritized prior to use. Although this might limit the preciseness in capturing the characteristics of the distribution of our continuous data, it allows us to learn a model that can be efficiently used for inference and optimal decision making [12]. Moreover, in the case of our medical problem at hand, the discretization process carried on some variables such as PSA values follows an intuitive and problem-specific discretization provided by the medical experts or domain literature. Some variables such as age, Gleason score, and others were also banded into suitable categories reflective of the expert knowledge of the disease.

4. EXPERIMENTAL RESULTS

Following the steps of the chainGA algorithm described above, we build our Bayesian network model that represents the data we have collected so far. For the purpose of this work, we only include a selection of the factors collected for each patient. In this case, 37 variables or factors are taken into consideration. These are illustrated in Table 2. The choice of factors was made on the basis of importance and on the extent of the actual amount of data available. For some prostate cancer patients, certain information is deemed sensitive or is for other reasons not collected. The 37 variables chosen span the collected patient database and characterize each stage of the patient journey from diagnosis to treatment and aftercare.
Our chainGA algorithm implementation was run 50 times with 100 generations containing a population of 10 network individuals at each time. Mutation and crossover rates were

0.1 and 0.9 respectively. The best scoring resulting network was then chosen as the optimal model for our problem at hand. Figure 3 illustrates this Bayesian network model (Ge-NIe snapshot [15]). As described earlier, the nodes in the network represent the factors in our problem and the links or edges between these nodes represent the relationships and dependencies between these variables. At first glance, one can already see some intuitive relationships formed in the model. For example, there is strong interaction between the values of PSA, DRE and Gleason score nodes and in turn the PSA node has an interaction with the node representing the decision to SCAN. In a Bayesian network, the direction of the arrows indicates the direction of variable dependency in the conditional factorisation but not necessarily causality. For example, the decision to SCAN is temporally dependent on PSA and DRE tests. In fitting Bayesian networks, there will always exist equivalent networks that have the same score while reversing some arrows. Such equivalent networks represent the same information regardless of the direction of the arrows. An account of equivalence classes in Bayesian networks can be found in [5, 26].

5. INFERENCE: APPLICATION AND USES OF THE BAYESIAN NETWORK MODEL

Once the structure for the Bayesian network is discovered and the probability distribution associated with it identified, the model can be used extensively not only for describing the existing relationships between the variables in our problem environment, but also for inferential exploration of any undetermined or hidden relationships among these variables. The idea is to update the probabilities of outcomes based on the relationships in the model and the evidence known about the situation at hand. Evidence about recent events or observations is applied to the model by instantiating or clamping a variable to a state that is consistent with the observation. The propagation of that evidence is then performed to update the probabilities of all the other variables that are connected to the variable representing the new evidence. After the inference, the updated probabilities reflect the new probabilities of all possible outcomes coded in the model. So, for prostate management in this case, with such a model, we would be able to not only answer a whole range of questions about the patient and his disease but also to get a very useful insight into a prostate cancer patient journey therefore assisting us in making decisions throughout this journey from diagnosis to treatment and aftercare. Investigation with our collaborators from the ARI has highlighted a range of important questions medical practitioners can address using this model, which we discuss in the next sub-sections.

5.1 Diagnosis and Early detection

Patient age and PSA levels are generally the first indicators medicals expert have to analyze to determine whether a patient has prostate cancer or not. However the PSA test suffers from false positives and therefore other information is needed to consolidate this result. To use the learnt model for diagnosis, a *hasCancer?* node can easily be introduced in the network where every known piece of information on the patient would help us predict the probability of the patient having cancer or not. The proliferation and availability of patient information stored nowadays also means more peripheral factors not used before such as patient diet for

Data Item and example	
Qualitative	Occupation: Fisherman
	Other Medical History: Renal failure
Quantitative	Any Family History of PC: Yes
	Age: 78
	PSA: 123.4
	Gleason Score: 3+4
	Tissue Involved: 30%
Other data (Categorical/ordinal)	DRE: Suspicious
	Tumour(TNM): T1a
	Recommended Treatment: Radical prostatectomy

