

Causal probabilistic modeling for malignancy grading in pathology with explanations of dependency to the related histological features

G. Weidl¹, J.R. Iglesias-Rozas² and N. Roehrl¹

¹Institute for Analysis, Dynamics and Modelling, University of Stuttgart, Stuttgart, Germany and

²Institut für Pathologie (Neuropathologie), Klinikum Stuttgart, Katharinenhospital. Stuttgart, Germany

Summary. This work demonstrates that histological grading of brain tumors and astrocytomas can be accurately predicted and causally explained with the help of causal probabilistic models, also known as Bayesian networks (BN). Although created statistically, this allows individual identification of the grade of malignancy as an internal cause that has enabled the development of the histological features to their observed state. The BN models are built from data representing 794 cases of astrocytomas with their malignant grading and corresponding histological features. The computerized learning process is improved when pre-specified knowledge (from the pathologist) about simple dependency relations to the histological features is taken into account. We use the BN models for both grading and causal analysis. In addition, the BN models provide a causal explanation of dependency between the histological features and the grading. This can offer the biggest potential for choice of an efficient treatment, since it concentrates on the malignancy grade as the cause of pathological observations. The causal analysis shows that all ten histological features are important for the grading. The histological features are causally ordered, implying that features of first order are of higher priority, e.g. for the choice of treatment in order not to allow the malignancy to progress to a higher degree. Due to the explanations of feature relations, the causal analysis can be considered as a powerful complement to any malignancy classification tool and allows reproducible comparison of malignancy grading.

Key words: Bayesian Networks, Astrocytomas, Classification and Interpretation

Introduction

The histological classification and grading of brain tumors and astrocytomas serves multiple purposes. 1: To establish a prognosis of the malignancy in each tumor. 2: To distinguish between different tumor processes with different phenotypic histological expressions. 3: To facilitate the histological, neuroradiological and clinical comparison between the different diagnosticians. 4: To establish relationships between the genetic and morphological changes of the tumors and clinical evolution of the patients. 5: To evaluate different techniques for clinical, neurosurgical and other therapeutic managements (Kaye and Walker, 2000; DeAngelis, 2001). From a historical point of view, malignancy was ascertained in a tumor if anaplasia was observed (Hansemann, 1920). Earlier, Cruveilhier (1829, 1835) had defined a neoplasm as “malignant” when it “recurred and killed the patient”. Later benignity and malignancy was defined by the histological evaluation of a neoplasm and there were two grading schemata, one by Bailey and Cushing (1926) and one by Kernohan and Sayre (1952). The Ringertz scheme (1950) was the most popular. This modified system was also used in Germany by Henschen (1955) and Schröder (Schröder et al., 1970). Zülch (1978, 1979, 1980a,b, 1986) closely defined the grade in terms of survival time: Grade I corresponds to cure or at least survival time of 5 and more years, Grade II is associated with postoperative survival time between 3-5 years, Grade III - with postoperative survival time between 2-3 years and Grade IV - survival between 6-15 months. The currently accepted grading according to the defined histological features is due to 1993/2000 WHO (Kleihues and Cavenee, 2000). Sometimes these classifications involve only histological aspects, in other cases architectural patterns, and the grading schemas retain a healthy amount subjectivity (Louis et al., 2001). Under these circumstances, subjective interpretations utilizing different techniques without homogeneous written

protocols lead, in many cases, to different grade assignments, which could also influence the diagnosis. Since well-delimited boundaries between histological grades of malignancy do not exist, the principal criteria for tumor classification and correlation with modern genetic findings are difficult to standardize. In this classification, feature definitions are verbal, and in many cases without established recorded protocols. They thus permit considerable interobserver variability and limited diagnostic reproducibility (The Childhood Brain Tumor Consortium 1989; Bruner et al., 1997; Conns et al., 1997; Prayson et al., 2000). The resulting tentative pathological diagnoses create significant clinical confusion. Objective statistical methods, and artificial neural networks (ANN) for diagnostic classification with multidimensional features have been useful in human and experimental pathology and laboratory medicine, in neuroradiological diagnosis of brain CT-images, meningiomas, astrocytomas, oligodendrogliomas and other tumors (Iglesias et al., 1983, 1986a,b, 1987, 1988 a-c, 1996, 1998 a,b, 2000; Kroh et al., 1988; Scarpelli et al., 1994; Martin et al., 1984; Brodbeck et al., 1997; Decaestecker et al., 1997; Sherf and Iglesias-Rozas, 1998; Papik et al., 1998; Conroy et al., 2000; Lennernäs et al., 2004; Wei et al., 2004).

The aim of the present study is to examine whether the histological grading of brain tumors can be accurately predicted and causally explained with the help of Bayesian networks.

Materials and methods

The present study was carried out with a large series of 794 human astrocytomas. The corresponding translation from the diagnostics code of astrocytomas to WHO-grading is given in Table 1. Ten histological characteristics (Table 2) were examined for the purposes of mathematical control of the diagnostic grading reached and for explanation.

Tumor grading and causal analysis

We treat the computerized grading of a tumor as

causal analysis (CA) and not only as an input-output classification task. For applications in the pathology, CA refers to searching for the grade of a brain tumor based on the set of observed histological features. In CA, we consider the malignancy grade as the cause of the effects, which are represented by the histological features. We consider causal analysis since we would like to order the importance of histological features for grading. Thus, causal analysis is supposed to be a methodology or systematic procedure for gathering and ordering of all relevant (in difference to available) data about features of a tumor. Although created statistically, CA identifies individually the grade of malignancy that has enabled the development of the individually observed histological features.

Modeling methods

There are many methods originating in artificial intelligence that have been used for medical diagnosis under uncertainty, e.g. neural networks, Bayesian networks (BN), and rule based algorithms. If the relations are uncertain, a neural network with proper training will be able to give the most probable diagnosis

Table 1. Distribution and number of astrocytomas in our study.

Tumor Group	Number of cases	Diagnosis code	WHO-grading
Astrocytomas			
Astrocytomas	56	21	2
Fibrillary A.	230	22	2
Protoplasmatic A.	50	23	2
Gemistocytic A.	6	24	2
Pilocytic A.	96	25	1
Juvenile A.	60	26	1
Subependymal Giant Cell As.	20	27	2
Anaplastic A.	238	28	3
Astroblastomas	38	29	4

Translation from the diagnosis code to WHO-grading.

Table 2. Histological definitions of the features collected in the present study.

1. Diffuse infiltration (DiffInfil):	Diffuse arrangement of tumor cells without demarcation or limitation to normal tissue.
2. Necrosis (Necro):	Localized destruction of tumor tissue.
3. Vascular abnormalities (VascAbno):	Any change in the vessel course and/or structure of the vessel wall.
4. Vascular occlusions (VascOccl):	Complete or incomplete occlusion of the vessel lumen by embolic or thrombotic matter without inclusions of tumor cells.
5. Cellular polymorphism (CellPoly):	Any variation in size or shape of the cell body.
6. Pericaryon size (Perica):	Average size of predominant cell forms.
7. Nuclear polymorphism (NucPoly):	Any variation in size or shape of the nuclei.
8. Typical mitoses (TypiMito):	Presence of normal figures of mitosis in tumor cells (with Ki-67: number of positive nuclei).
9. Atypical mitoses (AtypiMito):	Presence of any abnormal figures of mitoses in tumor cells.
10. Dedifferentiated cells (Undiff):	Presence of unidentifiable cell elements in tumor tissue.

In brackets is the shorthand notation used for modeling.

Causal probabilistic modeling for malignancy grading

given a set of symptoms, e.g. (Dytch and Wied, 1990; Astion and Wilding, 1992; Christy et al., 1995; Kolles et al., 1996; Burke et al., 1997). However, in contrast to Bayesian networks, it is neither possible to read the uncertainty of the conclusion from a neural network, nor to obtain a list of probable diagnosis in decreasing order. The probability tables of Bayesian networks can be assessed using a combination of data estimation (training), empiric studies, knowledge, experience, and various comparatively subjective estimates. A comparison of rule-based methods and methods based on Bayesian networks show that the essential difference is the way of handling the variables. While in rule based systems one is trying to model the expert's way of reasoning, with Bayesian networks one tries to model dependences in the medical domain and to use them for explanation and understanding of underlying phenomena.

Bayesian networks (BN) have gained acceptance in many application areas as summarized in Weidl (2002). In particular, they have been successfully applied in the medical domain, e.g. for diagnosis of muscle and nerve diseases (Olesen et al., 1989), diabetes advisory system (Andreassen et al., 1994), antibiotic treatment (Lucas, 2001), Triage and Radiology advisory systems (Shachter et al., 1987; Horvitz et al., 1989; Nathwani et al., 1990; Burnside et al., 2004).

BN have their origin in the combination of probability theory and graph theory. A Bayesian network is a knowledge base of a problem domain. It models the underlying structure of the domain, which is expressed by cause-effect relations among the domain variables. The causal relations are stochastic and not deterministic. They are expressed as conditional probabilities. BN allows estimation (update) of probability distributions of the unobserved variables, given the observed variables. For details see (Pearl, 1988; Jensen, 2001). For the considered application in pathology, the observed histological features are the effects of the cause, which is the grade of malignancy. One can represent this as an input-output classification task by the use of the naïve Bayes classifier, see Fig. 1, which has been described, e.g. by Duda and Hart (1973) and by Langley et al. (1992). It has been shown to have high predictive performance, although all effect variables are considered independent, which is a very strong assumption.

Friedman et al. (1997) have generalized the naïve

Bayes classifier based on the theory of learning Bayesian network. As a result, tree augmented naïve Bayes classifier (TAN) with interdependency and restrictions on the allowed interactions between the attributes (at most two parents), has been shown to outperform naïve Bayes, while still preserving the computational simplicity and robustness.

