Visual Verification of Cancer Staging for Therapy Decision Support

M. A. Cypko 1, J. Wojdziak 2,3, M. Stoehr 4, B. Kirchner 3, B. Preim 5, A. Dietz 4, H. U. Lemke 1, and S. Oeltze-Jafra 1

1ICCAS, Medical Faculty, University of Leipzig, Germany
2Media Design, Technische Universität Dresden, Germany
3Gesellschaft für Technische Visualistik mbH, Germany
4Head and Neck Department, University of Leipzig, Germany
5Department of Simulation and Graphics, University of Magdeburg, Germany

Abstract
It is generally accepted practice that each cancer patient case should be discussed in a clinical expert meeting, the so-called tumor board. A central role in finding the best therapy options for patients with solid tumors plays the Tumor, lymph Node, and Metastasis staging (TNM staging). Correctness of TNM staging has a significant impact on the therapy choice and hence on the patient’s post-therapeutic quality of life or even survival. If inconsistencies in the TNM staging occur, possible explanations and solutions must be found based on the complex patient records, which takes the costly time of (multiple) physicians. We propose a more efficient visual analysis component, which supports a physician in verifying the given TNM staging before forwarding it to the tumor board. Our component comprises a Bayesian network model of the TNM staging process. Using information from the patient records and Bayesian inference, the model computes a patient-specific TNM staging, which is then explored and compared to the given staging by means of a graph-based visualization. Our component is implemented in a research prototype that supports an understanding of the model computations, allows for a fast identification of important influencing factors, and facilitates a quick detection of differences between two TNM stagings. We evaluated our component with five physicians, each studying 20 cases of laryngeal cancer.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—Applications; J.3 [Computer Applications]: Life and Medical Sciences—Life and Medical Sciences

1. Introduction
Multidisciplinary expert meetings in oncology (also called tumor boards) improve the medical management of cancer patients [SP06]. Clinical studies from, e.g., head and neck [WMZ10], lung [LVP07], and colorectal cancer [SW08], have demonstrated and quantified this improvement. In a tumor board, many clinical experts contribute their expertise and formulate a comprehensive treatment plan. For patients with solid tumors, staging the cancer plays a central role in finding the best therapy options. Cancer stagings relate to classifications of primary cancer extents and spreads. The UICC (Union for International Cancer Control) and the AJCC (American Joint Committee on Cancer) jointly develop and maintain the Tumor, lymph Node, and Metastasis staging standard (TNM staging) [SGW11].

TNM staging is performed by the attending physician based on the fragmented and complex results of multidisciplinary diagnostic examinations stored in the Hospital Information System (HIS) and the paper-based medical record (PBMR). The TNM classification code is saved then to HIS and PBMR. An inaccurate code may result from pending, missing, or overlooked information. In preparation of the tumor board, the code is copied by a physician to a tumor board case sheet together with other patient-specific information. Inaccuracies in the TNM staging may remain unnoticed until the tumor board. Their discussion in the meeting, and generally an explanation of the TNM staging requested by a clinician, require a lengthy search in HIS and PBMR taking the costly time of multiple experts.

We propose a more efficient visual analysis component, which supports the physician in verifying the given TNM staging before forwarding it to the tumor board. Furthermore, it prepares the physician for potential questions asked in the meeting. Our component comprises a Bayesian network (BN) model of the TNM staging process. Using information from the patient records and Bayesian inference, the model computes a patient-specific TNM staging, which is then explored and compared to the given staging by means of a graph-based visualization. Our component is implemented in a research prototype that supports an understanding of the model computations, allows for a fast identification of important influencing factors, and facilitates a quick detection of differences between two TNM stagings. We evaluated our component with five physicians, each studying 20 cases of laryngeal cancer.
Existing application-independent BN tools, such as the widespread GeNi/SMILE [Dru99], focus on the modeling itself and require BN knowledge, which is not part of medical education. BN approaches in the context of clinical decision support mainly focus on model and prediction correctness (see [SNBA+13] for a survey). Hence, we have developed a dedicated graph view for the visual analysis of BN model computation results and their comparison with a given outcome. The view is integrated in a prototype for visual TNM verification. In summary, our contributions are:

- Interactive graph view for structured TNM staging inspection
- Tailor-made glyphs for encoding probabilities
- Interaction techniques for exploring complex BN graphs
- Comparative visualization of two TNMs
- Prototype for visual TNM verification

