Computers & Graphics 91 (2020) 1–11

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on VCBM 2019

A visual approach to explainable computerized clinical decision support

Juliane Müller[a,b,*], Matthaeus Stoehr[c], Alexander Oeser[a], Jan Gaebel[a], Marc Streit[d], Andreas Dietz[c], Steffen Oeltze-Jafra[a,b,e]

[a] ICCAS, Medical Faculty, University of Leipzig, Germany
[b] Department of Neurology, Otto-von-Guericke-University Magdeburg, Germany
[c] Head and Neck Department, University Hospital of Leipzig, Germany
[d] Johannes Kepler University Linz, Austria
[e] Center for Behavioral Brain Sciences, Magdeburg, Germany

ARTICLE INFO

Article history:
Received 17 January 2020
Revised 11 June 2020
Accepted 12 June 2020
Available online 26 June 2020

Keywords:
Information systems (hypertext navigation interfaces
decision support
etc.)
Applications to biology and medical sciences
Medical applications (general),

ABSTRACT

Clinical Decision Support Systems (CDSS) provide assistance to physicians in clinical decision-making. Based on patient-specific evidence items triggering the inferencing process, such as examination findings, and expert-modeled or machine-learned clinical knowledge, these systems provide recommendations in finding the right diagnosis or the optimal therapy. The acceptance of, and the trust in, a CDSS are highly dependent on the transparency of the recommendation's generation. Physicians must know both the key influences leading to a specific recommendation and the contradictory facts. They must also be aware of the certainty of a recommendation and its potential alternatives.

We present a glyph-based, interactive multiple views approach to explainable computerized clinical decision support. Four linked views (1) provide a visual summary of all evidence items and their relevance for the computation result, (2) present linked textual information, such as clinical guidelines or therapy details, (3) show the certainty of the computation result, which includes the recommendation and a set of clinical scores, stagings etc., and (4) facilitate a guided investigation of the reasoning behind the recommendation generation as well as convey the effect of updated evidence items. We demonstrate our approach for a CDSS based on a causal Bayesian network representing the therapy of laryngeal cancer. The approach has been developed in close collaboration with physicians, and was assessed by six expert otolaryngologists as being tailored to physicians' needs in understanding a CDSS.

1. Introduction

Clinical Decision Support Systems (CDSS) provide assistance to physicians in clinical decision-making about a specific patient at the time of care. Applications range from more focused decisions in response to a concrete clinical question (such as medication recommendation) over the selection of the optimal therapy given a set of evidence items triggering the inferencing process (available patient-specific information and findings), up to differential diagnosis based on fuzzy complaints and undifferentiated clinical features. This work focuses on therapy decision support within clinical routine.

A crucial prerequisite for the acceptance and the adoption of a CDSS in clinical routine is establishment of physicians' trust in the computed recommendation [1]. The recommendation needs to be explained and justified by providing evidence items impacting the computed recommendation, and research evidence [2]. Physicians particularly benefit from such explanation facilities when they disagree with the system's recommendation, or when they are at an early level of their professional training and have a limited understanding of the rationale for the line of reasoning for a computed recommendation [3]. In this context, they need to know the relevance of evidence items for the computed recommendation (global relevance) and must know if an item is supporting or contradicting the recommendation (local relevance).

The demand for *explainable* computerized clinical decision support has recently been reinforced with the increasing application of artificial neural networks (ANNs) in non-knowledge-based CDSSs [4], which are learned from data without including previous knowledge. However, *explainability* has already been researched before in the context of knowledge-based CDSSs frequently applying *Bayesian networks* (BNs), which are in focus of this work [5].

* Corresponding author at: Department of Neurology, Otto-von-Guericke-University Magdeburg, German.
E-mail address: juliane.mueller@med.ovgu.de (J. Müller).

https://doi.org/10.1016/j.cag.2020.06.004
0097-8493/

BNs model probabilistic relationships between components of a complex system, and are particularly suited for clinical decision-making under uncertainty and missing information [6]. They have been applied, e.g., in diagnostic reasoning, classification of findings, identification of risk factors, and therapy management [7,8]. The reasoning in BNs is difficult to explain since it does not imitate human reasoning but follows a normative approach, i.e., it assumes an ideal and informed decision-maker performing computations with perfect accuracy [9]. Moreover, the reasoning is based on conditional probabilities, which are hard to assess by humans [10].

Visualization can assist in elucidating the result of, and the reasoning behind, a decision-making process and hence support cognition in CDSSs [11]. However, most existing works in the context of BN-based decision support focus on a visualization of the underlying directed acyclic graphical model (see the work of Cypko et al. [12] for a brief overview). They neglect the application-specific presentation and explanation of the model's computation result and its reasoning via a sophisticated human–computer interface. For instance, Cypko et al. [12] focus on the presentation of the BN network structure for exploratory analysis with less emphasis on a concise interface with explanation facilities supporting the decision-making process in clinical routine. The need for such interfaces has been identified as the grand challenge in clinical decision support [13,14]. We address this challenge with a visual approach to explainable computerized clinical decision support and contribute:

- a scoring function for computing the relevance of an evidence item for the recommendation and for the classification into supportive, contradictory, and recommendation-changer evidence items,
- an interactive multiple views system for explainable clinical decision support inspired by decision-making within clinical routine, and
- capabilities to modify evidence items with the advent of new clinical information or in hypothetical reasoning and a comparative glyph-based visualization of anterior and posterior results.

Parts of our approach can be transferred to any kind of CDSS that gives weighted recommendations based on weighted evidence items for which an influence on the recommendation can be computed. We demonstrate our approach for a BN-based CDSS in endometrial carcinoma management [15]. Our visual approach has been developed in close collaboration with physicians and was evaluated by six expert otolaryngologists. The source code of our visual approach is publicly available online (https://github.com/JulianeMu/explainable_cdss_for_therapy_planning).

The remainder of this work starts with a brief introduction of BNs (Section 2), followed by an overview of related work (Section 3). Then, we present an analysis of requirements on explainable computerized clinical decision support (Section 4), introduce level of relevance computations (Section 5), and detail our visual approach in compliance with the requirements (Section 6). Finally, we describe our evaluation study (Section 7), discuss limitations (Section 8), and draw conclusions (Section 9).

2. Background on Bayesian networks

Bayesian networks are *directed acyclic graphical* (DAG) models that represent conditional probabilistic relations between random variables [16]. Nodes represent variables and edges represent relations. Depending on the variable type, nodes can be discrete or continuous. Discrete nodes have a certain number of *states*, e.g., a node representing the boolean variable "Tobacco" can have the states "true" and "false". Each node is assigned a *probability function* that takes the variable values of its parent nodes as an input and outputs a *probability distribution* over the states of the random variable. The distribution is stored in the form of a *conditional probability table* (CPT). For instance, the probability distribution of a Boolean node with m Boolean parent nodes is stored in a CPT with $2^m \times 2$ entries. Fig. 1 shows a simple BN illustrating these concepts.