In this work we use BN as causal probabilistic models in order to naturally combine two tasks: classification and causal explanation. We use the fact that a BN gathers prior knowledge on the domain. This knowledge is incorporated in the BN through constraint learning, as explained later. So we enforce in this BN-application, the qualitative causal structure of the model as consisting of three symbolic layers: the grade of malignancy as the cause of the effects (i.e. the histological features HF). We have structured the effects HF into two symbolic layers (after the learning algorithm), i.e. the observed histological features of first order $HF^I_i (i=1..m)$ (with the cause as the only influence) and of second order $HF^{II}_i (i=1..n)$ (with influences from both the cause and at least one feature of first order), see Fig. 2a). The interdependency between the variables in these three layers is estimated based on structure learning and knowledge constraints.

In addition, we have extended the structure of the BN model with one more variable, namely the finer grading of tumor malignancy (Iglesias et al., 1986a,b, 1988), (denoted as Mk-fine in the BN of Fig 2b). The finer grading of tumor malignancy has 6 states in the scale from 0 to 4, with a step size of 0.5. It is expected to be useful in the choice of more individual therapy, taking into account patient-individual observations on the features and due to the hierarchical causal ordering of

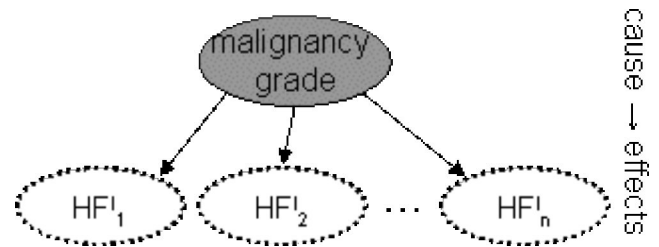


Fig. 1. Naïve Bayes classifier.

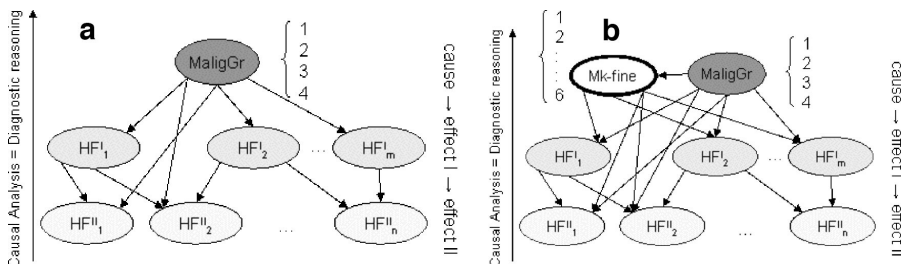


Fig. 2.a. Structure of the BN-model for causal analysis. Notations: MalignGr (parent node) is the malignancy grade (from 1 to 4) according to the WHO classification; HF^I, HF^{II} (child nodes) are the histological features of order I and II; b. Extended BN-model for causal analysis. Mk-fine is the finer grading coefficient, which is of interest for the choice of treatment.

their priority for treatment.

From the above one can see that a BN is a combination of two parts: qualitative (representing the knowledge, structure or causal relations between the variables in the domain) and quantitative (representing the strength of dependency between the variables expressed by conditional probability distributions CPDs). The qualitative causal structure is encoded in a directed acyclic graph (DAG) with a set of parent (e.g. MaligGr, Fig. 2) and child nodes (e.g. HF^I , HF^{II}), which are connected by directed causal links without loops (i.e. no paths, which end at their starting node). The DAG models the dependence and independence relations between the variables. The nodes in BN correspond one-to-one with the domain variables of the probability distributions such that there is one conditional probability distribution for each child node given its parents. We use the developed BN models in diagnostic applications, where we reason in a direction opposite to the causal arrow, i.e. observing the features (or measured variables), we conclude on the unobserved variables, see Fig. 2. This causal analysis is also referred to as diagnostic reasoning.

Causal Analysis (CA), d-separation and explanation of grading

We have chosen Bayesian networks for CA, since their causal structure allow mimicking human reasoning, just in the way a pathologist would have reasoned about the grade of malignancy given the histological features. The applicability of causal analysis, e.g. the use of causal ordering of features requires discussion on the causal mechanisms involved and about the validity of the Markov assumption at this abstraction level. Moreover, transparency of inference conclusions (i.e. the grading of malignancy) is required in order to provide a trustworthy causal analysis and decision support to the pathologists. Preferably, a user explanation should include a list of the grade of malignancy as the cause and its dependency on the evidence (i.e. the histological features HF) on which the conclusions were reached. For both, motivation of causal analysis and for explanation of grading, we utilize the notion of d-separation.

Essentially, the d-separation criterion and the Markov assumption are the two alternatives of relating the qualitative graphical structure of a BN to the conditional (in)dependencies underlying probability distributions. The Markov assumption implies that the state of a variable depends only on the state taken by the variable in its vicinity, i.e. the last shield this variable has from the influence of the other variables in the domain. For reasons of consistency, it can be shown that the d-separation criterion and the Markov properties are equivalent provided that the probability distribution is strictly positive, see Lauritzen et al., 1990. We focus in this work on the d-separation criterion as it is more easily applicable for the considered application. If one deals with the analysis of time series it would be

preferable to use directly the Markov assumption stating that the current state of the domain has an impact on its future state and it is independent of its past state. A detailed discussion on the causal relations in the domain with or without evidence is given in the appendix.

Learning of causal structure and probability distributions

The causal structure of the BN model can be built based on knowledge and experience. Alternatively, the structure can be learned from data representing various diagnostic cases with the corresponding histological features. The learning process is significantly improved when pre-specified knowledge from the pathologist is taken into account. This is known as structure learning from data cases under knowledge constraints. In general, there are two basic approaches for structure learning: a) Independence tests (in the following, referred to as constraint based procedures); b) Score and Search procedures. In real world applications one can access only limited data sets. Therefore, we have used the advantages of the NPC (Necessary Path Condition) algorithm due to Steck (2001) and Steck and Tresp (1999). It is an extension of the PC algorithm due to Spirtes et al. (2000), which allows incorporating the available knowledge from the pathologists. PC is a constraint-based learning algorithm relying on statistical independence tests (except for those pairs of variables for which a knowledge constraint has been specified). To have a valid independence statement, a number of links should be present in the graph. Thus, as a consequence of the statistical tests, an undirected link is added between each pair of variables for which no conditional independences were found, resulting in an equivalence class of DAGs. Furthermore, the edges are directed based on logical reasoning induced by the absence of colliders (i.e. links meeting in a node) and the fact that the underlying graph is assumed to be a DAG. The NPC is used for resolving ambiguous regions of uncertain edges induced by inconsistencies in the test results. The PC algorithm ignores such inconsistencies. For our case study, this means that based on medical expertise the pathologist has the opportunity to decide which variable is to be considered as a cause and which one as its effect.

The learning of the quantitative part of the BN model (i.e. the probability distributions) is done in this work with maximum likelihood estimation, since we utilize complete cases. A case is an assignment of values to some or of all the nodes of a domain. If values have been assigned to all nodes, the case is said to be complete; otherwise, it is said to be incomplete. In the case of missing values, it is useful to apply the EM algorithm (known also as EM learning or batch learning (Spiegelhalter and Lauritzen, 1990; Cowell et al., 1999)). The algorithm uses data (i.e., a set of cases) to estimate the conditional probability distributions when only the graphical structure is given. When no values are missing the EM algorithm performs a simple maximum likelihood estimate in a single iteration.

Methodology overview. Astrocytomas case study

To extract a proper structure of causal dependencies between the astrocytomas (the brain tumor under diagnosis, and the histological features (as consequences, Table 2), as well as to extract the strength of these dependencies (probabilities) and for inference (update of probabilities in light of new evidence) we have used the Hugin tool, see (Madsen et al. 2003). To learn the causal structure and probabilities of the diagnostic model the following steps have been performed: Data acquisition, Data preprocessing, Structure constraints for causal domain modelling, Structure learning from database cases, Extraction of dependency relations, Learning of probability distributions. Each of these steps is described below.

Data acquisition and preprocessing: Discretization of the malignancy coefficient

In the learning of model structure and probabilities, 794 astrocytomas cases have been utilized. The presence of each of 10 histological features has been rated on a scale from 0 (not present) to 3 (abundant presence) by visual inspection of the sections under a microscope. The type of astrocytoma and the WHO-grade (Table 1) have then been determined by an expert. A detailed description of the histological features used has been given in previous studies (Iglesias et al., 1986a,b, 1988; Cruz-Sánchez et al., 1988). These studies also introduced the malignancy coefficient Mk-fine, which takes continuous values between 1.00 and 4.00 and is computed from these 10 histological features with GRADO-IGL (a software utilizing discriminant analysis, i.e. a statistical technique for finding the linear combination of features which best separate two or more classes of object or event). GRADO-IGL has been developed by one of the authors (Iglesias, 1988) and can be downloaded from <http://www.igl-rozas.de/4.Grado->

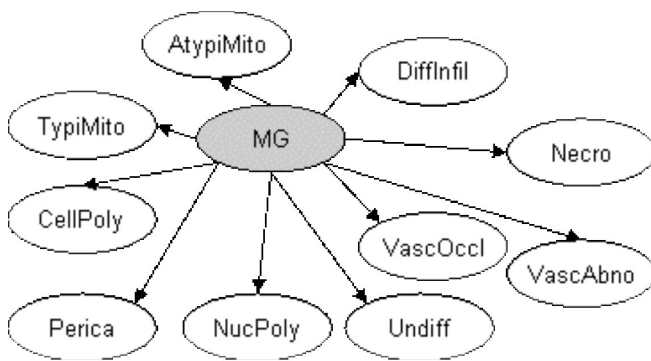


Fig. 3. Specification of structure constraints: the malignancy grade (MG denotes both MaligGr and Mk-fine) is causally affecting the observed histological features. The causal dependencies are drawn as links, directed from cause to effect.