Our working example throughout the paper is decision support for laryngeal cancer treatment in head and neck oncology. While the TNM staging and hence the BN model are specific for this application domain, our component is directly transferable to other domains. In Section 2, we provide further details of TNM staging and introduce BNs. In Section 3, we describe the clinical decision making workflow incorporating TNM staging and we deduce requirements on a TNM staging verification embedded in this workflow. In Section 4, we discuss related work. In Section 5, we describe our visual verification component and its prototypical implementation. In Section 6 and 7, we elaborate on the evaluation of the prototype and present the evaluation results. Finally, in Section 8 and 9, we discuss the limitations of our component and conclude with a brief summary and thoughts on future work.

2. Medical and Technical Background

We provide details of TNM staging in head and neck oncology and briefly introduce BNs.

2.1. TNM Staging

In head and neck oncology, several clinical practice guidelines exist. The National Comprehensive Cancer Network (NCCN) guidelines provide widely accepted evidence-based treatment recommendations [PSB+14, Nat16]. For laryngeal cancer, NCCN’s recommendations are based on the TNM staging system, which describes the extent of the primary tumor and its infiltration of surrounding structures (T), the affection of locoregional lymph nodes (N), and distant metastases (M) [SGW11]. Each of the three parameters is assigned a range of possible values:

- \( T \in \{ T_x, T_0, Tis, T_1, T_2, T_3, T_4 \} \),
- \( N \in \{ N0, Nx, N1, N2, N3 \} \), and
- \( M \in \{ M0, M1 \} \).

The suffix \( x \) indicates that the tumor and the lymph nodes, respectively cannot be evaluated. \( Tis \) represents a carcinoma in situ, whose classification as cancer is subject to an ongoing debate. The numerical values represent the course from low- to high-grade cancer. The TNM staging of laryngeal cancer allows for additional values facilitating a more fine-grained gradation [PSB+14].

2.2. Bayesian Networks

A BN [Pea98] is a probabilistic graphical model that has the potential to represent highly complex decisions with all relevant characteristics, and infer case-specific decision options. In detail, a BN model is described by a directed acyclic graph, which consists of (1) a set of nodes, (2) directed edges between pairs of nodes, and (3) conditional probability tables (CPT). Nodes represent random variables, each with a set of possible states. Directed edges represent direct causal dependencies (from a cause to its effect). Conditional probabilities are required for each node quantifying influences from the nodes’ direct causes. Based on a BN decision model, observations (patient information) can be set for a subset of the model’s variables, and the network computes for this specific case the probability distributions over all unobserved variables. Therefore, a BN may allow, e.g., for studying potential therapy options, their outcomes as well as relevant decision influences and identifying unobserved variables with a high diagnostic value.

Our long-term vision is to support tumor board therapy decisions by BN models. Stoehr et al. [SCD+14] presented such a model in the context of laryngeal cancer treatment (Fig. 1). In this paper, we use the model’s recently validated sub-network of TNM staging [CSK+17]. The model consists of 303 variables (nodes) with 334 dependencies (edges). Some of the variables comprise thousands of conditional probabilities.

3. Clinical Workflow and Requirement Analysis

We describe the clinical decision making workflow incorporating TNM staging and deduce the requirements on a TNM staging verification and its embedding in this workflow.

3.1. TNM Staging in Clinical Decision Making Workflow

An overview of the clinical decision making workflow in head and neck oncology is shown in Figure 2 based on [PWN+15, RDD+11]. The initial consultation and examination step comprises the anamnesis and multidisciplinary medical exams and imaging procedures carried out adhering to the NCCN guidelines. For instance, a panendoscopy is conducted to evaluate the resectability of the tumor and to collect information required for the TNM staging.

The following initial treatment assessment step by the attending physician includes a TNM staging and the composition of a diagnostic report incorporating an initial treatment suggestion (Sec. 2.1). The TNM classification code and the report are stored in the HIS and appended to the PBMR. In preparation of the next step – the tumor board – a physician copies all relevant information, including the TNM staging and the treatment suggestion, for 10 to 15 patients from HIS and PBMR to the tumor board case sheet.

In the tumor board meeting, all patients on the case sheet are discussed. Inconsistencies in the TNM staging, which result from pending, missing or overlooked information, delay the tumor board and cannot always be resolved. Thus, cases may have to be repeatedly introduced to the board after correction (arrow 1).