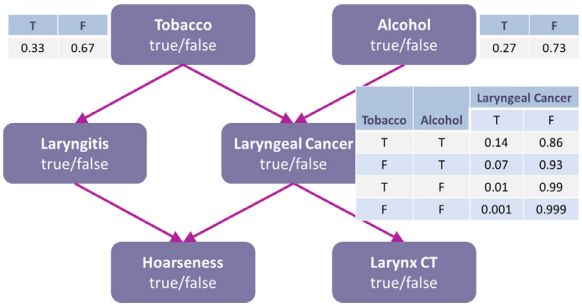

Fig. 1. Oversimplified Bayesian network for differential diagnosis of laryngitis and laryngeal cancer. Smoking and alcohol influence the development of laryngeal cancer. This influence is characterized by the node's conditional probability table (CPT).

BNs are used to answer probabilistic queries about the represented domain. Given the values of variables that can be observed, so-called *evidence items*, the *posterior marginal probability distributions* of *unobserved variables* can be computed using probabilistic *inference methods* [17]. In a medical context, BNs are often constructed based on clinical guidelines and results of clinical and epidemiological studies. BN variables are clinical concepts and relations are *causal dependencies*, i.e., an effect ("Laryngeal Cancer") is dependent on the causes ("Tobacco" and "Alcohol"). Evidence items are typically those representing patient attributes, e.g., sex, age, and the outcomes of medical examinations, e.g, tumor size and blood levels. Unobserved variables comprise unknown patient attributes, outcomes of medical examinations that have not been conducted, and *target variables*, such as the therapy recommendation and a clinical score or staging. The corresponding node of a target variable has as many states as therapy options available and the inferred probability distribution indicates the optimal option as well as the uncertainty about it.

3. Related work

This section describes the related work on explanation methods for BNs and reviews visualization in CDSSs, with a focus on the visualization of recommendation and reasoning.

3.1. Explanation methods for Bayesian networks

Work on explanation of BNs is scarce and mostly in the narrow scope of a particular system with a few exceptions [18]. For instance, a classification of existing research in terms of explanation properties, such as "focus", "purpose", "level", and "causality", has been proposed by Lacave and Diez [5]. The "focus" property is particularly relevant for our work since the focus of explanation can be on evidence, model or reasoning. The explanation of evidence is tightly related to inference, and refers to finding the configuration of unobserved variables that best explains the available evidence. The explanation of the model consists of displaying the content of the underlying knowledge base. The explanation of the reasoning refers to explaining the reasoning process and its results. This work is focused on the latter and links to an explanation of the model.

Linguistic expressions of probability are easier to comprehend than numbers. Hence, the explanation of reasoning is often expressed linguistically, e.g., "Increased probability of B makes A more likely" [19], and presented together with a graphical representation of the BN [20–22]. Especially in case of large BNs, how-

ever, too much text strains the user, the space consumption is too high, and relevant explanations could be overlooked in the process of reading all explanations. Therefore, we favor a graphical representation.

The relevance of individual evidence items for the computation result of a target variable can be determined by means of a sensitivity analysis computing the cost of omitting each evidence [23]. Based on cross-entropy as cost function, the level of relevance and the strength of so-called *chains of reasoning*, i.e., network paths from an evidence to a target variable, can be computed. A graphical approach showing how evidence is propagated through a causal BN has been proposed by Madigan et al. [24], where for each evidence, the level of relevance is computed and displayed on the BN graph representation by modifying the visual properties of nodes and edges. A *graphical evidence balance sheet* shows all evidence items together with their level of relevance. Chains of reasoning are also computed and displayed similar to the work of Suermondt [23]. A visual encoding of evidence relevance and a chain of reasoning visualization have been integrated in the ELVIRA tool for building graphical probabilistic models [25]. The relevance of evidence items within BNs has also been visually addressed by Champion and Elkan [26] using force-directed node-link diagrams and pie chart-inspired glyphs for probability distribution presentation. However, this approach lacks in providing a sufficient overview of given evidence items and is not tailored to applicability within healthcare. Especially when exploring large BNs of more than 1000 nodes, evidence items can be missed despite the emphasis of nodes of interest due to visual clutter. Cypko et al. [12] present a comparative, graph-based exploration of the computed probability distributions obtained from different evidence configurations for supporting the preparation of a head and neck tumor board – a multidisciplinary expert meeting discussing and giving a recommendation about the best therapy for a patient diseased with cancer. Neither of these approaches provides sufficient information about the relevance of an evidence item for the computed recommendation (global relevance) nor whether the item is supporting or contradicting the recommendation (local relevance). However, both are crucial to understand the recommendation generation as well as in assessing the recommendation's robustness or certainty and hence, in gaining trust into the CDSS. Furthermore, the approaches do not provide any information regarding the clinical guidelines and the possible treatments for verification of the computed recommendation. Finally, they do not support exploring the underlying BN in a more structured way along the causal flow.

The evidence view, which shows the available evidence items in a sorted manner regarding their global relevance for the recommendation (see Section 6.1), is inspired by the graphical evidence balance sheet proposed by Madigan et al. [24]. It is, however, tailored to visualizing the relevance of evidence items in deciding between multiple therapy options. The level of relevance is computed by means of a sensitivity analysis but in contrast to the approach proposed by Suermondt [23], a bounded metric is used as cost function (Section 5). Linguistic expressions and chains of reasoning could be useful extensions, but so far have not been considered.

3.2. Visualization and human–computer interfaces in Clinical Decision Support Systems

Visualization and human–computer interfaces have been acknowledged as key components of CDSSs, but are insufficiently studied [2,11,13]. Recently proposed design recommendations stress the importance of a simple interface, a clean and concise presentation, explanation facilities, and easy interaction [2]. Existing application-independent, BN-based decision support systems, such as HUGIN [27], Netica [28], Ergo [29], and GeNIe [30], support the entire pipeline from model building over verification to inference. Hence, their interfaces are complex and not suited for the tight schedule of clinical routine. Further, the incorporated explanation facilities are often limited and/or not tailored to the clinical domain. The visualization of BNs in these systems focuses on drawing the underlying graphical model including the adequate representation of causality and its properties as well as the display of large CPTs (see the work of Cypko et al. [12] for a brief overview).

A sunburst, parallel set, and tree visualization, for example, are employed in a CDSS for patient-specific antibiotic stewardship and compared to traditional *antibiograms* concluding in tree visualizations to be favored by users due to their familiarity [31]. Donut charts are used for monitoring antibiotic resistance in an intensive care unit [32]. The equally-sized donut chart segments represent possible healthcare-associated infections, whereas the color of segments encodes the cases of resistance. Within this CDSS, physicians are only provided with the computed recommendation while information about relevant reasons and underlying uncertainty are lacking. The BN-based visual verification of a tumor's classification has been proposed by Cypko et al. [12] in the context of tumor board preparation. To avoid the cumbersome inspection of the entire large BN in an overview visualization, an adjustable sub-network can be defined. In a drill-down of the BN, nodes can be dragged to and removed from a focus region, causing a restriction of the visualization to this focus and its causally related child nodes. The latter are arranged in the periphery by means of a force-directed layout algorithm, which yields an uncluttered visualization but at the cost of an inconsistent and less predictable node positioning. This hampers orientation in the large network. The probabilities of the node states are encoded by circular glyphs. A dedicated encoding allows for the comparison of a given tumor classification from the patient's paper record to the classification computed by the system. User interfaces in the form of coordinated multiple views have been implemented in CDSSs for diagnostics and treatment planning of traumatic brain injury and for prediction of highly prevalent heart and cardiovascular diseases [33,34]. Both CDSSs include a computation of the relevance of available evidence items for a recommendation, but neglect whether an evidence item is supportive or contradictory for specific states of the recommendation. Lastly, the web tool *NeuroSuites* (https://neurosuites.com/) presents BNs using force-directed or circular node-link diagrams. Additional information, such as the variable's name or most probable state, is only provided on-demand through interaction methods.