Table 1: Nature of Data Available

Patient Age	Family History
PHS (Patient Health Score)	DRE(Digital Rectal Examination)
PSA	IPSS
PSAD (PSA Density)	Gleason Score
Scan (MRI- CT or Bone scan)	Tumour (none, bilateral, right etc.)
Metastasis (yes or no)	Staging Tumour (of the TNM system)
Staging Node	Staging Metastasis
Recommended Treatments	Treatment received by patient
PSA Progression (After treatment)	Death?
Prostatectomy (Retropubic or perineal)	
Pathology Tumour (after prostatectomy)	Pathology Node (after prostatectomy)
Major Diseases:	
Arrhythmias	Arthritis
Cerebral Vascular Disease	Congestive Heart Failure
Diabetes Mellitus	Gastrointestinal Disease
Hepatobililiary Disease	Hypertension
Ischemic Heart Disease	Malignancy
Myocardial Infraction	Other Heart Disease
Other Significant Disease	Peripheral Vascular Disease
Respiratory Disease	Other Related Disease

Table 2: Selection of factors used

instance can be modeled in the Bayesian network for diagnostic purposes. In the case of the model illustrated in this paper, diagnosis is somewhat not possible as the data used to learn our model is retrospective and that of patients already diagnosed with prostate cancer. Consolidating the data collection with patients where cancer was not found, the model can be easily adapted to be used for diagnosis.

5.2   Scanning and Biopsy Decision

Part of the patient journey involves invasive procedures (e.g. prostate biopsy) which can be associated with possible complications, expensive and create discomfort for the patient. The Bayesian model provides us with accurate conditional probabilities based on each individual patient's data which will greatly assist in the decision to undergo any scanning or biopsy procedure.

5.3   Treatment Choice and Patient Quality of Life Post Treatment

As is for most cancers, deciding on the most appropriate treatment for an individual patient is not straightforward. A team of medical experts in various disciplines relating to prostate cancer gather to decide on an optimal treatment for prostate cancer patients. The decision is often a balance between how to best beat the disease without hindering the after care life of the patient. It is based on the disease

stage, general health state of the patient, his life expectancy, etc. Possible side effects of the treatment on the patient's quality of life are also considered. This can be seen from the network where the treatment node is influenced by various other factors including the disease stage of the patient, their age, their health state, etc. The Bayesian network model can be used to enhance decision-making at this stage of the patient journey.

5.4   Hospital Resources Planning

The model can constitute the basis for a robust and flexible model for assisting the medical team in predicting the journey a patient goes through from their personal and clinical data. An audit of retrospective patient data can therefore help the clinicians as well as hospital managers to forecast their needs for biopsies, scans and other treatment equipment in order to make provision for their prospective patients.

5.5   Patient Disease Education

Prostate cancer is a serious disease and every decision made from diagnosis to aftercare can affect the patient's quality of life. It is therefore very important for the patient to be aware of every aspect of his disease as to work along his medical expert to manage the disease and prevent any unnecessary aggravation or discomfort. A tool based on the
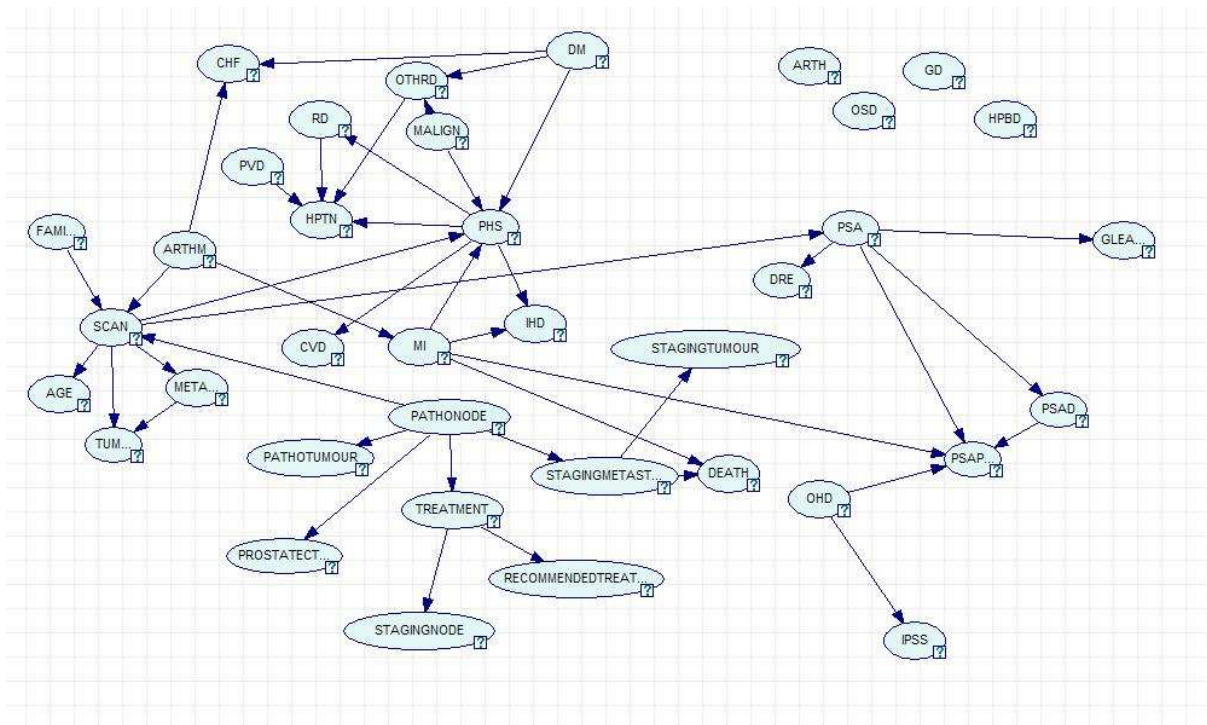
Figure 3: The Bayesian Network Model