After the tumor board, the therapy decision and possible complications are communicated to the patient in a consultation. If the
patient disagrees with the decision, an alternative treatment is discussed. In rare cases, a reintroduction in the next tumor board may be necessary (arrow 2). If the patient agrees, the therapy starts and may include one or multiple TNM restagings to monitor the treatment success. In case of a surgical approach, the patient case is discussed again in the tumor board to decide on a possible adjuvant therapy (arrow 3). Otherwise, the patient is followed up usually for five years (arrow 4).
Section 4.2. The overview visualization of large CPTs has been approached by Chiang et al. [CSLG08]. It is not required in our case, however, since we are rather interested in the outcome of Bayesian inference based on the CPTs. This reduces the challenge to visualize a probability distribution per node, i.e., one computed probability per state of the associated variable. Understanding conditional probabilities is inherent to Bayesian reasoning and clinical decision making. Since users perform poorly at this task, assistance by visual aids has been investigated [MDF12, OPH16].

Hasse diagrams are the earliest approach to visualize causalities represented by a directed-acyclic graph. However, they fail in conveying causal semantics and are hard to read for complex graphs due to many edge intersections. A widespread framework for creating and analyzing BN models is GeNi/SMILE [Dru99]. On demand, the GeNi graphical user-interface draws a bar chart glyph for each user-selected node to show the probability distribution. The direction of causality is indicated by directed arcs connecting the nodes. Koiter [Koi06] extended GeNi such that causal strength is mapped to the thickness of the connecting arcs and the causal sign – a cause can have a positive or a negative effect – is mapped to their color. Zapata-Rivera et al. [ZRNG99] developed techniques for mapping marginal probabilities to node color or size and for adjusting node proximity according to causal strength.

Kadaba et al. [KIL07] presented dynamic representations of causal semantics. For instance, causal direction and sign are encoded by circles on edges that move along the direction of causality, leading to a growing (+) or shrinking (−) effect node. The Growing Polygons technique of Elmqvist and Tsigas [ET03] is tailored to visualize dynamic processes and provides an alternative to traditional node-link diagrams. In an animation, it depicts the gradual change in a set of connected causal events, the processes, by animating size changes of n-sided polygons.


4.2. Graph Drawing and Interaction

Major challenges in drawing the large graph of a BN model (cf. Fig. 1) are the avoidance of node overlaps and edge crossings while at the same time generating edges with approximately equal length. In our case, the readability of node names and of information encoded by glyphs representing the nodes as well as the requirements R3 and R7 pose additional challenges.

In order to gain screen space for the glyph display, we adopt the graph aggregation strategy of Moscovitch et al. [MCH09]. It is based on the pre-defined node hierarchy of the BN model. Sub-networks can be opened and collapsed facilitating an iterative and interactive level-of-detail exploration concurrent with requirement R7. Alternative approaches are the zoom mechanisms and local magnification techniques proposed by Sarkar and Brown [SB92] and Furnas [Fur86]. They show detail while maintaining an overview of the graph structure. However, they offer a less structured exploration compared to aggregation strategies and require frequent pan&zoom.

To fulfill requirement R3, we define a focus region in the center of the graph view (Fig. 3). This is inspired by theBring Neighbors Lens of Tominski et al. [TAVHS06] which dynamically adjusts the graph layout to show local connectivity. Alternatively, Munzner [Mun97] suggests to use a hyperbolic focus+context view.

BN graphs are generally drawn in a topological layout from cause to effect. The resulting layout is often imbalanced and suffers from many edge crossings (cf. Fig. 1). In accordance with Coleman and Parker [CP96], we aim at a better readable and more aesthetic layout that is symmetrical, minimizes edge crossings, and produces edges of approximately equal length. Thus, we employ the widespread force-based method by Fruchterman and Reingold [FR91] and adapt it by introducing two constraints. First, nodes/glyphs must not overlap with the focus region and the nodes currently in focus. Second, nodes must be drawn close to the canvas border to guarantee maximum space for the nodes in focus and the display of their probability values (Fig. 4).

An important aspect of our graph drawing is an appropriate glyph size that guarantees readability. Each glyph and its associated text label can be considered as an object. A packing strategy is then required that optimally distributes all objects within a given area. Lodi et al. [LMM02] provide a general taxonomy of packing principles not specifically concerned with graphical models. We adapt the method of Donev et al. [DSCT04], but employ bounding rectangles instead of ellipsoids to approximate object shape. For a given number of nodes, we compute the maximum rectangle sizes such that all objects can be packed within the given rectangular canvas minus the focus region. This process is repeated once the number of nodes changes, e.g., a node is dragged into the focus region.