Our work proposes an interactive multiple views approach presenting the different aspects of the underlying data, e.g., the therapy guidelines and the system's recommendation, and bringing the relationships between these aspects to attention by means of interaction. Similar to the developed method of Cypko et al. [12], we display only a subset of the BN. Inspired by Wang and Mueller [35], however, we trade the less-predictable, force-directed graph layout and the free exploration for a layout and exploration based on the BN's *causal flow*. We use donut charts for encoding the probability distribution of node states and their design is refined for comparing two distributions resulting from differing sets of evidence items.

4. Requirements and design considerations

To develop an effective visualization for explainable CDSSs, we first must understand the common decision-making process within clinical routine. We need to know which types of information are available about the decision problem, and how humans process these information entities [36,37].

In clinical routine, decision-making is based on clinical guidelines and the physicians' knowledge and experience. The guidelines

take patient-specific data into account, such as examination findings, gender, and age. These data are available from various sources in an often unstructured form and unsorted regarding their relevance for the decision. Physicians integrate the available information mentally and generate their own mental decision model based on knowledge and experience. In this process, they automatically filter all relevant information and weight information entities by their expected impact on the outcome. We have analyzed this process in numerous regularly-conducted work shadowings over four years [12,38,39] and by means of discussions with two collaborating physicians from different disciplines and with mixed levels of experience in BNs. One physician was reassigned from clinical duties for one year to help us implementing a CDSS in laryngeal cancer management (Section 7) [39]. Based on our experiences and taking into account recently proposed design and functional criteria of CDSSs [2], our approach must include representations of:

- **R1** *Patient-specific data*. Evidence items representing the key influences need to be emphasized and sorted regarding their relevance for the recommendation. In that regard, physicians must know supporting as well as contradictory facts for the recommendation and its potential alternatives.
- **R2** *Clinical Guidelines*. Physicians must know the body of medical evidence used for recommendation computation.
- **R3** *Reasoning model*. To improve the acceptance and trustworthiness of a recommendation, physicians have to understand the underlying reasoning process. This can be supported in a BN-based CDSS by facilitating a guided drill-down of the network along the causal flow. Thereby, physicians can relate the BN's reasoning to their mental decision-making model and reasoning.
- **R4** *Decision*. The certainty of the computed recommendation needs to be shown. The state with the highest probability (e.g., optimal therapy option) must be emphasized but at the same time, its uncertainty, i.e., probabilities of other states (e.g., alternative therapy options), must be conveyed.

Furthermore, the decision-making within clinical routine can be classified as *decision-making under uncertainty and risk* [40]. Physicians are trained to decide under a certain level of uncertainty but they must at least know its level, i.e., by which magnitude does, e.g., the computed optimal therapy option outperform its alternatives? Moreover they must be able to evaluate the robustness of a recommendation in "what-if" scenarios investigating what cause provokes what effect by what means at what rate if a particular information is added or updated [41]. Hence, we propose the following additional requirements:

- **R5** *Conveyance of uncertainty*. Most recommendations of a CDSS are associated with uncertainty due to the uncertainty in the underlying knowledge base. The surety of the optimal option given all available options must be conveyed to the physicians.
- **R6** *"What-if" scenarios*. During the decision-making process, new patient information may become available, e.g., a very recent pathological report, or existing information is temporarily adjusted in hypothetical reasoning, e.g., in understanding how further reducing tumor size would affect the recommendation. Thus, our visual approach for explainable CDSS must allow for a fast recommendation update and a pre-post comparison.

5. Level of relevance computation

In decision-making, physicians mentally sort all available evidence items according to their relevance for the addressed question and incorporate the order in developing a recommendation. In the following, we distinguish between *global* and *local* relevance and contribute a simple measure of the latter. We determine both types of relevance in a sensitivity analysis computing the cost of omitting each evidence, similar to Suermondt et al. [23]. Instead of cross-entropy, the *Jensen–Shannon Distance* (JSD) is employed as cost function measuring the dissimilarity between the probability distributions of the target variable before (P) and after (Q) omission [42]. A small dissimilarity corresponds to a low level of relevance. In contrast to cross-entropy, the JSD is a metric, it is zero for equal distributions, and has an upper bound of one ($0 \leq JSD \leq 1$):

$$JSD(P||Q) = \sqrt{\frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)} \tag{1}$$

$$with\ D\ (X||Y) = \sum_i x(i) \log_2 \left(\frac{x(i)}{y(i)}\right)$$

$$and\ M = \frac{1}{2}(P+Q) \tag{2}$$

The JSD provides a *global level of relevance* but does not convey the *local relevance* of an evidence for each state of the recommendation, e.g., each therapy option, and whether the most probable state has shifted after omitting the evidence. Hence, the signed pre-post probability differences are computed and shifts are detected during the sensitivity analysis (*local relevance*). Evidence items yielding a positive difference for a specific state are *supportive*, those yielding a negative difference are *contradictory*, and those causing a state shift are *recommendation-changers*.

6. Visual approach to explainable computerized clinical decision support

Our approach is implemented as an interactive multiple views system (Fig. 2), with one view displayed on-demand only (Fig. 2 (4)) as well as several on-hover views, e.g., Fig. 2 (5). For simplification, we introduce our approach using a small BN for endometrial carcinoma management (http://www.cs.ru.nl/~peterl/endomcancer.html). It consists of 18 nodes, e.g., the patient-specific "pre-operative grade", "recurrence of the tumor", and "lymph node metastasis", as well as 36 relations, and computes a recommendation for the most suitable patient-specific "adjuvant therapy" [15].

The *evidence view* provides an overview of all evidence items sorted by their level of relevance for the computed outcome, e.g., an "Adjuvant therapy" recommendation (Fig. 2 (1)). The *document view* introduces the patient through doctor's letters and provides therapy-relevant information such as the clinical guidelines (Fig. 2 (2)). The *outcome view* overviews the computed probability distributions of the recommendation and an adjustable set of clinical scores, stagings, etc. (Fig. 2 (3)). By scrolling down or selecting an evidence in the evidence view, the *network view* is revealed facilitating a structured exploration of the BN starting at the selected evidence (Fig. 2 (4)).

