Multicenter prospective testing to predict malignancy in adnexal masses using Bayesian network models

O Gevaert[1,*], C Van Holsbeke[2,3], R Fruscio[4], S Guerriero[5], A Czekierdowski[6], L Savelli[7], A Testa[8], D Fischerova[9], D Jurkovic[10], T Bourne[2,11], P Neven[2], L Valentin[12], B De Moor[1], D Timmerman[2]

1 Dept Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, Leuven, Belgium; 2 University Hospitals, Leuven, Belgium; 3 Ziekenhuis Oost-Limburg, Genk, Belgium; 4 San Gerardo Hospital, Monza, Italy; 5 Ospedale San Giovanni di Dio, Cagliari, Italy; 6 Medical University, Lublin, Poland; 7 Reproductive Medicine Unit, Bologna, Italy; 8 Università Cattolica del Sacro Cuore, Rome, Italy; 9 General Faculty Hospital of Charles University, Prague, Czech Republic; 10 King's College Hospital, London, UK; 11 St Georges Hospital Medical School, London, UK; 12 University Hospital, Malmö, Sweden;

Short title: IOTA 2: Bayesian models
Keywords: ovarian cancer, prospective performance, IOTA, Bayesian network, ultrasound

* Corresponding author:
Olivier Gevaert
K.U.Leuven
Dept. Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10
3001 Leuven
Belgium
+32 16 328646
+32 16 321970
olivier.gevaert@esat.kuleuven.be

Three referees:
-
-
-

Abstract (250 words max)

Objectives

To investigate prospectively the performance of Bayesian networks built using data from the International Ovarian Tumour Analysis (IOTA) study phase 1.

Methods

The first phase of IOTA resulted in a data set of 1066 patients from 9 centers in 5 countries. The data were randomly stratified using 70% of patient data (i.e. the training set) to construct a Bayesian network and 30% of the patient data to estimate the generalization performance (i.e. the test set). After building a Bayesian network model (referred to as BN1) on the training set, this resulted in an Area Under the ROC curve (AUC) of 0.942 (SE 0.017). Subsequently, new data were gathered in IOTA phase 2. This data set was used to estimate the prospective performance of the BN1 model.

Results

In total 1940 patients from 19 centres, including old and new centres from 8 countries, were recruited. The BN1 model has an AUC of 0.940 (SE 0.006) on all IOTA phase 2 data. After splitting the IOTA 2 data into data from the old centres and new centres, the AUC was 0.939 (SE 0.008) and 0.940 (SE 0.009) respectively.

Conclusions

The prospective performance of the BN1 model is similar to the initial performance on the IOTA phase 1 data and thus BN1 is able to generalize to new data. Analysis of the performance observed in old and new centers, showed almost no difference in the prospective performance of new and old centers, indicating that BN1 can be used in other centers in addition to the ones that were used to train the model.

Introduction

The International Ovarian Tumour Analysis consortium (IOTA) is a multi-centre study on the pre-operative characterisation of ovarian tumours based on artificial intelligence methods. The ultimate goal is to be able to distinguish benign and malignant adnexal masses. To accomplish this data was gathered consisting of more than 50 parameters based on transvaginal ultrasound and colour Doppler imaging in a standardized way (Ref White journal). The outcome measure was the histological classification of excised tissue as malignant or benign. Phase 1 of the IOTA study resulted in 1066 patients which were gathered from 9 centres in 5 countries. Based on this data set logistic regression, Least Squares support vector machines, and also Bayesian networks were used as models to distinguish benign from malignant adnexal masses. Each model has its own advantages/disadvantages.

Recently, phase 2 was finished resulting in 1940 patients from 19 centres. This includes the old centra which were also present in the first phase of the IOTA study. Besides these, the IOTA phase 2 study also contained ... new centra. The IOTA phase 2 data set contained ?? patients from old centres and .. patients from new centres (...%). Table ?? lists the old and new centra and their contribution to IOTA phase 2.

Here, we report on the prospective performance of Bayesian networks. Bayesian networks provide a more complex framework than for example logistic regression and allow arbitrary relations between all the variables (HR paper en andere klinishe paper met BNs). This allows modeling non linear relations in the data. Moreover a Bayesian network can be interpreted . Prospective validation allows to assess the reliability of this model in clinical practice. This validation will be investigated both on new data from the old centres and on data from new centres. This will allow to systematically evaluate if a Bayesian network model is able to distinguish benign from malignant masses.