Conclusion. Existing approaches to visualize BN models show the entire graph and require exhaustive panning & zooming for the exploration of large graphs. In favor of a graph aggregation strategy, we decide to facilitate a more structured analysis. Instead of the traditional topological layout in Figure 1, we employ a screen-filling circular layout to create a more visually balanced representation.

5. Visual Verification of TNM Staging

We present our visual analysis component and demonstrate its application to a patient case of laryngeal cancer.

5.1. Visual Analysis Component

The visual analysis component has been developed in due consideration of the requirements in Section 3.2. Its functional workflow, overall design, and the compliance with the requirements are summarized in the following and detailed in Section 5.2.

The visual analysis component is integrated into a therapy decision support system based on BN models [CSK17]. The graphical interface of our component is separated into four areas (Fig. 3): a graph view (1), a basic patient information summary (2), a variable selector (3), and a tool set (4). For decision support, a user selects
both the TNM staging model and a patient case. On this request, the BN receives the required patient data from patient databases, and computes the desired patient-specific BN (PSBNs). Next, the user uploads a PSBN into the visual analysis component. The basic information summary then shows the patient’s name, date of birth and id number, and is always visible during the analysis for double-checking the correct patient case at any time (R1).

The graph view initially presents the PSBN’s decision graph restricted to the nodes of the T, N, and M states and only their direct influences, i.e. adjacent nodes (R3). The node identifiers are displayed next to the nodes using a minimum font size that guarantees legibility (R2). The probability distribution as well as the type of information (observed or computed) are encoded by the node filling pattern (R6). On demand, detailed information of the probabilistic values can be displayed for individual nodes (R5, Fig. 4).

The user can verify a given TNM staging by comparing it with a computed staging. Instantiating a comparison causes a colored emphasis of influences/nodes, which exhibit a difference in probability distribution (R7, blue = given staging and yellow = computed staging). To better understand these differences, the user can drag the nodes from proximity into focus, thereby triggering the display of their associated influences.

The variable selector comprises a list view of all variable identifiers and a search bar which may be employed for quick access (R4). The tool set allows for switching between two different graph analysis modes: model view and value comparison view.

5.2. Graph View Design

The graph view provides a visual representation of the PSBN’s decision graph for automatic TNM classification.

Graph Layout. In the initial display of the graph, the nodes of the T, N, and M states are displayed in the center and strung along a ring representing the circular focus region (Fig. 3). Only their adjacent nodes (R3). The node identifiers are displayed next to the nodes using a minimum font size that guarantees legibility (R2). The probability distribution as well as the type of information (observed or computed) are encoded by the node filling pattern (R6). On demand, detailed information of the probabilistic values can be displayed for individual nodes (R5, Fig. 4).

The user can verify a given TNM staging by comparing it with a computed staging. Instantiating a comparison causes a colored emphasis of influences/nodes, which exhibit a difference in probability distribution (R7, blue = given staging and yellow = computed staging). To better understand these differences, the user can drag the nodes from proximity into focus, thereby triggering the display of their associated influences.

The variable selector comprises a list view of all variable identifiers and a search bar which may be employed for quick access (R4). The tool set allows for switching between two different graph analysis modes: model view and value comparison view.

Node Glyphs. The nodes of the graph are represented by circular glyphs of equal size. Circles suggest the completeness of an entity and the radial distribution of their possible anchor points for edges offers a high flexibility. This is particularly important in interactive visualizations where nodes can be dragged around freely and anchor points must move along.

Each glyph is partitioned into equally-sized slices starting at 12 o’clock and proceeding in a clockwise fashion (Fig. 4). The states and hence the slices are sorted from best to worst, e.g., from no to high-grade larynx T state. The number of the slices corresponds to the number of states of the represented variable plus an “absurd” state. The latter is added to decision nodes in order to reflect intercepting probabilities of paradox influences (e.g., a tumor does not exist if surrounding tissue is infiltrated). Paradox influences are caused by incorrect or incomplete patient data. The TNM model contains three decision nodes corresponding to the T, N, and M state.