6.1. Evidence view

Inspired by the regular decision-making within clinical routine, we provide medical decision-makers with all available patient-specific information. In a tabular *evidence view*, we list all evidence items as rows sorted from top to bottom in descending order of global relevance (Eq. (2)) for the computation result of a chosen target variable (**R1**). This assists physicians in identifying the key information for a specific recommendation.

We focus on two levels of impacts instead of just one: the global relevance for the computed recommendation and the local relevance for each target node's states, which are important to represent the underlying uncertainty within the model (**R5**). This assists users in understanding the reasoning process, and helps to

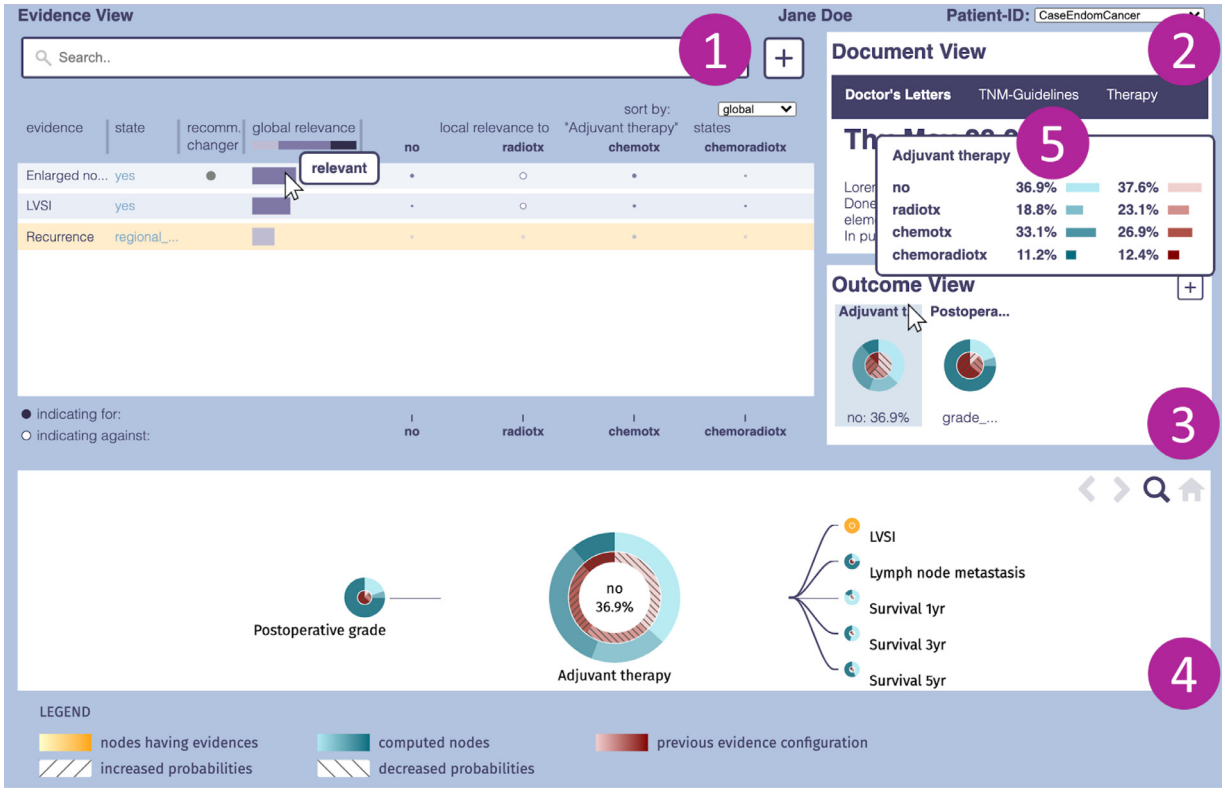

Fig. 2. Interactive multiple views approach to explainable computerized clinical decision support demonstrated on the ENDORISK model [15]. The *evidence view* (1) shows the observed evidence items (Enlarged nodes in CT observation, Lymphovascular Space Invasion (LVSI), and Recurrence) sorted from top to bottom in descending order of their global relevance in computing the target variable (*Adjuvant therapy*). The evidence item's name, state, global level of relevance (colored bars), and local level of relevance for each state of the target variable (circles) are conveyed. The *document view* (2) shows doctor's letters and clinical guideline parts related to the target variable. A placeholder text is displayed here. The *outcome view* (3) overviews the computed probability distributions of the recommendation and an adjustable set of clinical scores, stagings etc. The target variable under investigation is having a light blue background color. Here, two possible target variables (*Adjuvant therapy* and *Postoperative grade*) have been defined. The *network view* (4) allows for a structured exploration of the underlying Bayesian network. The user can drill-down the network along its causal flow. The current node of investigation (*Adjuvant therapy*) is displayed in the center with its causes and effects displayed to the left and right, respectively. Donut and pie chart-based glyphs represent the probability distributions of the nodes' states. For comparing distributions after an evidence update, the outer (blue) and inner (red) donut chart depict the post and prior probability distribution. The sign of difference in probability is encoded by line textures. Hovering facilities provide further insights on the computed probability distributions (5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

build trust in the system. Therefore, and because of the effortless representation of additional information, we decided to use a tabular-inspired view to present evidence items.

In Fig. 2 (1), sorting of evidence items is according to the global relevance for the computed "Adjuvant therapy" recommendation, where the presence of "Enlarged nodes in CT observation" has a little higher relevance level than "lymphovascular space invasion (LVSI)" and "recurrence" (Fig. 2 (1)). The evidence item's name and state are shown in the first two columns. A gray circle in the third column indicates that this evidence is a *recommendation-changer*, e.g., Enlarged nodes in CT observation in Fig. 2 (1). Thus, without this evidence, another recommendation would have been given. Hovering methods allow for deeper investigation of the related recommendations with and without this evidence (Fig. 3). The fourth column shows the evidence item's global relevance for the computed recommendation encoded by colored bars. In agreement with our collaborating physicians, we propose a segmented color scale with values from *less relevant* (0–25%; light purple) over *relevant* (25–75%; purple) to *highly relevant* (75–100%; dark purple). The abstracted categories instead of numerical values with no clinical meaning support interpretability. The remaining columns represent the evidence item's local relevance for all individual states ("no (therapy)", "radiotherapy", "chemotherapy", and "chemoradiotherapy") of the target variable ("Adjuvant therapy") sorted from left to right according to clinical severity. The circles' area encodes the local relevance for the respective target node state whereas its

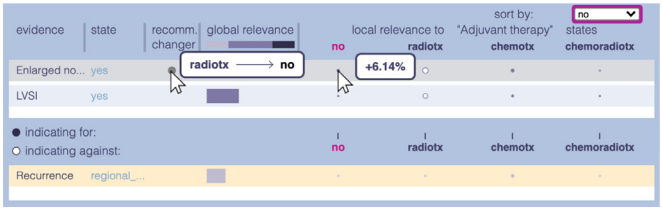

Fig. 3. Evidence view sorted by evidence items' local relevance for a selected outcome state ("Adjuvant therapy – *no (therapy)*", first column in local relevance multicolumn). Supporting and contradictory evidence items are shown in the top and bottom table, respectively. The level of local relevance is encoded by circle area. Hovering allows for detailed effect analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

color is linked to the global relevance of this evidence. In this context, we distinguish between filled and blank circles to represent supporting and contradictory effects on the particular target node's states, respectively (**R1**, **R5**).