Methods

Data

For both IOTA phase 1 and IOTA phase 2 data on more than 50 parameters was gathered. This includes medical history (personal and family history of breast and ovarian cancer, age, etc.), serum tumour marker CA125, colour Doppler imaging and blood flow indices (PSV, RI,

TAMX and PI), ultrasonography, morphology and echogenicity. We used for patients with masses in two ovaries the dominant tumour. Furthermore patients who were pregnant and patients with missing values were excluded from the analysis. Phase 1 IOTA data has been previously reported (ref JCO). Briefly, the phase 1 IOTA data set contained 1066 patients from 9 centres in 5 countries of which 266 (25%) were malignant. In phase 2, 1940 patients were gathered from 19 centres, including old and new centres. ???? patients had a malignant mass (???%). Table ?? contains patient characteristics for IOTA phase 2. Based on results from previous studies and with the help of a medical expert 35 variables were selected as relevant to the domain. We used discrete Bayesian networks therefore continuous variables were discretized into bins according to a gynaecological expert.

Comment [OG1]: Update met uiteindelijke aantal

Bayesian networks

A Bayesian network is a probabilistic model (Pearl 1988) that consists of two parts: a dependency structure, encoding the (in)dependencies and local probability models encoding the detailed relation between the variables. Usually a variable only depends on a few other variables which is reflected by a sparse graph. Figure 1 shows a toy example of a Bayesian network that models an ovarian mass.

The second part of this model, the local probability models, specifies how the variables depend on their parents. We used discrete-valued Bayesian networks, which means that these local probability models can be represented with conditional probability tables (CPTs). Such a table specifies the probability that a variable takes a certain value given the value of its parents. Figure 1 shows an example of a small Bayesian network with three variables and the corresponding CPTs.

Learning Bayesian networks

Learning discrete-valued Bayesian networks from data proceeds in two steps. First the dependency structure that best explains the data is constructed. This is done using a heuristic search strategy combined with a scoring metric (Heckerman et al. 1994, Cooper and Herskovits 1992). The scoring metric reflects the fit of the structure to the data, possibly combined with prior knowledge. The best dependency structure is found by iteratively making all small changes to the current dependency structure and adopting the change with the highest score. Because of the greedy nature of this search strategy, it is likely to come up with a sub-optimal model and therefore the model building process is repeated several times with different initialization parameters. Then the model with the highest score is selected. The second step consists of estimating the parameters of the local probability models for the selected model. This amounts to filling in a CPT for every variable and every possible value of its parents using the data.

Model building and prospective testing

Figure 1 shows graphically the model building process. First the IOTA phase 1 data set was split in a training set and a test set. The training set was used to build a model and the test set was used to estimate the classification performance. IOTA phase 2 data was used to determine prospective performance in both old and new centres.

Results

After building a Bayesian network model on the training set (referred to as BN1), this resulted in an Area Under the ROC curve (AUC) of 0.942 (SE 0.017) on the test set. On the complete IOTA phase 2 data, the BN1 model has an AUC of 0.940 (SE 0.006). After splitting the IOTA 2 data into data from the old centers and new centers, the AUC was 0.939 (SE 0.008) and

0.940 (SE 0.009), respectively. Figure 2 shows the Bayesian network that was built using only the training data set from IOTA phase 1.
Table xx shows the performance for each centre separately.

Discussion

The prospective performance reported here increases our performance that models based on IOTA phase 1 data generalize well to new data. Moreover new means here both new data in the same centres as the phase 1 study as well as completely new centres.

Acknowledgments

OG is supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Research further supported by research Council KUL: GOA AMBioRICS, CoE EF/05/007 SymBioSys, PROMETA, several PhD/postdoc & fellow grants. Flemish Government: FWO: PhD/postdoc grants, projects G.0241.04 (Functional Genomics), G.0499.04 (Statistics), G.0232.05 (Cardiovascular), G.0318.05 (subfunctionalization), G.0553.06 (VitamineD), G.0302.07 (SVM/Kernel), research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, GBOU-McKnow-E (Knowledge management algorithms), GBOU-ANA (biosensors), TAD-BioScope-IT, Silicos; SBO-BioFrame, SBO-MoKa, TBM-Endometriosis Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet, Bioinformatics and Modeling: from Genomes to Networks, 2007-2011) ; EU-RTD: ERNSI: European Research Network on System Identification; FP6-NoE Biopattern; FP6-IP e-Tumours, FP6-MC-EST Bioptrain, FP6-STREP Strok

References

Figures

Figure 1 workflow for prospective testing of Bayesian network models. AUC stands for Area Under the ROC curve. IOTA stands for International Ovarian Tumour Analysis.