Node Glyph Filling. The sliced display can be interpreted as a circular bar chart showing the probability distribution of the corresponding variable adding up to an overall of 100%. We tested two approaches to encode the probabilities: 1) filling each slice proportionally to the associated absolute probability value and 2) filling the slice with highest probability value completely and the remaining ones proportionally to this value. Both approaches offer trade-offs between legibility and comparability. 1) allows for a comparison of fill levels within and across nodes, but suffers from lower legibility of small values and values close to 100%, which may be misinterpreted as fully filled, i.e. “absolutely certain”. 2) increases the visibility of small values thereby improving both the legibility of the probabilities and the comparability of probability values within nodes. However, it renders the comparability across nodes impossible. Moreover, it may cause the same misinterpretation of absolute certainty for a fully filled slice, which is however less likely due to other slices also being filled. Despite an anticipated slower learning curve, we give preference to 2), since the legibility of individual states was rated as more important than the comparability across nodes in our requirement analysis (R2, Sec. 3.2). In order to compensate for the loss in comparability and for an inspection of the absolute probability values, the user may click on a glyph causing an integration of its detail view shown in Figure 4 into the graph view (Fig. 7 (2)).

In order to distinguish nodes representing observed values, e.g., medical examination results, from those representing computed values, the former are shown with a small inner circle being cut out (Fig. 5 (a)) This loss in potential encoding space is acceptable for the former. As of now, uncertainty is not included as a concept in the TNM staging system. Hence, an observation in the clinical context was either not made or it was made with a 100% outcome/state resulting in a single clearly perceivable, fully filled slice segment. A single fully filled slice in a glyph representing computed information conveys a certain prediction (Fig. 5 (b)) while various partially filled slices convey an uncertain prediction (Fig. 5 (c)).
TNM Model Comparison. The model comparison mode allows for the visual verification of the TNM staging given in the patient record by comparing it to a computed TNM staging. The comparison requires two runs of TNM model computation. (1) In the first run, the T, N and M state of the patient record are employed. The model computes the probability distributions for all internal nodes between the leaves of the PSBN graph (examination methods) and its roots (decisions, i.e. the TNM states). We refer to these internal nodes as patient states. (2) In the second run, the patient states and also the TNM states are computed based on the observations. A comparative visualization of the runs’ results facilitates the detection of discrepancies (Fig. 3). Two colors are used to distinguish between run (1) based on the given clinical TNM (blue) and run (2) computing the TNM (yellow). The node glyph filling is adapted such that it reflects different degrees of agreement between the results (Fig. 6).

In case of a high degree of agreement of more than 90%, the filling of the glyph’s slices is set to black in order to avoid their visual emphasis (Fig. 6 (a)). This threshold has been obtained from the network modeling. BNs perform reasonably accurate even with a less precise model quantification [OD13]. Therefore, the model’s probability parameters were assessed by seven probabilistic values from 1% to 99% in irregular 10% to 15% steps [CHK15]. For this reason, we also accept an inaccuracy of up to 10% in the PSBNs. For lower degrees of agreement, the slices are filled and distinctly colored according to the results of the two model runs (Fig. 6 (b-d)). In c) for instance, the T-state given from the patient record was T3, resulting in a 100% blue filling of the respective slice. The computed T-state (yellow), however, is uncertain assigning only a probability of 53.4% to T3. Please note the blue, semi-transparent center circle in all glyphs of a) to d). It indicates that the corresponding information was set by the physician in run (1), similar to indicating an observed medical examination by a blank circular cut-off as explained before.

Graph View Interaction In order to browse the PSBN hierarchy, the user may drag additional nodes into the focus region causing their arrangement on the focus ring and the display of their adjacent nodes in the proximity. In order to remove a node from focus, the user drags the node away from the ring causing a disappearance of its direct neighbors. A node may be clicked causing a detailed
Figure 4: Detail view of a node glyph representing the T-state variable. The possible states of the variable are represented by slices and sorted from best (no tumor; T0) to worse (high-grade tumor; T4b) in a clockwise fashion starting at 12 o’clock. A segment’s fill level indicates the associated relative probability value. A label per segment shows the identifier of the state and the absolute probability value.

Figure 5: Example glyphs for the encoding of observed (a) and computed information (b,c). Observed nodes have one completely filled slice and an empty inner circle. Computed nodes have one completely filled slice of the most probable state while the remaining slices are filled relative to the most probable state.

Figure 6: Example glyphs for the comparison of two TNM model computations (blue and yellow) showing different degrees of agreement of the computed probability distributions: a) high, b) acceptable, c) low, and d) no agreement. A high degree of agreement is indicated by a black slice filling. Detailed information regarding the encoding is given in the text.