Visual perception studies have shown that humans can more accurately read off quantitative values from length than from area [43]. For encoding local relevance, we still favor circles over bars to avoid confusion with the bars encoding global relevance. We employ perceptual scaling to compensate for the known non-linear relationship between an increase in circular area and the perceived increase [44]. This facilitates coarse areas estimates, i.e.,

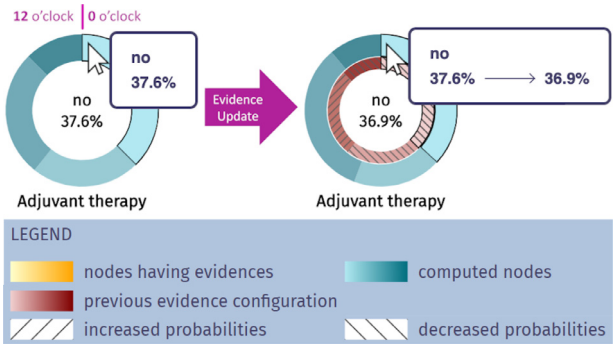

Fig. 4. Donut chart presentation of the probability distribution of a node. The arc length of each segment encodes the probability of the corresponding node state (accessable via hovering). Severe states are displayed in a dark color and positioned close to 12 o'clock while non-severe states are bright and closer to 0'o clock. The blue outer arc depicts the current probability distribution. In case new/more recent evidence is entered or in hypothetical reasoning (Section 6.5), an additional red inner arc is displayed depicting the probability distribution before evidence adjustment. Line textures then, encode the sign of probability differences. Hovering methods emphasize related states and display the probability changes. The clinical severity of a state is mapped to color saturation. Inside the donut chart, the state having the highest probability and the corresponding probability value are shown. The node label is positioned below the glyph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

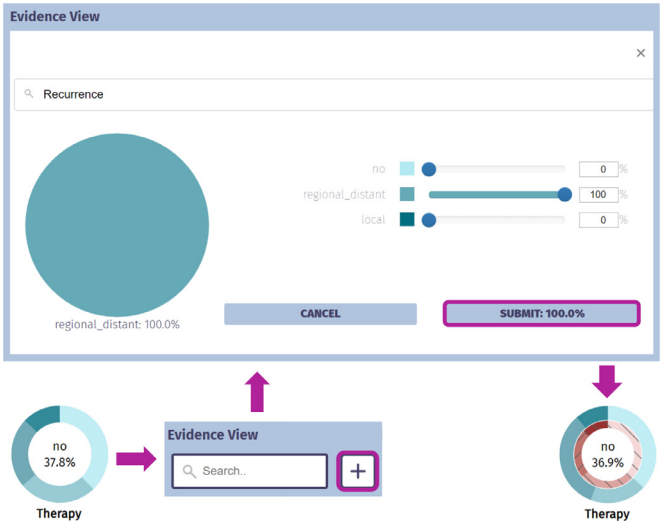

Fig. 5. Evidence updates can be accomplished in a temporary canvas view facilitating the definition of a new probability distribution via sliders. By submitting the new evidence, the Bayesian inference engine is triggered and all views are updated.

small, medium and large, that are sufficient to understand the approximate local relevance. Hovering allows for obtaining the precise numerical value (**R1**, **R5**).

Above the table, a text field with auto-complete functionality allows for the search of a specific evidence whose row is then temporarily highlighted in red. The plus icon allows for inserting new and updating existing evidence items to obtain more information in "what-if" scenarios (**R6**, Section 6.5). The row of the corresponding evidence item is highlighted in yellow for quick identification. A drop-down list allows for switching the sorting criterion of the table from global relevance to local relevance for a specific state, e.g., "no (therapy)" in Fig. 3. This yields a split view of supporting and contradictory evidence items in a sorted top and bottom table, respectively. For example, the "regional distant recurrence of the tumor" finding represents the strongest evidence contradicting the "Adjuvant therapy – *no (therapy)*" outcome, whereas the "Enlarged nodes in CT observation" presence and "LVSI" finding are supporting factors for the application of no therapy.

6.2. Document view

Next to the patient-specific information, physicians are provided with clinical guidelines, doctor's letters, and information about the therapy. We present this information dynamically in the document view (**R2**, Fig. 2 (2)). Text related to the currently-selected outcome variable is displayed and all occurrences of the variable's most probable state are highlighted.

6.3. Outcome view

To communicate the recommendation, we introduce the outcome view. It shows the computed probability distributions of the recommendation and an adjustable set of clinical scores, stagings, etc. (**R4**). In Fig. 2 (3), the recommendation for "Adjuvant therapy" is shown. The most probable state ("no (therapy)") and its probability are displayed as text below the glyph.

A segmented circular glyph shape (Fig. 4) is favored to encode the probability distribution among node states over easier to compare rectangular bars, since it allows for a more dense information representation in case of many variable states, conveys the part-whole relationship, and is visually more pleasing. This is inspired by the proposed approaches of Cypko et al. [12] and Champion et al. [26], but we use color (yellow) to emphasize evidence items to improve their pre-attentive perception [45]. Since these glyphs shall provide a brief presentation of the underlying probability distribution to convey the underlying uncertainty (**R5**) and to avoid overwhelming the user, we decided to not label all possible states, as proposed by Cypko et al. [12]. Hovering facilities, however, provide the actual probabilities on-demand (Fig. 4) or detailed labels and probabilities of all states (Fig. 2 (5)). Additionally, a comparative view can be displayed in an extra, temporary canvas to allow for an in-detail investigation (similar to Fig. 5 with disabled sliders). The size of the glyph segments encodes the values of the computed state probabilities. The color saturation and clockwise position of segments redundantly encode the order of the states in the underlying BN. In causal BNs, the order corresponds to the clinical severity of a state. Thus, severe states, e.g, a high number of lymph nodes affected by cancer, are displayed in a dark color and positioned close to 12 o'clock while non-severe states, e.g., small tumor size, are bright and closer to 0'o clock (Fig. 4). Since no significant difference in proportion assessment accuracy between pie and donut charts could be observed [46], we advocate for donut charts in case of sufficient space.

In case an evidence has been modified by the user, the glyph is split into an inner ring (red) and an outer ring (blue) encoding the pre- and post probability distributions resulting from the original and modified set of evidence items, respectively (**R6**) (Fig. 4). Again, the color saturation and position order correspond to the predefined state order in the underlying BN. Thus, corresponding states in the inner and outer ring can be mentally assigned to each other. Inferring the exact numerical difference from the mutually shifted ring segments is difficult. However, inferring whether probabilities are increasing or decreasing was deemed sufficient by our collaborators. In order to avoid confusion by adding another color scheme, a stripe texture is applied for encoding the sign of the numerical difference. In case the evidence modification did not change the probability distribution, the inner ring is also, drawn in blue. This facilitates a quick identification of nodes affected by the newly entered evidence.