Bayesian network model and endorsed with prostate cancer information could be developed to assist the patients by giving them an intuitive interface with their disease, where they can explain what their PSA level, or Gleason Score etc. mean for them.
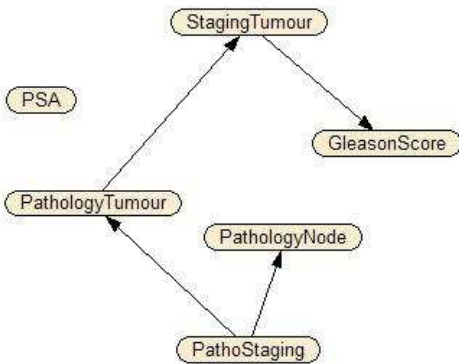

Figure 4: Pathology Staging Bayesian Network

5.6   Pathological Staging : Partin Tables

Another area highlighted by the clinicians was prediction of the final pathological staging of the disease. The Partin tables are one technique which is used for this purpose. It was originally developed by Partin et al [24,25] at the Brady Institute of Urology at the John Hopkins University, USA. The original study examined data from 703 patients with clinically localized disease undergoing radical prostatectomy between 1982 and 1991. The study evaluated the utility of
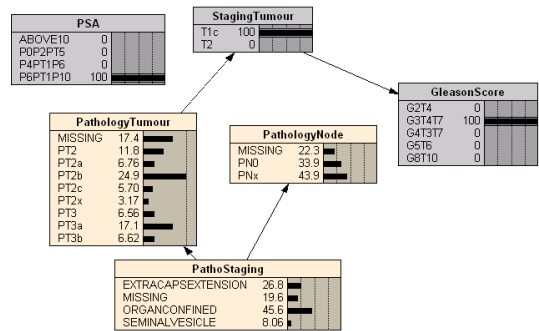

Figure 5: Pathology Staging after Evidence (visualized in Netica)

Figure 6: Partin Table for Clinical Stage T1c

logistic regression analysis combining PSA, Gleason score, and clinical tumour stage as a predictor of the final pathological staging. The results of the study proved the hypothesis that the combination of the indicators gave better prediction than any singly used indicator. These results were then validated with a bigger patient dataset in a multi-institutional study. Updated again in 2001, to incorporate a larger set of 5079 patients, the Partin tables or nomograms are nowadays the most popular and widespread predictive tool for prostate cancer pathological staging used all over the world. Doctors can use these nomograms with PSA, Gleason score and estimated clinical staging information to determine a representative probability of the disease extent i.e. organ-confined, extraprostatic extension, seminal vesicle invasion and pelvic lymph node invasion. This is important to counsel patients in deciding the most appropriate course of treatment and management of the disease. Although the Partin tables were based on results obtained from a highly regarded institution, some concerns with regards to their validity when applied to a new and heterogeneous dataset still remain. This is particularly important when applying them to a patient set from outside the USA where the ethnic mix, and other social, dietary and environmental factors might be different. Several studies have therefore been carried out to validate the tables further. Results from these investigations vary in their confirmation of the good performance of the tables on one hand and the conclusion as to the limitation of the information used in these nomograms when compared to the proliferation of measured patient data that could be used nowadays on the other hand. Research has indeed shown that there is room for improvement as the Partin tables do not provide information beyond pathologic stage and use a very limited set of prostate cancer features. A study by Crawford et al. [9] has also confirmed this result by investigating the use of Artificial Neural Networks for prostate cancer management. This study and a similar one carried out by Djavan [11] have shown that the inclusion of more input variables result in a higher accuracy. This however, is also claimed to result from the use of ANNs proving a promising technique compared to linear regression for the problem at hand. Nevertheless, the black-box effect of ANNs makes the understanding of the resulting model highly complex and therefore difficult to use.