5.3. Verification Workflow

We demonstrate the visual analysis component for a verification workflow comprising eight steps (Fig. 7).

(1) The verification starts with the nodes of T, N, and M state being in focus and their adjacent nodes in proximity.

(2) The physician gains a quick overview of the clinically predetermined T, N, and M state, which were imported from the patient record into the model. He or she clicks on individual nodes to further investigate the states (T2, N2c, and M1).

(3) To verify the TNM staging, the physician switches to the comparison mode. The M state shows a very high degree of agreement (black filling), the N state an acceptable degree of agreement, and the T state a low degree of agreement: T3 inferred by the system as opposed to T2 from the patient record.

(4) The physician decides to further investigate nodes with a low degree of agreement. She starts by dragging the T state node into the focus region. This causes a disappearance of the N and M state and their adjacent nodes.

(5) She further drills down the network by repeating step 4 until she arrives at examination methods (graph leaves). Starting with the initial focus on the TNM staging, each examination method can be reached within a maximum of three focus shifts.

(6) At this point, the physician is only interested in the results of actually carried out examinations (not in the computed probabilities of the remaining ones). To highlight these, she selects the check box "observed values" in the variable selector area.

(7) She studies the examinations’ details and takes notes. While one examination reported a paralyzed vocal cord (step 7), a later examination reported a fixed vocal cord (step 8). This has a significant impact on the T state changing from T2 to at least T3, which was lighted in the list. Nodes in focus are emphasized in the list by a pulled out node name to the left.

5.3. Verification Workflow

We demonstrate the visual analysis component for a verification workflow comprising eight steps (Fig. 7).

(1) The verification starts with the nodes of T, N, and M state being in focus and their adjacent nodes in proximity.

(2) The physician gains a quick overview of the clinically predetermined T, N, and M state, which were imported from the patient record into the model. He or she clicks on individual nodes to further investigate the states (T2, N2c, and M1).

(3) To verify the TNM staging, the physician switches to the comparison mode. The M state shows a very high degree of agreement (black filling), the N state an acceptable degree of agreement, and the T state a low degree of agreement: T3 inferred by the system as opposed to T2 from the patient record.

(4) The physician decides to further investigate nodes with a low degree of agreement. She starts by dragging the T state node into the focus region. This causes a disappearance of the N and M state and their adjacent nodes.

(5) She further drills down the network by repeating step 4 until she arrives at examination methods (graph leaves). Starting with the initial focus on the TNM staging, each examination method can be reached within a maximum of three focus shifts.

(6) At this point, the physician is only interested in the results of actually carried out examinations (not in the computed probabilities of the remaining ones). To highlight these, she selects the check box "observed values" in the variable selector area.

(7) She studies the examinations’ details and takes notes. While one examination reported a paralyzed vocal cord (step 7), a later examination reported a fixed vocal cord (step 8). This has a significant impact on the T state changing from T2 to at least T3, which was lighted in the list. Nodes in focus are emphasized in the list by a pulled out node name to the left.
Figure 7: Verification workflow of a TNM staging given in a patient record (see Section 5.3 for details).
correctly inferred by the system. The TNM staging in the patient record was performed prior to the second examination. (8) Finally, the physician focuses on further nodes with high discrepancies and starts over at step 4. To find a specific node/variable without having to search the network, she uses the list or search bar in the variable selector area.

5.4. Implementation

The interactive visual analysis component is implemented as a web application, based on HTML5, SVG, CSS3, and JavaScript frameworks. In particular, AngularJS [Goo16] was used to implement the data structure, in combination with the data-driven library D3.js [Bos15] for creating the SVG elements.

6. Study Set-Up of TNM Verification

We conducted an evaluation to study the TNM model’s impact on supporting clinicians in verifying these TNM states. The following section describes the study set-up, participants and test cases.

Prior to the study, 21 patient cases were prepared; one test case to introduce our component and 20 to evaluate our component. To study the participant’s behavior in case of 1) an incorrect clinical TNM staging or 2) a system’s wrong suggestion, nine of the 20 cases have been manipulated by 1) changing the clinically recorded T, N and M states, or 2) deleting single examinations. Furthermore, we prepared a one page paper sheet explaining the visualization of nodes, which included the Figures 4, 5, and 6. Five clinicians participated in the study: three resident physicians (R1/2/3) and two senior physicians (S1/2). Resident physicians had a clinical experience of nine months, four years, and five years, whereas each senior had more than 10 years of clinical experience. All participants had no background knowledge about BNs, the TNM staging BN model, or our component. Each one answered questions about self-confidence with both staging TNM and in using computers (with a possible ranking of 1 = little, 2 = ordinary and 3 = considerable).