The glyph representing the variable under investigation ("Adjuvant therapy") is drawn with a light blue background

color (Fig. 2 (3)). Clicking on a glyph updates the evidence view accordingly, so that highly relevant evidence items for the selected target node are positioned at the top and the local relevancies are updated regarding the chosen target node's states. The plus icon allows for adding a variable to the outcome view.

6.4. Network view

Physicians use their mental model of the patient and the decision process to reason about the optimal therapy option. To improve acceptance and trustworthiness of a computed recommendation, they must be enabled to relate their mental model to the BN and the inference processes of the CDSS. Thus, we support a structured exploration of the BN along the causal flow of recommendation generation in a dedicated network view (**R3**, Fig. 2 (4)). The view can be particularly useful in case physicians disagree with the computed outcome or relevance of an evidence as well as for investigating the impact of a newly entered or modified evidence. This view is displayed on-demand only by clicking on an evidence in the evidence view (Fig. 2 (1)) or by scrolling down on the interface.

Inspired by a filmstrip metaphor, the user can explore the BN structure forward and backward, starting at the selected evidence [47]. The exploration is guided by the BN's causal flow, which in the representation is from left (cause) to right (effect) as proposed by Wang et al. [35]. The selected evidence is initially displayed in the center and drawn enlarged for better recognition. Its causes and effects are connected by arcs, and sorted alphabetically from top to bottom to ensure a consistent layout and simplify the search for nodes.

If the user clicks on a cause, the filmstrip is animated from left to right. The old node of investigation becomes the effect, the selected cause becomes the centered node of investigation, and its causes smoothly fade in. Navigation facilities are displayed in the top right of the view. Arrow icons allow for moving in either direction along the taken causal path by a sequence of node clicks (**R3**). Moreover, the user can jump to a specific node by a text search for its name (magnifier icon) and return to the home position (home icon).

The donut chart-based glyphs introduced in Section 6.3 and Fig. 4 represent the probability distributions of the node states (**R5**). The glyph encoding is similar to the outcome view with the difference of showing the most probable state and its probability in the center of the glyph instead of positioning it underneath. Additionally, a bar chart-inspired visualization to show detailed information on the probability distribution, similar to Fig. 5 with disabled sliders, can temporarily fade-in.

A yellow background color and ring indicate nodes representing evidence, e.g., the "LVSI" node in Fig. 2 (1) and (4), respectively. Its saturation encodes the clinical severity of the evidence (Section 6.3). We apply a yellowish color instead of blueish to support differentiation between given evidence items and computed variables. In Fig. 2 (4), we simulated that new evidence regarding the "regional distant tumor recurrence" became available during the use of the CDSS. The updated and the previous computed probability distributions of the "Adjuvant therapy" node are then encoded by its outer and inner ring, respectively. In case of multiple evidence updates, the inner circle always represents the initial probability distribution.

6.5. Evidence updates

New evidence can be inserted, and existing evidence can be modified, after clicking on the plus icon in the upper part of the evidence view (**R6**, Fig. 2 (1)). A temporary active canvas view opens and via text search, the desired node is selected (Fig. 5). Sliders and text inputs facilitate the evidence definition while the pie or donut chart-based glyph on the left side is updated automatically to represent the probability distribution. In case of a hundred percent certain evidence, the respective state is simply set to 100%. If the clinical information is affected by uncertainty, e.g., resulting from a measurement device, manual assessments, or the age of information [48], a probability distribution is defined. A sanity check verifies that the overall probability sums up to 100%. By submitting the evidence update to the system, the BN inference engine is triggered. The resulting changes are reflected in the evidence view (Section 6.1) and can be investigated in the outcome (Section 6.3) and the network view (Section 6.4).

6.6. Implementation

Our approach is implemented as a web-based solution using HTML5, and D3 JavaScript [49]. Since the installation of additional software in clinical environments is hard to accomplish due to security issues, we chose a web-based solution. Furthermore, it allows for a quick exchange with users and is available from all kinds of modern web browsers without installing additional software. The internal BN representation and the inference engine are implemented as a FLASK REST web service using the SMILE engine for python (pySMILE) [50].

7. Evaluation

The therapy of a laryngeal cancer patient must be discussed by multiple experts from different disciplines in a tumor board. In preparation of the tumor board, one physician prepares the case including an initial suggestion of how to treat the patient, which will then be discussed by the board. In a perfect evaluation scenario, we would compare the preparation and discussion with and without using the CDSS. However, we focus the evaluation on tumor board preparation and leave the tumor board discussion for future work. It would require a separate complex study with a large organizational effort to imitate a realistic tumor board scenario with many clinical experts.

For visual evaluation purposes, we use a BN representing the treatment of laryngeal cancer [51], which is still under development by medical experts and has not yet been fully validated. Hence, we decided to use a validated sub-network [38]. This subnetwork consists of 303 nodes and 334 causal relations, and represents the *TNM classification* (*T*umor size and type, lymph *N*ode in-filtration, and *M*etastasis spreading) in laryngeal cancer management [52]. It automatically computes the classification based on the available evidence items. The TNM classification or staging has a high influence on the therapy decision-making and is determined manually prior to the tumor board by adhering to the guidelines and based on the patient paper records. The TNM staging is announced in the tumor board meeting by the presenting physician and may be questioned and has to be reworked, e.g., due to very recent updated evidence from imaging or pathology. Thus, in our evaluation, we can see the determination of the TNM classification and its discussion in the tumor board meeting as a reasonable substitute for preparing and discussing the therapy recommendation.

In the evaluation, we investigate three complementing aspects that match the following evaluation codes proposed by Isenberg et al. [53]: *Visual Data Analysis and Reasoning*, *User Performance*, and *User Experience*. Furthermore, according to Gunning [54], the "measures of explanation effectiveness" of artificial intelligence approaches include *user satisfaction (US)*, *mental model (MM)*, *task performance (TP)*, *trust assessment (TA)*, and *correctability (Co)*. Before we elaborate on these aspects, we briefly describe our evaluation procedure.