IGL.htm). In the preprocessing procedure we have discretized the coefficient of malignancy into 6 discrete states: from 1 to 4, with a step size of 0.5.

Structure constraints for causal domain modelling

The purpose of specification of any known dependences or independences in the data set is to improve the causal structure of the BN models, which are learned from data cases with help of the NPC-algorithm. Initially, we imply only one type of constraint, specifying the diagnosis (WHO-grading) as the cause of the observed histological features. This reflects the causality of the event appearance, i.e. the features of a tumor are observed only if the patient suffers from that disease and consults with a clinic for assessment. No explicit dependency relations between the histological features have been defined. These assumptions reflect the certainty of knowledge about the causality relations in the problem domain. Here, we have applied structure constraints as shown in Fig. 3.

Structure learning from database cases. Extraction of dependency relations

After gathering of the required information (data, preprocessing/discretization and specifications of structure constraints), the learning algorithm is prepared to learn the structure of the model for the specified data. The structure learning may be performed using different algorithms, e.g. PC or NPC. We have chosen the NPC algorithm (with probability of rejecting a true independence hypothesis given by the level of significance = 0.05), since we have a limited data set with 794 cases. Thus, to resolve the uncertainties, the NPC algorithm relies on interaction with the user (i.e. the pathologist) where the medical expert gets the opportunity to decide on directionality of undirected links and to resolve the ambiguous regions. The learning

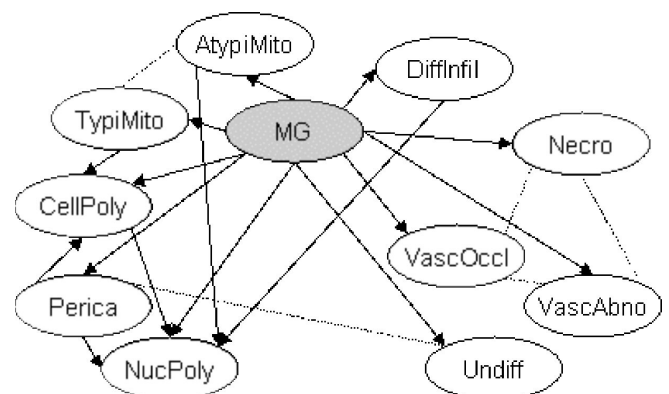


Fig. 4. Uncertainties in BN-structure found by the PC algorithm. The directionality of undirected links can be determined interactively by the user. The directed links represent certain causal relations.

algorithm (NPC) incorporates user specified constraints (see Fig. 3) for enforcing of known dependency relations and determines the structure of the model. The undirected edges are induced due to equivalence classes of DAGs. At this step, the joint probability distribution will be the same no matter how edges are directed. Fig. 4 shows the obtained model structure with unresolved uncertainties.

The user (pathologist) may help to resolve these structural uncertainties based on his own experience/expertise/knowledge of the causal dependencies between the histological features (i.e. the dependent variables). In the case of astrocytomas, shown in Fig. 4, the pathologist input on causal dependences implied that the state of vascular abnormality is influencing the state of vascular occlusion, and these together are affecting the state of necrosis. Moreover, we have selected the presence of unidentifiable cell elements in tumor tissue (represented by the variable “Undiff”) as the cause of the average size of predominant cells forms (variable “Perica”). The resulting model is shown in Fig. 5. The node colored in dark grey indicates the cause (the tumor according to the WHO-grading of malignancy $\text{MaligGr} = 1-4$). In addition, the model incorporates the dependency of the finer malignancy grade Mk-fine (1-6) (dark grey) for more individual choice of treatment. Grey indicates the histological features of first order (dependent only on the malignancy grade MaligGr and Mk-fine), i.e. for the case of astrocytomas: $\text{HF}_i^I (i=1..m, m=4) = \{\text{DiffInfl}, \text{VascAbno}, \text{Undiff}, \text{AtypiMito}\}$. The nodes colored with light grey indicate the histological features of second order (dependent on the malignancy grade and only one first order histological feature), i.e. for the case of astrocytomas: $\text{HF}_i^II (i=1..n, n=3) = \{\text{VascOccl}, \text{Perica}, \text{TypiMito}\}$. The white nodes are causally dependent on the malignancy grade and at least three other histological features, i.e. for the case of astrocytomas, these are $\{\text{Nekro}, \text{NucPoly}, \text{CellPoly}\}$.

The learning NPC-algorithm has estimated from the data cases the probabilistic dependency between MaligGr (WHO-grading in 4 states) and the finer grading (Mk-fine). In addition, we have chosen the causality direction from MaligGr to Mk-fine . This does not affect the probability distributions, but ensures the correct mapping between the two grading scales. After the qualitative part (i.e. structure) of the model is learned, the maximum likelihood algorithm estimates the probability distributions from the data, as shown in Fig. 6. For the BN model structures obtained in Fig. 5 we have estimated with the maximum likelihood algorithm the probability distributions from the same data, i.e. 794 complete cases (all variables have been assigned values) of astrocytomas. The resulting probability distribution is shown in Fig. 6a). Thus, this completes the learning of the qualitative and quantitative part of the BN model for computerized grading of astrocytomas. To provide reproducible results and to speed up the diagnostic process, we utilize computerized reasoning by inference

algorithms. The inference results for the diagnosis are presented in the next section.

Results

Inference

The inference algorithm is used to draw conclusions on the grade of astrocytomas based on the provided findings (evidence on the 10 histological features) and on specific information learned as a BN model of the medical domain (in our study - from 794 astrocytoma cases). Fig. 6a) shows the distribution of probabilities in grades 1-4 based on 10 histological features for the learned default configuration of the BN model, while Fig. 6b)-e) for WHO-grades 1-4, as well as for the finer grading Mk-fine , together with the corresponding probability distribution of 10 histological features. The histological features, which are influenced by those of first and second order, play an essential role in the inference on the malignancy grade. At the same time they are of lower priority while choosing the proper treatment of the malignancy grade. The flexibility of the method is in the possibility to consider also combinations of other probability configurations, which is obvious from Fig. 6b-e). The symbol “e” indicates evidence. Here we show the selection of MaligGr in a certain (WHO-) grade of malignancy and the corresponding results of the probability distribution of the finer grading Mk-fine together with the histological features. As a reminder: histological features (HF) in scale 0 means absence of the feature with some probability.

Fig. 6b) shows the probability update for all HF after selecting grade 1 for the malignancy (MaligGr). This update corresponds to the finer grading coefficient Mk-fine for grade 1.5-2 with highest probability ($p=47.44\%$), followed by grade 1-1.5 with $p=32.05\%$ and grade 2-2.5 with $p=11.54\%$. This grading of the inference algorithm is confirmed in the presence of the following combination of histological features: here we state only

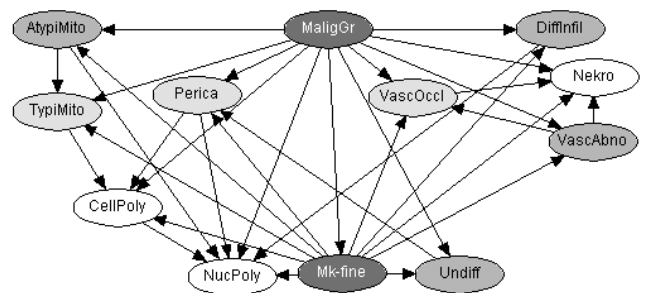


Fig. 5. Structure of the BN model for grading of astrocytomas learned from data cases, where the pathologist experience on grading has also been taken into account in order to resolve the undirectedness of some causal links.

Causal probabilistic modeling for malignancy grading

the combinations of highest probability, represented by the longest bars: {DiffInfil=3(62%),Nekro=0(92%),VascAbno=0(33%),VascOccl=0(82%),Undiff=0(70%),NucPoly=1(53%),CellPoly=1(55%), TypiMito= 0(96%),AtypiMito=0(100%), Perica=1(76%)}. For malignancy grade MaligGr=1, the most probable (and not absent) HF^I is only DiffInfil=3 (62%), which is causally affecting only one of the most probable (and not absent) HF^{II}: Perica=1 (76%). These HF are affecting in their turn the HF: NucPoly=1 (53%) and CellPoly=1 (55%).

In analogy to Fig. 6b, we show in Fig. 6c the probability update for all HF for MaligGr=2 corresponding to Mk-fine=1.5-2 with p=45%; Mk-fine=2-2.5 with p=26% and Mk-fine=1-1.5 with p=17%. Similarly Fig. 6d shows the probability update for all HF for MaligGr=3 corresponding to Mk-fine=3-3.5 with

p=57%, Mk-fine=2.5-3 with p=15% and Mk-fine=3.5-4 with p=13%. Likewise Fig. 6e shows the probability update for all HF for MaligGr=4 corresponding to Mk-fine=3.5-4 with p=50%, Mk-fine=3-3.5 with p=34% and Mk-fine=2.5-3 with p=8%.

Note that the BN should only be used in the context for which it was developed. Namely, the probability corresponds to the strength of extracted dependency, which represents only pathological biopsies (not normal tissue). We also note that the distribution of probability for the 10 histological features is in very good agreement with the pathologist's experience.