The study was conducted in two sessions during the clinicians’ working hours. During the study, each participant was more or less often interrupted by clinical calls and needed to leave the study for several minutes. During the study, clinicians separately evaluated TNM stagings of the 20 patient cases using only our component. A knowledge engineer introduced our component as described in Section 5.1, and presented an example patient case based on the workflow as described in Section 5.3. Each clinician was asked to 1) study the patient’s TNM states, 2) reconstruct the rationale, and 3) confirm or correct the TNM states. Upon request, participants were allowed to use a printed table of the TNM staging system. Finally, the participants took notes of the T, N, and M states and their conclusions. This procedure was repeated for the remaining 20 patient cases. Clinicians were supposed to work autonomously, but in case of a question the knowledge engineer was available.

The study was video and audio-recorded. From these records and notes we evaluated the required time for each patient case, the number of guideline requests, and the participants’ notes including correct TNM verification, conclusions, and uncertain verification and conclusions.

7. Results

We evaluated the TNM states and conclusions, the time, the participants’ questions, and their comments during the study (for more details see Table 1 and Figure 8).

Initially, all participants were introduced to our component, which took between 8 and 23 minutes depending on a participant’s requests. Afterwards, each participant required support from the knowledge engineer for the first three to four cases, and was able to work autonomously for the remaining sixteen cases. During the study, all participants requested the table of the TNM staging system, either to obtain information or to verify an assessment. Specifically, the less experienced participant R1 requested the table to conclude or verify the T, N, and M states. The experienced participants R2 and S1 requested the table more often but usually to approve their estimates. Compared to the other participants, they both were also more frequently interrupted by clinical calls. Especially participant S1 was interrupted several times at the first three cases and needed repeated introductions.

Verification took between 1.5 and 10.25 minutes, and included the participants’ requests and discussions as well as an explanation of the current procedure. On average, participants required the same time of about four minutes. From the 12th patient case, all participants verified correctly and the average time decreased to less than 3 minutes. Misunderstanding of our component’s content included distinguishing, interpreting and arguing observed and computed values. Initially, within the first patient cases, two participants wrongly assumed that all nodes represent clinically available patient information. Furthermore, two resident and one senior clinician wrongly assigned a high confidence to the clinically available information, which are always presented in the model with 100%.

Finally, all participants felt comfortable with the space-filling arrangement of nodes. Also, after presenting the informational paper sheet with visual node descriptions, the clinicians understood the node design. However, initially, one participant related the circle bar slices to a tumor size. Therefore, he expected information about larger tumors at Boolean valued nodes (e.g., the M state node) compared to nodes with more values (e.g., the T state node). The participant explained this initial conjunction by studies of tumors in clinical images, which appear as usually round, black and differently sized areas.

8. Discussion

Our component enables an autonomous TNM verification (R8 in Section 3.2). Furthermore, participants were able to verify or correct a TNM staging in a reasonable time of 4 minutes on average (R9). In the following, we discuss the study results as well as issues and future work related to our component’s layout, functionality and its applicability.

Study results and limitations The study shows that a user can successfully use our component and correctly verify a TNM staging after about 10 training cases. Anecdotal feedback by the clinicians indicates that the component promotes an understanding of the represented decisions and their influences. A comparison of our component’s efficiency to traditional methods is difficult since no traditional verification methods exist, except
Table 1: Personal details (self-assessments) and results from the five participants of the verification study.