We propose an intuitive alternative with Bayesian networks which offer a sound probabilistic model based on concepts such as causality and inference; well suited to the medical practice.

5.7 Bayesian Networks and Partin Tables

In this section, we investigate the case for using a Bayesian network model as a flexible alternative to Partin tables for predicting prostate cancer pathological staging. One approach is re-doing the same exercise of building the previous Bayesian network model, but this time the model is induced from a reduced dataset focusing only patients having undergone surgery and only on the variables used to generate the Partin tables, notably: PSA, Gleason Score and Clinical Staging information. This results in the network illustrated in Figure 4. The Pathology Staging node was added in order to simplify the reading of the pathology staging outcomes described by the Pathology Tumour node. From our clinicians, we know that a result up to PT2 means the disease is organ confined. PT3a mean extra-capsular extension and PT3b means the cancer invades the seminal vesicle. For

lymph node involvement, we look at the pathological node variable, with PN1 meaning lymph node involvement and PN0 the opposite. We should note that some low risk early prostate cancer patient did not have lymph node removed and therefore the regional lymph node could not be assessed (i.e. PNx). Consequently, in this study, an assumption was made that PNx is equivalent to PN0. Figure 5 displays the compiled network after evidence from a patient case with PSA value of 7, clinical tumour stage of T1c and a Gleason score of 3+4. The outcome predictions from the appropriate Partin table, depicted in Figure 6 are 54% for organ-confined disease, extraprostatic extension at 36%, seminal vesicle 8%, and lymph node involvement 2%. The predicted staging probabilities from the Bayesian network model as we can see in Figure 5 are 45.6% organ-confined disease, extraprostatic extension at 26.8% and seminal vesicle 8.06%. The Pathology Node shows a 77.8% of negative lymph node involvement (PNx=PN0). The inference was done using the Netica tool for Bayesian networks [23]. As we can see, the resulting probabilites from the model are specific to the small sample used to generate the Bayesian network (only 46 prostatectomy patient cases) and the institution the data was collected from. Moreover, the patient mix the Partin tables are based on is quite different to the mix in our cohort. The advantage and flexibility of Bayesian networks in this case, is that we can make the predictions more precise and more powerful by using the general model built for prostate cancer mangement in Figure 3 to infer pathology staging as well instead of building separate models for each application. Since the model built earlier includes the factors we are interested in; PSA, Gleason Score and the clinical tumour stage, then we could freely use the same network for this task too. As a consequence we also profit from the other patient and disease factors that could influence the staging decision, or moreover predict those factors from available pathology staging evidence to be used for explanation purposes for instance.

6. CONCLUSION AND FUTURE WORK

In this paper we explored the versatile use of Bayesian network for prostate cancer management in general and as a Partin tables alternative in particular. We have built a Bayesian network model to represent prostate cancer patient data from the ARI, using a chain-model Genetic Algorithm based on node orderings. The resulting model could be used to answer various queries relating to prostate cancer management from diagnosis to treatment decision making and pathology staging. For the latter purpose, we compared some results to Partin tables and saw promising results in spite of the small and limited data sample used. The discrepancies in the model predictions can also be attributed to the different patient sample used, Scottish patients in this case. The potential of the use of Bayesian networks in this case is to support decision making in a more intuitive and population-centered approach to Partin tables. Models can be developed to suit the patient mix and the medical institution at hand. As a future step, one would need to explore this by including more patient data in the model as to do a large scale comparison. Furthermore, the predictive potential for each of the prostate cancer applications of the Bayesian network model proposed in the paper can be investigated in its own right.

7. REFERENCES

[1] R. R. Bouckaert. Probabilistic network construction using the minimum description length principle. *Lecture Notes in Computer Science*, 747:41–48, 1993.

[2] W. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 2:159–225, 1994.

[3] CancerResearchUK. http://www.cancerresearchuk.org.

[4] D. Chickering, D. Heckerman, and C. Meek. Large-sample learning of Bayesian networks is NP-Hard. *J. Mach. Learn. Res.*, 5:1287–1330, 2004.

[5] D. M. Chickering. Learning equivalence classes of Bayesian-network structures. *J. Mach. Learn. Res.*, 2:445–498, February 2002.