Validation

For validation of our model we ask how reliably it predicts the WHO grade, given all 10 histological

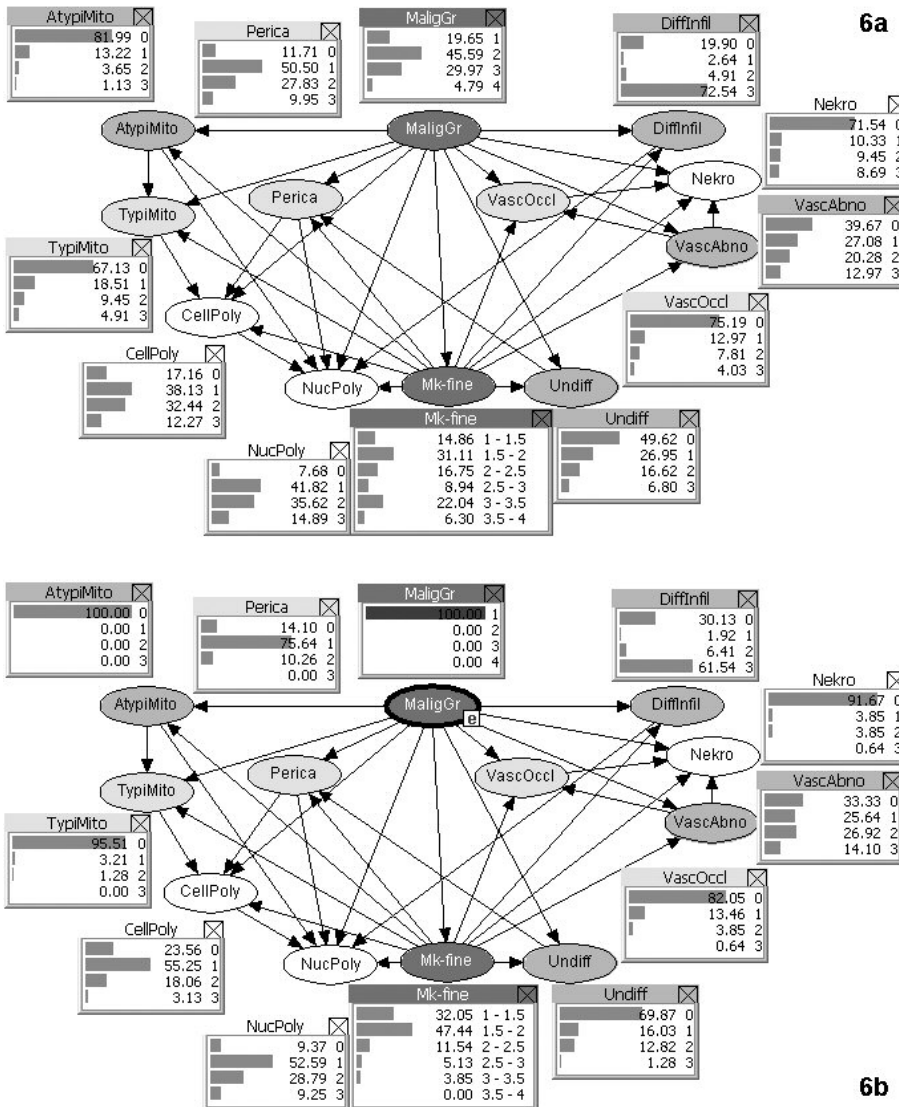
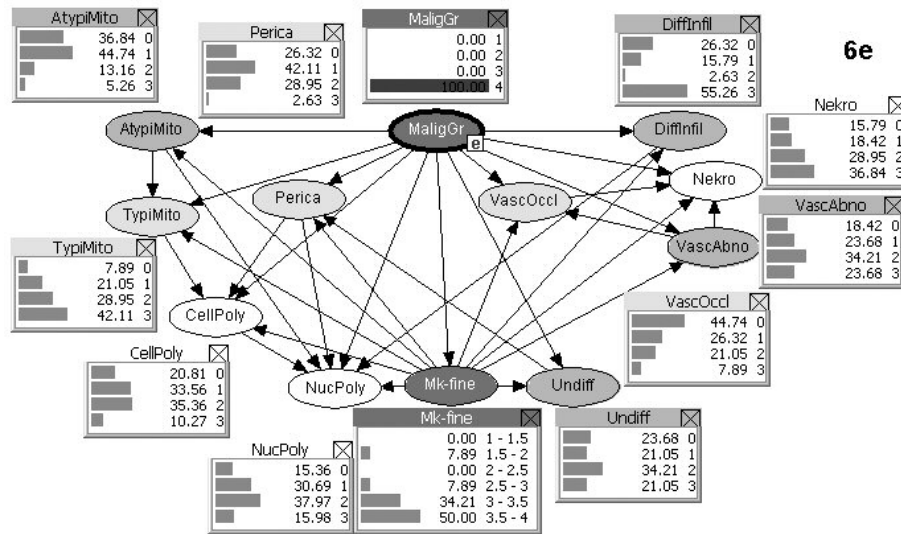
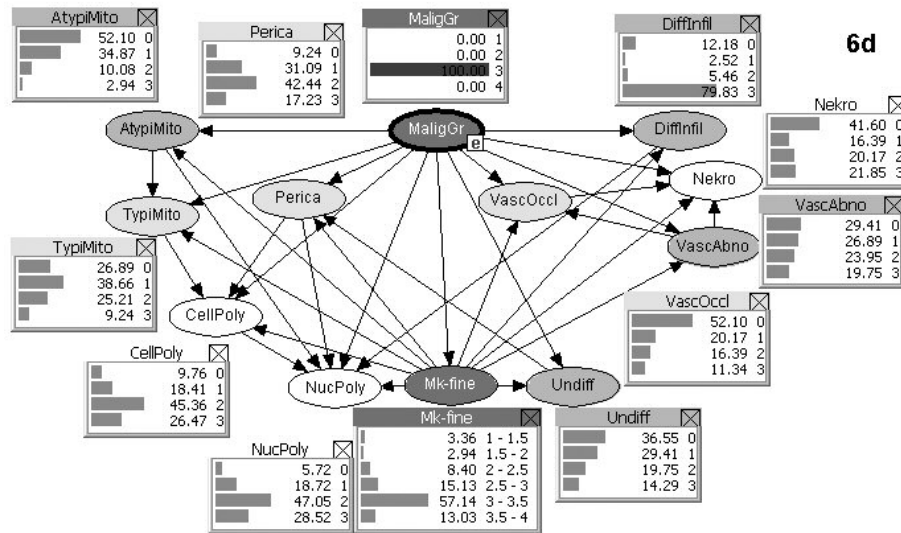
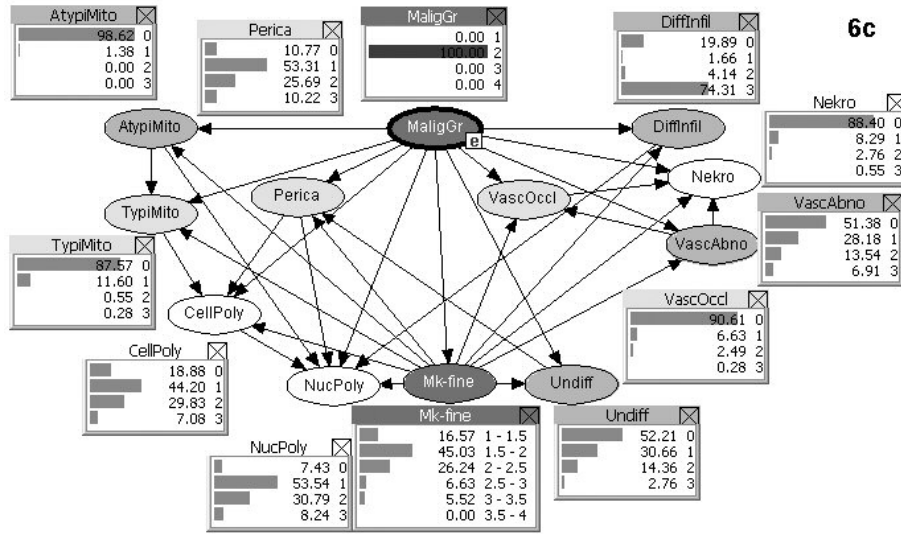


Fig. 6. Inference for grading of astrocytomas. **a.** Default distribution of probability in the cause variable: malignancy grade MaligGr, together with the histological features. **b.** (WHO) Malignancy grade 1 and probability distributions for HF and Mk-fine. **c.** (WHO) Malignancy grade 2 and probability distributions for HF and Mk-fine. **d.** (WHO) Malignancy grade 3 and probability distributions for HF and Mk-fine. **e.** (WHO) Malignancy grade 4 and probability distributions for HF and Mk-fine.

Causal probabilistic modeling for malignancy grading



Causal probabilistic modeling for malignancy grading

features. To this end, we divided the 794 cases into 10 subsets of approximately equal size (about 79 cases) and grade distribution for cross-validation, i.e. we trained the algorithm with 9 of the subsets and tested the classification performance with the remaining one. Table 3 shows the AUC (area under the receiver operating characteristic) and the accuracy (percentage of correct predictions) of each of the cross-validation runs and the corresponding averages for testing and training sets. We use the multi-class AUC definition of Hand and Till (2001), which measures how well each class is separated from all the other classes.

Sensitivity to evidence (SE) analysis

In this work, we have analyzed the quality of the learned domain model by the so called Sensitivity to Evidence (SE) Analysis. This means that the results of a probability update (propagation of evidence) obtained out of the inference are examined with respect to how sensitive they are to variations in the provided evidence (histological features).