<table>
<thead>
<tr>
<th>Part. no.</th>
<th>Clinical experience (in years)</th>
<th>Experiences with TNM staging (1=little; 2=ordinary; 3=considerable)</th>
<th>Computer skills (1=little; 2=ordinary; 3=considerable)</th>
<th>Requests for TNM guidelines</th>
<th>Averg. time per case in min</th>
<th>Complete verification (from 20 possible)</th>
<th>Correct stagings (from 20 possible)</th>
<th>Correct conclusions (from 20 possible)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.75</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3.8</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>R2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3.7</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>R3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4.7</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>S1</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4.6</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>S2</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3.6</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 8: Study results of the five participants (R1, R2, R3, S1, and S2) with 21 patient cases; one to introduce our component and 20 for the evaluation. This diagram combines a bar chart and a table presenting the required time for each patient case and the answer quality of stagings with conclusion.

for studying the patient records. Patient data are usually not verified prior to the multidisciplinary expert meetings. However, with one clinician (author of this paper) we tested the TNM verification of ten patient cases using two other methods; first, using regular electronic and paper-based patient records and, second, using a list with all of the network’s nodes and their probabilities. In general, each patient record provides on average 78 information items (ranging between 36 and 154 items) [CSK∗17]. Specifically, a patient record provides the information usually in running text through multiple pages of performed examinations. A network’s list provides 12 pages of information for all 303 variables including the available patient information and probabilistic parameters. Using the lists, the clinician stopped the verification after the first case being overwhelmed by too many list items. Using the patient records, the TNM verification took the clinician an average of 10 minutes per patient case (ranging between 6 and 15 minutes). Studying the TNM stagings was exhausting due to the extensive documents. In all 10 cases the clinician verified correctly. However, the verification time was more than twice as long as with our component. Further, the chance of overlooking information is higher than with our guided approach.

All studies were performed in a relaxed atmosphere without consequences in case of mistakes. This limitation should be tackled in a prospective study by letting the clinicians present their verification results at the tumor board.

**Layout** Clinicians benefit from a clear overview and a simple exploration. They neither expect nor desire a causal graph structure. On the one hand, clinicians know the direct influence between diagnostic information due to their daily practice of studying guidelines and making complex decisions. On the other hand, we experienced a clear understanding of the variable relations from the participants’ workflow and explanations.

After a short learning phase, the node glyphs are easy to read and understand. However, Boolean valued nodes appear prominently in the graph view due to their large slices. Furthermore, a Boolean node that is set to True surrounded by nodes that are set to False is difficult to distinguish.

Our hierarchical graph navigation approach impedes the user in gaining an overview of the entire TNM classification model. The graph layouting technique does not aim at continuity of individual node positions which hampers the recovery of a node of interest after layout changes.

**Functionality** During the TNM verification process, clinicians protocoled their insights by taking notes on the tumor board sheet. Protocoding should be improved by adding functions to the visual analysis component for marking and annotating relevant nodes as...
well as for storing intermediate views of the graph. The clinicians frequently used the table of the TNM staging system during verification. In clinical practice, such tables are attached to walls of the doctor’s room or stored on the clinical workstation. Our prototype should be extended by a table view that provides nodespecific information on request. Finally, the study indicated potential to automate uniform, frequently performed user actions, for instance, highlighting the observed values when reaching examination results during network exploration (cf. step 6 in Section 5.3).

Applicability The TNM staging system for head and neck cancer is annually updated to reflect new evidence. The most recent update comprises some major changes of the lymph node staging including new states as well as additional influences [AGE17]. Still, such an update can easily be implemented. It requires only small adaptations of the BN TNM model, e.g., adding, modifying and removing variables or influences. On the other hand, TNM staging systems are organ-specific, including different states and influences. Therefore, applying our approach to other clinical domains requires a new TNM model. However, the visual analysis component is independent of the TNM model and can readily be transferred.

9. Conclusion
This work presents a visual component for probabilistic graphical models for tumor therapy decision support. The approach is focused on this application but may serve as a role model for a wide variety of complex treatment decisions, e.g. related to severe vascular diseases, where precise diagnosis is essential, as well as to select the right combination of treatment options. Usability issues could be recognized and partially solved by a clear introduction of the design and available tools. Misleading design issues will be addressed in further development. Finally, our component meets the clinical requirements, is simple to understand with a clear introduction, and after a few patient cases also autonomous to use.

In future work, we will add a tutorial to our component presenting the initial introduction as well as tooltips at nodes, states and tools to provide additional assistance.

Acknowledgements
ICCAS is funded by the German Federal Ministry of Education and Research (BMBF). Grant number: 03Z1LN11. The statements made herein are solely the responsibility of the authors. The authors would like to thank the clinicians of the ENT department at the University of Leipzig who evaluated our system as well as the anonymous reviewers and Marek J. Druzdzel and Agnieszka Onisko for or their valuable comments and suggestions.

References


[Goo16] GOOGLE: AngularJS by google, 2016. 9