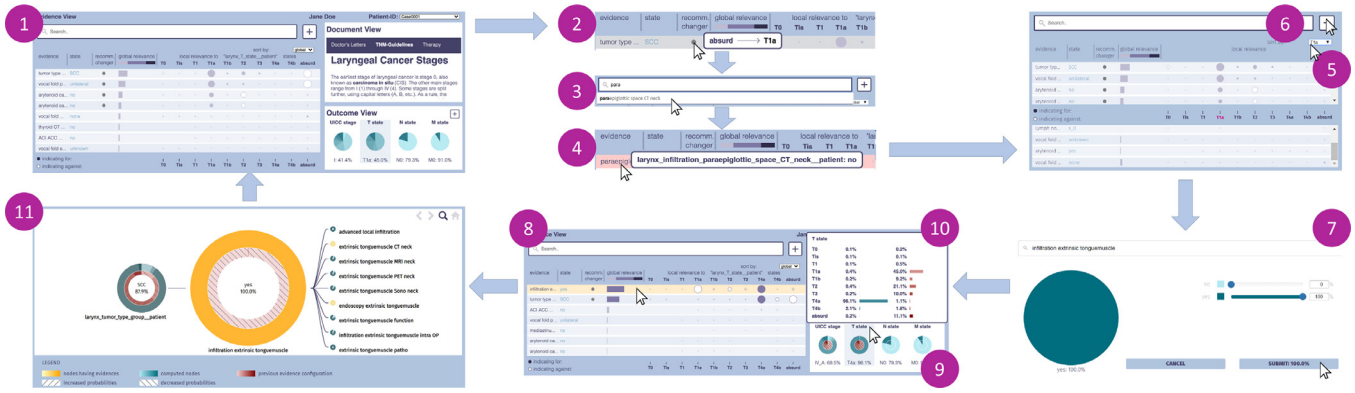

Fig. 6. Conceptual workflow. The participating physicians were asked to comprehend the computed TNM classification, determine and classify the key influencing evidence items, and explain the differences caused by an evidence update. The physicians started with gathering an overview of a patient's condition by collecting all patient-specific information and investigating the predefined outcome variables in the outcome view as well as their key influences in the evidence view (1). Highly important evidence items resulting in a recommendation change are highlighted through a gray circle (2). Since numerous evidence items can be present, a visual auto-complete functionality allows for the search of a specific evidence whose row is then highlighted (3, 4). As physicians must know both, the supporting as well as the contradictory facts for a recommendation, a visual separation of these evidence items is provided (5). During the evaluation process, the physicians added a new evidence item by clicking on the plus icon (6, 7). The physicians then, investigated the resulting changes in the evidence view and the outcome view (8, 9). To fully understand the reasoning behind the TNM classification, they read the guidelines in the document view (10) and explored the BN structure in the network view (11).

7.1. Evaluation procedure

We conducted the evaluation of our visual approach with six experienced otolaryngologists – two male and four female – with one, two, five, six, eight, and ten years of professional experience, respectively. One physician is a co-author of this paper and has expertise in BNs, whereas the other participants are unfamiliar with BNs.

After giving a short introduction to the topic and the tasks that need to be fulfilled, we split the evaluation study into two parts, starting alternatingly to avoid learning effects. In the first part, we asked the participants to determine the TNM-staging for two individual anonymized real patient cases by investigating an unsorted case sheet of patient-specific examination results, which resembles the TNM-staging process within clinical routine. In doing so, we asked them to name the key influences supporting and contradicting their recommendation. These information entities are important for justification when presenting the TNM-classification in the tumor board. We then extended the list of evidence items by one additional examination result. In clinical routine, a newly-aggregated evidence can appear anytime, e.g., during the tumor board from another physician, and needs to be taken into account. Given this extended set of evidence items, we asked the physicians whether they would see a need to change their recommendation and especially, whether the new piece of evidence had an impact on the classification of tumor type. The tumor type, e.g., a squamous cell carcinoma, is highly important for the determination of the T stage, and tumors of different type demand unique treatment.

The second part comprises the same tasks using our new visual approach with the difference that this time they were asked to comprehend the TNM classification of the system instead of determining it themselves. They also used two comparable patient cases recorded within clinical routine. They followed the workflow shown in Fig. 6, beginning with an overview of the patient's condition and the computed recommendation, followed by an exploration of the underlying BN, and an investigation of the consequences caused by the evidence update.

7.1.1. User performance

We recorded the time needed to fulfill all tasks. On average, the participating physicians spent 10.1 min using the unsorted case-sheet of patient-specific examination results. Utilizing our visual approach, however, the average amount of time spent to accomplish the same tasks was 6.6 min; 3.5 min faster. Although five of the physicians had not seen the proposed visual approach before, we observed that after one patient case of learning, all physicians were able to independently operate the prototype. One could argue that the time savings compared to the case-sheet consultation are related to the sorted presentation of evidence items (*TP*). However, the participating physicians described the TNM-classification process using paper records and the hospital information system as comparable to the case-sheet presentation. Only their experience with the hospital information system allows them for a quick search for examination results on a regular basis. Therefore, our comparison is valid.

7.1.2. Visual data analysis and reasoning

Although our evaluation study was not performed to validate the correctness of the utilized BN, enhancements were suggested by the physicians. For example, the physicians noted that the tumor type pathohistological finding is overweighted in the TNM-staging process (*Co*). In this context, we emphasize that the physician who was not familiar with BNs was also able to identify this overweighting. Thus, our visual approach can assist in the validation of BNs.

7.1.3. User experience

During the justification process of the computed recommendation, the physicians especially valued the well-structured and clear presentation of evidence items sorted by their relevance for the recommendation in the *evidence view* (Section 6.1, Fig. 2 (1)). It allows getting an overview of the patient's condition, indicating the important patient-specific information leading to the recommendation, and justifying the computed outcome (*US*, *MM*). Furthermore, the usage of familiar visualizations, such as pie and bar chart-inspired presentations, to represent evidence items and the associated uncertainty within the system, were stated to be very helpful during the comprehension process. Additionally, the selection of outcomes of interest in the *outcome view* (Section 6.3, Fig. 2 (2)) and the resulting updated sorting of evidence items by their global relevance in an animated fashion in the *evidence view* (Section 6.1, Fig. 2 (1)) were described as improving the usability (*TA*).

By providing additional information in the form of doctor's letters and clinical guidelines in the *document view* (Section 6.2,

Fig. 2 (3)), the justification process is simplified since physicians can, e.g., relate the relevant evidence items to their description in the corresponding clinical guidelines. For example, one physician compared the evidence items leading to a N-state – *N0* with the regarding TNM-guidelines from the *document view* (Fig. 2 (3)) and confirmed the computed outcome.

We conducted a questionnaire to get insights into the clinical relevance of the proposed approach, the effectiveness of the visualizations, and the convenience of the interaction facilities. It comprises 26 questions, 7 inquiring participant details and 19 employing a five-point Likert scale to rate various aspects of relevance, effectiveness and convenience (−−, - ,o, + ,++). While the six physicians filled in the questionnaire, the spoken comments were recorded. We provide the questionnaire, as well as the detailed answers, in the supplementary material and summarize the most important insights in the following:

All physicians emphasized the importance and clinical relevance of visual explanation and guidance in a CDSS. They confirmed that the proposed visual approach fulfills their demands on comprehension and justification of the generated recommendation (*TA*). They particularly valued the structured, clear, and familiar presentation of all evidence items, which resembles their regular decision-making process. The sorting of all evidence items regarding their local relevance for a specific state of an outcome variable as well as the emphasis of supporting and contradictory facts using filled and unfilled circles, however, were seen as more controversial. One physician claimed the contradictory evidence items are irrelevant whereas the other physicians rated this information to be helpful (see in Fig. 3). The presentation of probability distributions using donut-charts was rated as clear and understandable.