We consider myopic hypothesis driven SE analysis on discrete random variables, as discussed in Kjærulff and Madsen (2005). Assume a situation where a pathologist has to make a decision based on the probability distribution of a hypothesis variable (the WHO-grading MaligGr or Mk-fine of a brain tumor). This could, for instance, be information to a surgeon deciding on whether to operate on a patient or not. Prior to deciding about an operation, the surgeon has the option to investigate the impact of the collected information on the posterior distribution of the hypothesis variable. Given a set of findings (histological features) and a hypothesis (diagnosis), we can analyze: 1) which sets of findings are in favor, against, or irrelevant for the hypothesis, 2) which sets of findings discriminate the hypothesis from an alternative hypothesis, 3) what if a variable had been observed to a different value than the one observed, etc. These questions can be answered by SE analysis. In particular, we analyze the impact, the findings, and the discrimination of competing diagnosis grades and consider what-if analysis for different or uncertain observations of the histological features, as shown below. Everywhere in the sensitivity to evidence analysis, we use as example the following evidence

(findings ϵ) on the histological features, denoted with (#), which is also shown in Fig. 7:

{DiffInfil=3,Nekro=0,VascAbno=1,VascOccl=0,Undiff=0,NucPoly=1,CellPoly=1,TypiMito=0,AtypiMito=0,Perica=1}. (#)

Impact

The impact of a subset of the evidence on a state of the hypothesis variable is determined by computing the normalized likelihood (NL) of the evidence given the hypothesis. It can also be computed from the prior and posterior probabilities, as shown in (1):

$$NL = \frac{P(\epsilon|x)}{P(\epsilon)} = \frac{P(\epsilon,x)/P(x)}{P(\epsilon)} = \frac{P(x\epsilon)P(\epsilon)/P(x)}{P(\epsilon)} = \frac{P(x\epsilon)}{P(x)}, \quad (1)$$

where $P(x)$, $P(\epsilon)$ are the prior probabilities of the state x and of the set of evidence ϵ . $P(x\epsilon)$ is the conditional probability of the state x under the condition of observing a set of findings ϵ and $P(\epsilon|x)$ is the likelihood. NL determines the subsets of the evidence (findings) acting in favor ($NL > 1$), against ($NL < 1$) or irrelevant ($NL = 1$) for each possible hypothesis ($MaligGr \equiv WHO\text{-}grade$). In our model of astrocytomas, we have studied the impact of various subsets of evidence on the state of the hypothesis variable $MaligGr \equiv WHO\text{-}grade = 1,2,3,4$, expressed by the normalized likelihood NL. This information shows which findings do or do not have an impact on the conclusion concerning the grading of a tumor.

Table 4 shows that some findings (for $MaligGr=2$) are irrelevant ($NL=1$) for grade 1 and 2, and almost all of them (with the exception of the first line in Table 4) are against grade 3 and 4. From Table 4 we can conclude that observing only $VascAbno = 1$ and $VascOccl = 0$ is evidence which is irrelevant for $MaligGr_WHO\text{-}grade = 1$, speaks for $WHO\text{-}grade = 2$ (since $NL > 1$) and rejects the possibility of $WHO\text{-}grade = 3$ or 4 ($NL < 1$).

Discrimination

The discrimination of a pair of competing hypotheses is based on the calculation of the Bayes' factor for all subsets of a selected set of evidence:

$$B = \frac{\text{posterior odds ratio}}{\text{prior odds ratio}} = \frac{P(x\epsilon)/P(y\epsilon)}{P(x)/P(y)} = \frac{P(\epsilon|x)}{P(\epsilon|y)} = \frac{L(x\epsilon)}{L(y\epsilon)}, \quad (2)$$

Table 3. Cross-validation results measured in AUC and accuracy.

AUC*100	1	2	3	4	5	6	7	8	9	10	av.
test	78.73	85.34	80.50	72.40	82.38	83.01	80.52	85.03	83.23	68.95	80.01
train	96.34	96.25	96.21	96.17	96.18	96.49	96.39	96.26	96.16	96.39	96.28
Accuracy	1	2	3	4	5	6	7	8	9	10	av.
test	70.13	72.50	67.09	77.50	64.20	62.34	67.50	78.48	71.25	66.67	69.77
train	82.15	81.79	83.22	82.49	83.31	82.85	83.89	82.24	82.49	82.33	82.68

Causal probabilistic modeling for malignancy grading

where $P(x|\epsilon)$ and $L(x|\epsilon)$ are respectively the conditional probability and the likelihood of the state x while observing a set of findings ϵ . In general, a bigger Bayes factor is more discriminative for the alternative hypotheses y , as obvious from (2). For example, the probability distribution shown in Fig. 7 is based on the findings on the 10 histological features, as specified in (#). We have used the Bayes factor to examine which of the two competing hypotheses variables, namely the WHO-grading (represented by $\text{MaligGr} = x$) and the more fine resolution of malignancy grade (represented by $\text{Mk-fine} = y$) is easier to discriminate by the provided evidence (with different random combinations of the findings). To discriminate between the WHO-grading

with $\text{MaligGr} = 1$ ($p=42\%$) and 2 ($p=58\%$) (see Table 1, Table 5 and Fig. 7), we have a choice to consider instead $\text{Mk-fine} = 1.5 - 2$ ($p=96\%$). Not only the higher probability, but also the Bayes factor ($B < 1$) for all possible combinations of findings is in favor of the more fine resolution. This combination of evidence also discriminates the closest grading classes of $\text{Mk-fine} = (1-1.5)$ or $(2-2.5)$. For all random combinations of features with $B > 1$, the Bayes factor provides a reliable discrimination criteria, see Table 5. Here, the smallest Bayes factor $B = 0.4$ in favor of $\text{Mk-fine} = 1.5-2$ is inferred from evidence due to a combination of the following histological features: {DiffInfil, VascAbno, Cellpoly, Perica, TypiMito, Undiff}. Considering the

Table 4. Normalized likelihood NL of the subsets of the evidence (findings) acting in favor ($NL > 1$) of, against ($NL < 1$) or irrelevant ($NL = 1$) for each possible hypothesis (the WHO-grade of MaligGr).

Histological features										MaligGr _ WHO-grade			
Diff Infil	Necro	Vasc Abno	Vasc Occl	Cell poly	Perica	Nuc Poly	Typi Mito	Atypi Mito	Undiff	1	2	3	4
X										NL<1	NL=1.02	NL=1.1	NL<1
		X								NL<1	NL=1.04	NL>1	NL<1
X		X								NL<1	NL=1.08	NL<1	NL<1
X			X							NL<1	NL=1.25	NL<1	NL<1
X		X	X							NL<1	NL=1.29	NL<1	NL<1
		X	X							NL=1	NL=1.26	NL<1	NL<1
no evidence										NL=1	NL=1	NL=1	NL=1
All other subsets of the evidence (findings) are acting in favor ($NL > 1$) of grade 1 and 2 and against grade 3 and 4.										1.01-2.7	1.01-1.61	NL<1	NL<1

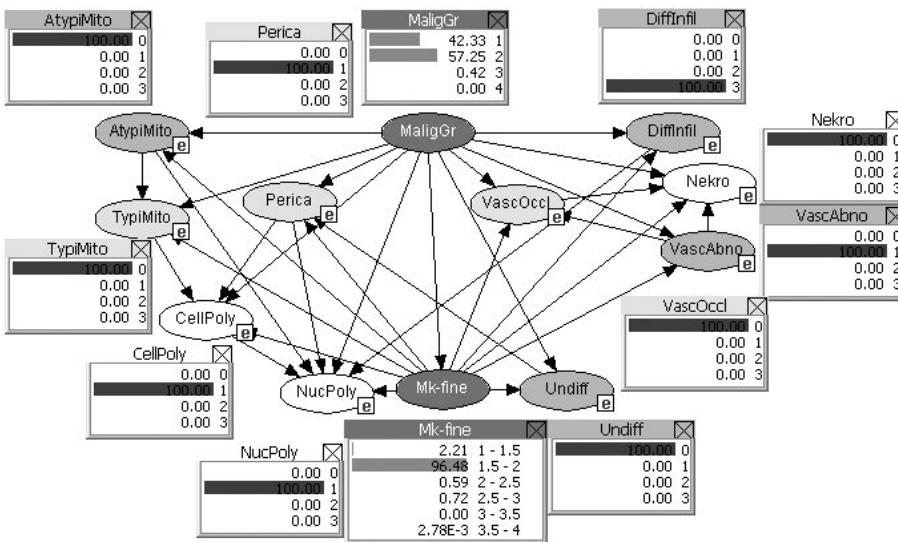


Fig. 7. Evidence (findings e) used in the sensitivity to evidence analysis, i.e. analysis of impact, findings, discrimination of competing diagnosis grades and what-if-analysis.