[6] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE transactions on Information Theory*, 14:462–467, 1968.

[7] G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[8] R. G. Cowell, S. L. Lauritzen, A. P. David, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1999.

[9] E. Crawford, E. Gamito, C. O'Donnell, A. Errejon, D. Raben, M. Han, A. Partin, and A. Tewari. Artificial neural network model to predict risk of non-organ-confined disease and risk of lymph node spread in men with clinically localized prostate cancer. *Journal of Urology*, pages 165–233, 2001.

[10] L. de Campos and J. Huete. On the use of independence relationships for learning simplified belief networks. *International Journal of Intelligent Systems*, 12:495–522, 1997.

[11] B. Djavan, M. Remzi, A. Zlotta, C. Seitz, P. Snow, and M. Marberger. Novel artificial neural network for early detection of prostate cancer. *Journal of Clinical Oncology*, 20(4):921–929, February 2002.

[12] N. Friedman and M. Goldszmidt. Discretizing continuous attributes while learning Bayesian networks. In *Proceedings of the International Conference on Machine Learning*, pages 157–165, 1996.

[13] N. Friedman and M. Goldszmidt. Learning Bayesian networks with local structure. In *Proceedings of the 12th Conference on Uncertainty in AI*, pages 252–262, 1996.

[14] E. J. Gamito, E. D. Crawford, and A. Errejon. *Artificial neural networks for predictive modeling in prostate cancer.*, chapter Handbook of Prostate Cancer: Biology, Epidem. and Therapeutic Modalities. 2002.

[15] GeNIe. GeNIe structural modelling tool http://genie.sis.pitt.edu/.

[16] J. Habrant. Structure learning of Bayesian networks from databases by genetic algorithms-application to time series prediction in finance. In *ICEIS*, pages 225–231, 1999.

[17] D. Heckerman. An empirical comparison of three inference methods. In *Proceedings of the 4th Conference on Uncertainty in AI*, pages 283–302, 1990.

[18] D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. In *KDD Workshop*, pages 85–96, 1994.

[19] R. Kabli, F. Herrmann, and J. McCall. A chain-model genetic algorithm for bayesian network structure learning. In *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 1264–1271, New York, NY, USA, 2007. ACM.

[20] R. S. Kirby, M. K. Brawer, and L. J. Denis. *Fast Facts: Prostate Cancer*. Health Press, third edition, 2001.

[21] M. Koivisto and K. Sood. Exact Bayesian structure discovery in Bayesian networks. *J. Mach. Learn. Res.*, 5:549–573, 2004.

[22] P. Larrañaga, C. Kuijpers, and R. Murga. Learning Bayesian network structures by searching for the best ordering with genetic algorithms. *IEEE Transactions on System, Man and Cybernetics*, 26:487–493, 1996.

[23] Netica. Netica Bayesian network software from Norsys http://www.norsys.com.

[24] A. W. Partin, M. Kattan, E. Subong, P. Walsh, K. Wojno, J. Oesterling, P. Scardino, and J. Pearson. Combination of prostate-specific antigen, clinical stage, and gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *Jama*, (277):1445–51, 1997.

[25] A. W. Partin, J. Yoo, H. B. Carter, J. D. Pearson, D. W. Chan, J. I. Epstein, and P. C. Walsh. The use of prostate specific antigen, clinical stage and gleason score to predict pathological stage in men with localized prostate cancer. *Journal of Urology*, (150):110–4, 1993.

[26] J. Pearl and T. Verma. Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in AI*, pages 220–227, 1990.

[27] R. Robinson. Counting labeled acyclic digraphs. *New Directions in the Theory of Graphs*, pages 239–273, 1973.

[28] L. H. Sobin and C. Wittekind. *TNM classification of malignant tumours*. Wiley-Liss, 6th edition edition, 2002.

[29] P. Spirtes, C. Glymour, and R. Scheines. Causation, Prediction and Search. *Lecture Notes in Statistics, New York: Springer Verlag*, 81, 1993.

[30] S. van Dijk, D. Thierens, and L. C. van der Gaag. Building a GA from design principles for learning Bayesian networks. In *GECCO'03*, pages 886–897, 2003.

[31] M. L. Wong, S. Y. Lee, and K. S. Leung. A hybrid data mining approach to discover Bayesian networks using evolutionary programming. In *GECCO '02: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 214–222, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[32] H. Zhang, Z. Zhang, and A. Partin. Neural network based systems for prostate cancer stage prediction. In *Proceedings of the IEEE-INNS-ENNS Conference*, 2000.