Causal probabilistic modeling for malignancy grading

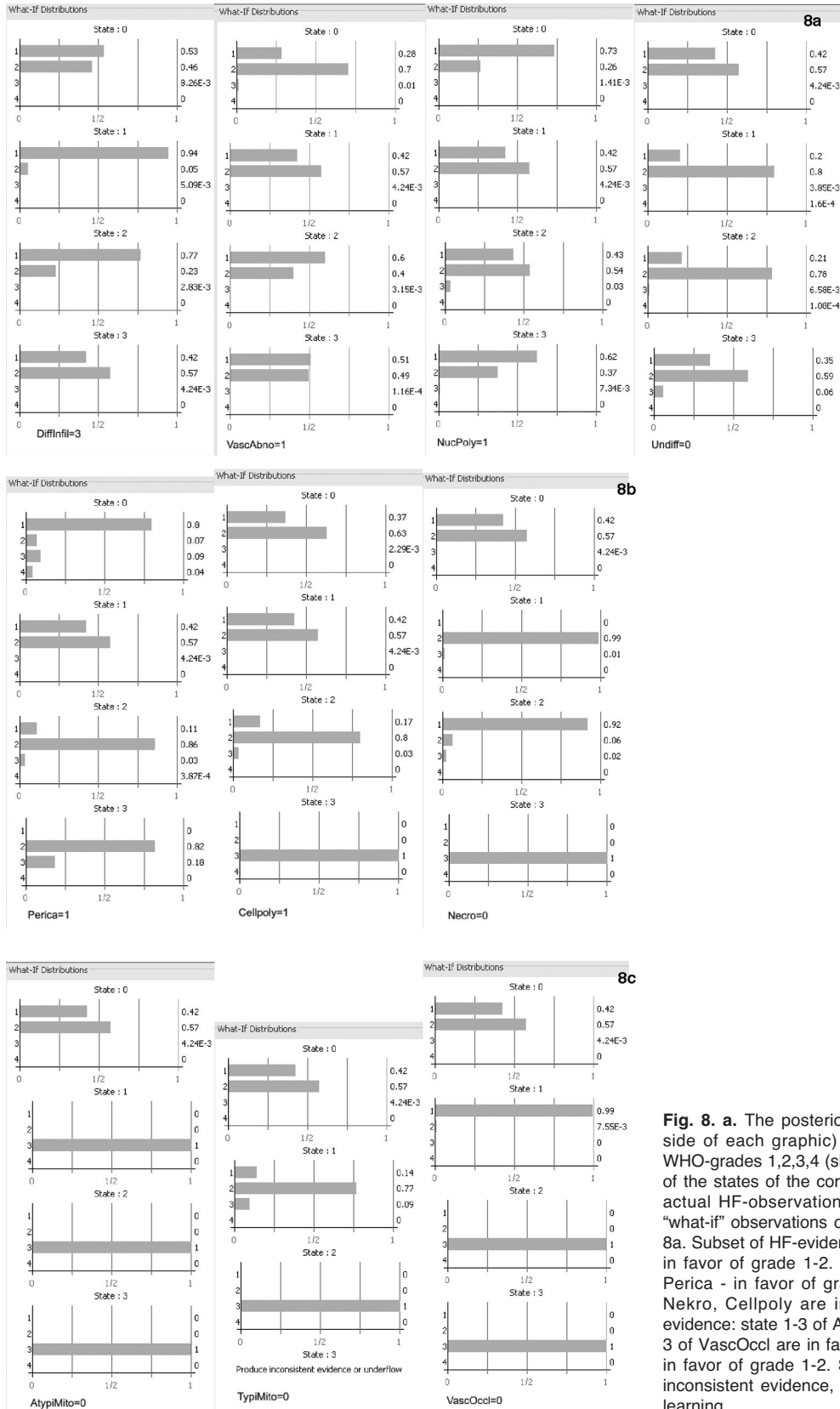


Fig. 8. a. The posterior distribution (shown on the right hand side of each graphic) of the hypothesis variable MaligGr in WHO-grades 1,2,3,4 (shown on the left hand side) as a function of the states of the corresponding ten histological features (the actual HF-observations are written below each graphic, the “what-if” observations of alternative HF-States are on top). Fig. 8a. Subset of HF-evidence {DiffInfil, VascAbno, NucPoly, Undiff} in favor of grade 1-2. **b.** Subset of HF-evidence: state 1-3 of Perica - in favor of grade 1-2, while state 3 of the variables Necro, Cellpoly are in favor of grade 3. **c.** Subset of HF-evidence: state 1-3 of AtpiMito, state 2 of TypiMito and state 2-3 of VascOccl are in favor of grade 3, the rest of the states are in favor of grade 1-2. State=3 of variable “TypiMito” produces inconsistent evidence, since it was not present in the data for learning.

Table 5. Discrimination of a pair of competing hypotheses (Mk-fine grading with respect to WHO-grading = 2) for the set of evidence (#).

Mk-fine grading	1 – 1.5	1.5 – 2	2 – 2.5	2.5 - 3	3 – 3.5	3.5 - 4
Probability p	0.02	0.96	0.01	0.01	0	0
Bayes factor B	0.63 -14.43	0.4 – 0.99(B<1)	0.78 - 38.2		B>1	

evidence on the rest of the histological features, as specified in (#), this still leaves $B < 1$, which supports the more fine resolution of the grade. Similar analysis on the other states of MaligGr confirms this conclusion on hypothesis discrimination.

Further refining (e.g. with 12 or more states) of the resolution of the malignancy grade does not necessarily provide more information in favor of the choice of consequent treatment. Thus a relatively small refining from 4 WHO-grades to 6 states (for Mk-fine) can already improve the diagnosis and prognosis on the development of malignant tumors. This is an argument in favor of GRADO-IGL, which provides continuous grading of malignant tumors. It has to be noted that the aim is not the absolute number, but rather the prognosis in relation to the treatment.

What-if Analysis

What if an observed discrete random variable had a different state? We consider a hypothesis driven approach to what-if analysis. Hypothesis driven what-if analysis is performed by computing the posterior distribution of the hypothesis variable for each possible state of the observed variable. Fig. 8 shows the results of the what-if analysis on the histological feature for values, different from those actually observed, specified in (#). Based on the what-if Analysis, we can conclude that all findings are important for grade 1-2. In particular: “DiffInfil, VascAbno, NucPoly, Undiff, Perica” are in favor of grade 1-2, while the state 3 of the variables Nekro, VascOccl, Cellpoly, TypiMito and state 1-3 of AtypiMito are in favor of grade 3.

We have also studied the sufficiency of the findings (i.e. evidence for the 10 histological features) on the hypothesis MaligGr (i.e. astrocytomas) distributed in 42% for grade 1 and 58% for grade 2. This set of findings was sufficient for grade 1, 2, 3 and insufficient for grade 4. This is reasonable since only 4% of the cases were rated as grade 4.

In summary, based on the performed sensitivity to evidence analysis, we can conclude that the quality of the BN models, learned in combination with pathology data and constraints representing domain knowledge, is good, and the inference results have been confirmed by an experienced pathologist.

Discussion

From the created model structure for astrocytomas

diagnostics we can provide interpretation of causal biological dependency between the histological features. From the structure of the BN model learned from the 794 cases, one can see several levels of hierarchy. The grade of malignancy of astrocytomas (MaligGr) is the cause with effects expressed by the histological features. The last can be divided into first and second order histological features. The first order (atypical mitoses, dedifferentiated cells, vascular abnormality and diffuse infiltration) are directly dependent only on the grade of malignancy, while the second order (typical mitoses, pericaryon size, vascular occlusions) are causally influenced by two variables: the grade of malignancy together with a histological feature of first order. The three remaining histological features (necrosis, nuclear polymorphism and cellular polymorphism) can be considered only as effects and their biological importance will probably not lead to the most efficient therapy for the corresponding grade of malignancy.

The requirements on the causal analysis include creating reliable advice to the pathologists. This reflects in requirements on the methodology, such as: a) visibility of causal dependency, b) measurement and comparison of different scenarios of feature combinations. Thus, causal analysis can be considered as an efficient tool for: A) reproducible comparison and reliable advice; B) understanding and explanation of the biological dependency between histological features, which allows focusing the research on the primary histological features of astrocytomas; C) education and training of young pathologists in order to avoid the overflow of data and speed up the learning process by visualizing the structure of dependency relations.

When to use bayesian networks

Based on the theory and our experience with applications of Bayesian Networks, we can conclude that BNs are suitable when 1) the problem domain is characterized by chain causality. That is, a cause has an impact or effect, which causes other effects and so on. 2) Well-established knowledge, experience and expertise on the causal structure of the selected medical domain is available or can be obtained. 3) Parts of the causal relations are of stochastic character. That is, different configurations of the same causes may have different impacts in different cases, or the same cause may have different effects in different cases. 4) The observed events in the domain are well-defined. 5) The set of all possible abnormality/failure cases in the problem

Causal probabilistic modeling for malignancy grading

domain is of well-limited and manageable size. 6) The inference task is of a diagnostic nature under uncertainties. That is, based on measured and calculated features, or pathologists' observations, one can estimate the probabilities of possible causes (e.g. grade of malignancy) and use the most probable cause to decide on suitable treatment or corrective/surgery actions. 7) Explanation of inference conclusions is needed in order to understand the hierarchy and dependency of histological features and to provide decision support for pathologists.

Outlook

We have discussed causal analysis, since it allows analyzing the priority of histological features. This can be linked to the treatment of the actual grade of a tumor. In the continuation of this work, we will show how the grade findings relate to the chosen treatment, e.g. hormonal, radio or chemotherapy. This can allow pathologists/surgeons to analyze the biological significance of the order of histological features and comparative benefits of alternative treatment options (or corrective/surgery actions). Improper treatment, e.g. due to wrong grading, might allow the tumor to progress to a higher grade and reduce the survival time for a patient.

Looking into the future, the use of Bayesian networks can also open the perspective for understanding and interpretation of the causal biological mechanisms behind malignant developments, if clinical data are available. This could help in the choice of strategic focus in the biological and clinical research of astrocytomas, whereby the priority would be on the study of the root causes behind malignant developments.

Appendix

Definition [*d-separation*], based on (Pearl, 1988) and (Jensen, 2001).

The independence relations induced by a set of nodes in a directed acyclic graph can be determined using the d-separation criterion Fig. 2, which states:

Two distinct variables X and Z in a causal network are d-separated (i.e. $X \perp Z$), if for all paths connecting X and Z there is an intermediate node Y such that one of the following statements is satisfied:

1) Y is the middle node in a serial ($X \rightarrow Y \rightarrow Z$ or $X \leftarrow Y \leftarrow Z$) or diverging connection ($X \leftarrow Y \rightarrow Z$) and Y is instantiated by evidence, i.e. $X \perp Z | Y$

2) Y is the middle node in a converging ($X \rightarrow Y \leftarrow Z$) connection and neither Y nor any of its descendants (effects) have received evidence, i.e. $X \perp Z | \emptyset$.

Two nodes are *d-connected* (denoted as \parallel), if they are not d-separated (denoted as \perp). The d-separation criteria makes it possible to determine whether or not evidence on a variable X can change the belief about another variable Y . An algorithm for determining whether or not two variables are d-separated given evidence on a subset of the remaining variables has been implemented in the

Hugin tool and has linear (in the size of the graph) time complexity.

Explanation of inference conclusions in all CA applications follows the above formulated d-separation criteria. In the general structure of Fig. 2, one can encounter the same types of connections between the variables:

- serial connection: $\text{MaligGr} \rightarrow HF^I_{i=1..m} \rightarrow HF^{II}_{k=1..n}$
- diverging connection: $HF^I_1 \leftarrow \text{MaligGr} \rightarrow HF^I_2$
- converging connection: $HF^I_1 \rightarrow HF^{II}_2 \leftarrow HF^I_2$ or $\text{MaligGr} \rightarrow HF^{II}_2 \leftarrow HF^I_2$

We use d-separation for modeling and explanation of inference. The graph of Fig. 2 is used in the explanations below.

- $HF^I_1 \perp HF^I_2 | \emptyset$: HF^I_1 and HF^I_2 are dependent, when there is no evidence. There are two paths between HF^I_1 and HF^I_2 , which go through the converging connection $HF^I_1 \rightarrow HF^{II}_2 \leftarrow HF^I_2$ and through the diverging connection $HF^I_1 \leftarrow \text{MaligGr} \rightarrow HF^I_2$. Since there is no evidence, condition 1) of the d-separation definition is violated. Generalized for the layer containing the histological features of first order HF^I_j , this implies pairwise dependency of HF^I_j when no evidence is provided.

$HF^I_1 \perp HF^I_2 | HF^{II}_2$: HF^I_1 and HF^I_2 are dependent, if HF^{II}_2 (as descendent of HF^I_2) is known, since condition 2) of the d-separation definition is violated. This implies pairwise dependency when the histological feature of second order HF^{II}_2 is instantiated.

$HF^I_1 \perp HF^I_2 | \text{MaligGr}$: HF^I_1 and HF^I_2 become independent, if MaligGr is known in the diverging connection $HF^I_1 \leftarrow \text{MaligGr} \rightarrow HF^I_2$ and there is no evidence on HF^{II}_2 in the converging connection $HF^I_1 \rightarrow HF^{II}_2 \leftarrow HF^I_2$. Thus, observing abnormality in HF^I_1 , will tell us nothing about whether or not HF^I_2 is also abnormal (and reverse).

$HF^I_2 \parallel \text{MaligGr} | \emptyset$: HF^I_2 and MaligGr are dependent, when the conditioning set is the empty set (i.e. \emptyset).

$HF^I_2 \parallel \text{MaligGr} \{ HF^I_1, HF^I_3 \}$: There are at least three paths from MaligGr to HF^{II}_2 : $P_{HF^I_1}^1 = \text{MaligGr} \rightarrow HF^I_1 \rightarrow HF^{II}_2$, $P_{HF^I_2}^1 = \text{MaligGr} \rightarrow HF^I_2 \rightarrow HF^{II}_2$, and $P_{HF^I_3}^1 = \text{MaligGr} \rightarrow HF^I_3 \rightarrow HF^{II}_2$. Both $P_{HF^I_1}^1$ and $P_{HF^I_3}^1$ are blocked by observation on HF^I_1 , HF^I_3 i.e. the condition 1) of d-separation is violated. But, looking at Fig. 2, one can see that the $P_{HF^I_2}^1$ path is not blocked by evidence. Therefore, HF^{II}_2 and MaligGr are dependent.

$HF^I_2 \perp \text{MaligGr} \{ HF^I_1, HF^I_2, \dots, HF^I_m \}$: If there were only the three paths $P_{HF^I_1}^1$, $P_{HF^I_2}^1$, $P_{HF^I_3}^1$ and if they all were blocked by evidence, then HF^{II}_2 and MaligGr are independent.

$HF^I_{i, i=1..m}$ are the histological features of first order derived as effects of the malignancy grade. Their risk assessment of being abnormal can provide an indication on pathological developments. In the serial connection, since $HF^I_{i, i=1..m}$ are instantiated by evidence, MaligGr and $HF^{II}_{j, j=1..n}$ are independent. Thus, d-separation reflects also the fact that histological features of second order HF^{II}_j are events which can be prevented from developing to a higher degree, if a suitable treatment on

the cause MaligGr is undertaken at the right time. For example, if abnormal levels of the first order histological features have been measured, they might allow (if not treated properly) the malignancy to progress to a higher degree.

If we check explicitly the condition of MaligGr, it becomes instantiated by evidence. Then $HF_1^I \perp HF_2^I \mid \text{MaligGr}$: HF_1^I and HF_2^I are independent in the diverging connection $HF_1^I \leftarrow \text{MaligGr} \rightarrow HF_2^I$, if either MaligGr are known. This allows simplification in BN modeling for CA. We assume conditional independence of effects (or histological features of first order) given their common cause (the grade of malignancy).

Similar causal argumentation applies in the extended model with the finer grading Mk-fine. Then we have two hypothesis variables, MaligGr and Mk-fine, where Mk-fine is related to the histological features in a similar causal manner to MaligGr.

The d-separation implies in the extended model the following properties, in addition to the ones discussed above:

$\text{MaligGr} \perp \text{Mk-fine} \mid \emptyset$: Observing MaligGr is in a certain grade, will tell us nothing about whether or not Mk-fine is also at a grade close to the one of MaligGr (and reverse). All paths between MaligGr and Mk-fine go through the converging connections $\text{MaligGr} \rightarrow HF_1^I$, $HF_2^I \leftarrow \text{Mk-fine}$. Since the conditioning set is the empty set (i.e. \emptyset) information cannot flow through this converging connection. Thus, the two hypotheses on the grade of malignancy (MaligGr and Mk-fine) are d-separated given \emptyset . Marginal independence between two hypotheses is not the general case, since we often have measurements (evidence) on some or all features HF_j^I .

$\text{MaligGr} \parallel \text{Mk-fine} \mid \{HF_1^I, \dots, HF_m^I, HF_n^I\}$, i.e. MaligGr and Mk-fine become dependent, if either HF_1^I, \dots, HF_m^I or HF_n^I (as descendent of HF_m^I) are known, since condition 2) of the d-separation definition is violated.

In the converging connection, HF_j^I are instantiated by evidence, meaning dependence between MaligGr and Mk-fine. This dependence is actually useful for the choice of treatment.

Acknowledgements. We would like to acknowledge HUGIN Expert A/S for giving the authors access to a release candidate of HUGIN Researcher 6.7, as well as Anders Madsen and the anonymous reviewers for their comments stimulating the improvement of this paper.

References

- Andreassen S., Benn J.J., Hovorka R., Olesen K.G. and Carson E.R. (1994). A probabilistic approach to glucose prediction and insulin dose adjustment: description of metabolic model and pilot evaluation study. *Computer Meth. Programms Biomed.* 41, 153-165.
- Astion M.L. and Wilding P. (1992). The application of backpropagation neural networks to problems in pathology and laboratory medicine. *Arch. Pathol. Lab. Med.* 116, 995-1000.
- Bailey P. and Cushing H. (1926). A classification of tumors of the glioma group on a histogenetic basis with a correlation study of prognosis. J.B. Lippincott. Philadelphia.
- Brodbeck A., Iglesias-Rozas J.R. and Zell A. (1997). Classification of oligodendrogliomas using neural networks. In: Classification and knowledge organization. Klar R. and Opitz O. (eds). Springer-Verlag. Berlin. pp 434-440.
- Bruner J.M., Inouye L., Fuller G.N. and Langford L.A. (1997). Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. *Cancer* 79, 796-803.
- Burke H.B., Goodman P.H., Rosen D.B., Henson D.E., Weinstein J.N., Harrell F.E. Jr, Marks J.R., Winchester D.P. and Bostwich D.G. (1997). Artificial networks improve the accuracy of cancer prediction. *Cancer* 79, 857-962.
- Burnside E.S., Rubin D.L. and Shachter R.D. (2004). Improving a Bayesian network's ability to predict the probability of malignancy of microcalcifications on mammography. In: Proceedings computer assisted radiology and surgery 2004. Lemke H.U., Inamura K., Doi K., Vannier M.W., Farman A.G. and Reiber J.H.C. (eds). Elsevier. Amsterdam. pp 1021-1026.
- Burnside E.S., Rubin D.L., Shachter R.D., Sohlich R.E. and Sickles E.A. (2004). A probabilistic expert system that provides automated mammographic-histologic correlation: initial experience. *Am. J. Roent.* 182, 481-488.
- Christy P.S., Tervonen O., Scheithauer B.W. and Forbes G.S. (1995). Use of a neural network and a multiple regression model to predict histologic grades of astrocytoma from MRI appearances. *Neuroradiology* 37, 89-93.
- Conns P., Jhonson P., Scheithauer B., Yates A. and Pearl D. (1997). Improving diagnostic accuracy and interobserver concordance in the classification and grading of Primary Gliomas. *Cancer* 79, 1381-1393.
- Conroy J.M., Kolda T.G., O'Leary D.P. and O'Leary T.J. (2000). Chromosome identification using hidden Markov Models: Comparison with Neural Networks, Singular Value Decomposition, Principal Components Analysis, and Fisher Discriminant Analysis. *Laboratory Investigation* 80, 1629-1641.
- Cowell R.G., Dawid A.P., Lauritzen S.L. and Spiegelhalter D.G. (1999). Probabilistic networks and expert systems. Springer. Berlin.
- Cruveilhier J.C. (1829/1835). Anatomie pathologique du corps humain. Baillière. Paris.
- Cruz-Sánchez F.F., Iglesias J.R., Rossi M.L., Cervós-Navarro J., Figols J. and Haustein J. (1988). Histological characterization of 41 Ependymomas with the help of a personal computer. *Cancer* 62, 150-162.
- DeAngelis L.M. (2001). Brain tumors. *New Engl. J. Med.* 344, 114-123.
- Decaestecker C., Salmon I., De Witte O., Camby I., Van Ham P., Pasteels J.L., Brotchi J. and Kiss R. (1997). Nearest-neighbour classification for the identification of aggressive versus non-aggressive low-grade astrocytic tumors by means of image-cytometry-generated variables. *J. Neurosurg.* 86, 532-537.
- Duda R.O. and Hart P.E. (1973). Pattern classification and scene analysis. John Wiley & Sons. New York.
- Dytch H.E. and Wied G.L. (1990). Artificial neural networks and their use in quantitative pathology. *Anal. Quant. Cytol. Histol.* 12, 379-393.
- Friedman N., Geiger D. and Goldszmidt (1997). *Machine Learning* 29, 131-163.
- Hand D. and Till R. (2001). A simple generalisation of the area under the ROC Curve for multiple class classification problems. *Machine*

Causal probabilistic modeling for malignancy grading

- Learning 45, 171-186.
- Hansemann D. von. (1920). Das Problem der Krebsmalignität. *Z. Krebsforsch.* 17, 172-191.
- Henschen F. (1955). Tumoren des ZNS und seiner Hüllen. In: *Nervensystem (Handbuch der speziellen pathologischen Anatomie und Histologie)*, vol. XIII/3, Scholz W. (eds). Springer. Berlin. pp 413-440.
- Horvitz E., Cooper G.F. and Heckerman D. (1989). Reflection and action under scarce resources: Theoretical principles and empirical study. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. Morgan Kaufmann. Detroit. pp 1121-1127.
- Iglesias J.R., Aruffo C., Esparza J. and Kazner E. (1988). Histological grading of brain tumours. *EDV Medizin Biologie.* 19, 38-44.
- Iglesias-Rozas J.R. and Esparza J. (1998). Histological grading of oligodendrogliomas with the help of a polynomial abductive network. *Eur. J. Pathol.* 4, 981-03.
- Iglesias J.R., Sánchez J., Sandra A. and Mohnhaupt A. (1983). Computer model of archive and diagnosis of brain tumours based on the WHO classification. *EDV in Medizin und Biologie* 14, 40-44.
- Iglesias J.R., Kazner E., Aruffo C. and Esparza J. (1986a). A model semiautomatic type-specific CT diagnosis of brain tumors. *Mathematical fundamentals and practical application.* *Br. J. Radiol.* 59, 895-900.
- Iglesias J.R., Pfannkuch F., Aruffo C., Kazner E. and Cervós-Navarro J. (1986b). Histopathological diagnosis of brain tumors with the help of a computer: mathematical fundamentals and practical application. *Acta. Neuropathol.* 71, 130-135.
- Jensen F.V. (2001). *Bayesian Networks and Decision Graphs*. Springer. New York.
- Kaye A. and Walker D. (2000). Low grade astrocytomas: controversies in management. *J. of Clin. Neurosci.* 7, 475-483.
- Kernohan J.W. and Sayre G.P. (1952). Tumors of the central nervous system. In: *Atlas of tumor pathology. Fascicle 35.* Armed Forces Institute of Pathology. Washington D C.
- Kjærulff U.B. and Madsen A.L. (2005). *Probabilistic Networks - An Introduction to Bayesian Networks and influence diagrams.* Unpublished manuscript available from <http://developer.hugin.com/Publications/>.
- Kleihues P. and Cavenee W.K. (2000). World health organization classification of tumors pathology and genetics. *Tumours of the nervous system.* IARC Press. Lyon.
- Kolles H., von Wangenheim A., Rahmel J., Niedermayer I. and Feiden W. (1996). Data-driven approaches to decision making in automated tumor grading. An example of astrocytomas grading. *Anal. Quant. Cytol. Histol.* 18, 298-304.
- Kroh H., Iglesias J.R., Matyja E., Aruffo C., Maeir-Hauf K. and Cervós-Navarro J. (1988). Computer classification of experimental brain tumors in mice. *Exp. Pathol.* 35, 41-46.
- Langley P., Iba W. and Thompson K. (1992). An analysis of Bayesian classifiers. In: *Proceedings of 10th national conference on artificial intelligence.* AAAI Press. Menlo Park. pp 223-228.
- Lauritzen S.L., Dawid A.P., Larsen B N. and Leimer H.G. (1990). Independence properties of directed Markov fields. *Networks* 20, 491-505.
- Lennernäs B., Sandberg D., Albertsson P., Silén Å. and Isacson U. (2004). The effectiveness of artificial neural networks in evaluating treatment plans for patients requiring external beam radiotherapy. *Oncol. Rep.* 12, 1065-1070.
- Louis D.N., Holland E.C. and Cairncross J.G. (2001). Glioma classification: a molecular reappraisal. *Am. J. Pathol.* 159, 779-794.
- Lucas P. (2001) Bayesian networks in medicine: a model-based approach to medical decision making. In: *Proceedings of the EUNITE workshop on Intelligent Systems in patient Care*, Adlassnig K-P. (ed). Vienna. pp 73-97.
- Madsen A.L., Lang M., Kjærulff U., and Jensen F. (2003). The hugin tool for learning bayesian networks. In: *Proceedings of the seventh european Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty.* pp 549-605.
- Martin H., Voss K., Hufnagl P. and Fröhlich K. (1984). Automated image analysis of gliomas - an objective and reproducible method for tumor grading. *Acta Neuropathol.* 63, 160-169.
- Nathwani B.N., Heckerman D.E., Horvitz E.J. and Lincoln T.L. (1990). Integrated expert systems and videodisc in surgical pathology: an overview. *Human Pathol.* 21, 11-27.
- Olesen K.G., Kjærulff U., Jensen F., Falck B., Andreassen S. and Andersen S.K. (1989). A MUNIN network for the median nerve - a case study on loops. *Appl. Artificial Intelligence* 3, 384-403.
- Papik K., Molnar B., Schaefer R., Dombovari Z., Tulassay Z. and Feher J. (1998). Application of neural networks in medicine - a review. *Med. Sci. Monit.* 4, 538-546.
- Pearl J. (1988). *Probabilistic reasoning in intelligent systems.* Morgan Kaufmann. Detroit.
- Prayson R.A., Agamanolis D.P., Cohen M.L., Estes M.L., Kleinschmidt-DeMasters B.K., Abdul-Karin F., McClure S.P., Sebek B.A. and Vinay R. (2000). Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. *Neurol. Sci.* 175, 33-39.
- Ringertz N. (1950). Grading of gliomas. *Acta Pathol. Microbiol. Scand.* 27, 51-64
- Scarpelli M., Montironi R., Magi Galluzzi C. and Diamanti L. (1994). Quantitative evaluation of nucleolar features on cytologic material in brain tumor diagnosis. *Clin. Neuropathol.* 13, 323-328.
- Schröder R., Müller W., Bonis G. and Vorreit W. (1970). *Statistische Beiträge zum Grading der Gliome. III: Astrozytome und Oligodendrogliome.* *Acta Neurochirurgica* 23, 1-29.
- Shachter R., Eddy D., Hasselblad V. and Wolpert R. (1987). A heuristic bayesian approach to knowledge acquisition: Application to the analysis of tissue-type plasminogen activator. In: *Proceedings of the 3rd Annual Conference on Uncertainty in Artificial Intelligence (UAI-87).* Elsevier Science., New York. pp 183-190.
- Spiegelhalter D. and Lauritzen S. (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks* 20, 579-605.
- Spirtes P., Glymour C. and Scheines R. (2000). *Causation, Prediction, and Search.* 2nd edition. MIT Press. Cambridge.
- Steck H. (2001). *Constrained-based structural learning in Bayesian Networks using finite data sets,* PhD Thesis. Institut für der Informatik der Technischen Universität München.
- Steck H. and Tresp V. (1999). Bayesian Belief Networks for data mining. In: *Proceedings of the 2nd Workshop on Data Mining und Data Warehousing als Grundlage Moderner Entscheidungsunterstuezender Systeme, DWDW99.* Universität Magdeburg. pp 145-154.
- The Childhood Brain Tumor Consortium (1989). *Journal of Neuro-Oncology* 7, 211-224.
- Wei J.S., Greer B.T., Westermann F, Steinberg S.M., Son C-G., Qing-Rong Chen Q-R., Whiteford C.C., Bilke S., Krasnoselsky A.L.

Causal probabilistic modeling for malignancy grading

- Cenacchi N., Catchpoole D., Berthold F., Schwab M. and Javed K.J. (2004). Prediction of Clinical Outcome Using Gene Expression Profiling and artificial Neural Networks for Patients with Neuroblastoma. *Cancer Res.* 64, 6883–6891.
- Weidl G. (2002). Root cause analysis and decision support on process operation, industrial applications of Bayesian Networks and influence diagrams. Technology Doctoral Dissertation. ISBN 91-88834-36-0, Mälardalen University, Sweden.
- Zülch K.J. (1978). Principles of the new WHO classification of brain tumors. *J. Neurosurg. Sci* 22, 1-5.
- Zülch K.J. (1979). Histological typing of tumours of the central nervous system. World Health Organization, Geneva (International histological classification of tumours, no. 21)
- Zülch K.J. (1980a). Principles of the new World Health Organization (WHO) classification of brain tumours. *Neuroradiology* 19, 59-66.
- Zülch K.J. (1980b). Pathologie und Biologie der raumfordernden Prozesse von Rückenmark und Wirbelsäule. *Radiologe* 20, 459-465.
- Zülch K.J. (1986). Brain tumors. Their biology and pathology. 3rd ed. Springer- Verlag, Berlin, Heidelberg, New York, Tokyo. Chapter 2. Biological Behavior and Grading (Prognosis). pp 27-40.

Accepted February 14, 2007