The physicians appreciated the visual emphasis of all changes resulting from an evidence update by double-ring donut charts. In the network view, they were able to investigate the impact of the update on the reasoning of the CDSS. The view linking and interaction facilities were rated as conforming with user expectations, and operating in a concise manner. All interactions provide immediate feedback. Running the BN reasoning engine after an evidence update yields results in around two seconds on a standard notebook, which was considered acceptable by the physicians.

The animation of changes in all views except the document view, e.g., animation of rows shifting in the table of the evidence view after choosing a different outcome variable and animation of graph navigation in the network view (see the supplemental video), were highly appreciated. They assisted the physicians in retracing an evidence's global influence on multiple outcomes and in general, in keeping track of changes in the visualization. Finally, all physicians commented that they would in general argue for using the proposed approach during tumor board preparation and board meeting in order to present their initial treatment recommendation.

8. Discussion

The evaluation indicates that our approach supports the verification and justification of recommendations provided by a CDSS. The combination of simple and familiar visualization techniques, such as pie chart and bar chart-inspired visualizations representing highly complex processes, allows physicians without knowledge in CDSS for the comprehension of generated recommendations. Using our prototype, the physicians were able to verify the computed TNM-staging recommendation by investigating the proposed information (**R1**, **R2**, **R3**, **R4**) while comprehending the associated uncertainty within the generated outcome (**R5**). All physicians emphasized the ability to update evidence items and inspect the related highlighted changes (**R6**) as well as the investigation of causal relations within the underlying network (**R3**). The interaction techniques and their response time were rated as fulfilling the required needs and suitable for the accomplished tasks. Our prototype serves as an introduction tool to the patient's condition, since it summarizes all given patient-specific information sorted by their relevance. With these advantages, it can assist during the tumor board preparation in recommendation generation and during the tumor board in discussing the recommendation.

Limitations. Currently, our approach is tailored to graphical causal models such as BNs. However, with the exception of the network view, it can be transferred to any kind of model incorporating uncertainty and any kind of function for specifying the global relevance of individual evidence items.

Furthermore, patient-specific information are provided in a csv file at present. For application in clinical routine, data needs to be fetched directly from the hospital information system, gathering all available information automatically.

Due to space restrictions, only a limited number of evidence items are visible at a glance in the evidence view. However, because of the sorting of the evidence items by their individual level of relevance for a recommendation, the most important patient-specific information entities are visible while the remainder is retrievable through scrolling methods.

Currently, only the local structure of the network is presented. Thus, users can investigate the direct causal reasons and causal effects of a single node of interest using the network view. One physician requested a global representation of the whole underlying network using a fish-eye-technique [55]. The causal chain of reasoning [23] between two nodes of interest further apart was also requested. However, within BNs there are multiple paths of reasoning, and not every node is of importance within this path. To convey the causal chain of reasoning, a suitable presentation needs to be found.

In hypothetical reasoning using our approach, only two sets of evidence items can be currently compared. A collaborative setup, where several experts discuss various alternatives, requires multiple of these comparisons.

Lessons learned. During the development of our proposed visual approach, we recognized that it was especially important to discuss our concept with many domain experts having different levels of expertise in causal modeling. In the beginning, we shared our ideas with two expert physicians. However, when conducting the evaluation with domain experts who did not contribute to the design phase, it became clear that improvements, such as the integration of global and local relevance, were essential for understanding and justifying the reasoning process.

Furthermore, in our initial development phase, we focused too much on "technical" explainability of the Bayesian inferencing by providing facilities for network exploration. After multiple discussions and an initial evaluation with domain experts, we changed our point of view after learning that "medical" explainability is very much related to convey the relevance of individual evidence items for the computed recommendation.

9. Conclusion and future work

For CDSS, the presentation and explanation of the underlying models computation result and its reasoning via a sophisticated human computer interface have been identified as the grand challenges. We address these challenges by presenting a novel visual approach to explainable computerized clinical decision support. Inspired by the decision-making process in clinical routine, our approach combines uncertainty-aware presentations of the computed recommendation with methods for justification presented through a list of patient-specific evidence items sorted by their relevance for the chosen outcome. Additionally, a structured exploration method for comprehension of the reasoning within the

CDSS depicting the local structure of the model/network is provided. Analysis of the impact of an evidence update allows for the investigation and justification of the generated results and provides methods for hypothetical reasoning regarding a patient's most probable future condition. During an evaluation study, physicians assessed our proposed approach to generate trustworthy and justifiable results, and have recommended its usage within clinical routine.

For future work, physicians proposed a representation of the global structure of the network and the visualization of the causal chain of reasoning within the CDSS. By providing this information, more insights into underlying reasoning can be gathered, which assists especially for inexperienced physicians in learning the decision-making process for a specific clinical question. Furthermore, we want to provide export facilities of evidence item sets to store the state after user modifications. As soon as the treatment decision model for patients suffering from laryngeal cancer is completed and validated, a clinical comparative evaluation of applying the proposed approach in the tumor board can be conducted.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Juliane Müller: Conceptualization, Methodology, Software, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Matthaeus Stoehr:** Formal analysis, Validation, Investigation, Resources, Data curation. **Alexander Oeser:** Conceptualization. **Jan Gaebel:** Conceptualization, Software. **Marc Streit:** Writing - review & editing. **Andreas Dietz:** Supervision. **Steffen Oeltze-Jafra:** Conceptualization, Supervision, Writing - review & editing.

Acknowledgments

The research leading to this work was supported by the German Federal Ministry of Education and Research (BMBF) (03Z1LN11) and the Federal State of Saxony-Anhalt (FKZ: I 88).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cag.2020.06.004

References

[1] Vourgidis I, Mafuma SJ, Wilson P, Carter J, Cosma G. Medical expert systems – a study of trust and acceptance by healthcare stakeholders. In: Lotfi A, Bouchachia H, Gegov A, Langensiepen C, McGinnity M, editors. Advances in computational intelligence systems. Cham: Springer International Publishing; 2019. p. 108–19. doi:10.1007/978-3-319-97982-3_9.

[2] Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inf Assoc 2018;25(5):585–92. doi:10.1093/jamia/ocx118.

[3] Darlington KW. Designing for explanation in health care applications of expert systems. Sage Open 2011;1(1). doi:10.1177/2158244011408618. 215824401140861

[4] Berner ES, La Lande T. Overview of clinical decision support systems. In: Berner ES, editor. Clinical decision support systems: theory and practice. Health informatics. Springer; 2016. p. 3–22. doi:10.1007/978-3-319-31913-1_1.

[5] Lacave C, Diez FJ. A review of explanation methods for Bayesian networks. Knowl Eng Rev 2002;17(02):107–27. doi:10.1017/S026988890200019X.

[6] Spooner S. Mathematical foundations of decision support systems. In: Berner ES, editor. Clinical decision support systems: theory and practice. Health informatics. Springer; 2016. p. 23–43. doi:10.1007/978-3-319-31913-1_1.

[7] Lucas PJ, van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artif Intell Med 2004;30(3):201–14. doi:10.1016/j.artmed.2003.11.001.

[8] Yet B, Bastani K, Raharjo H, Lifvergren S, Marsh W, Bergman B. Decision support system for warfarin therapy management using Bayesian networks. Decis Support Syst 2013;55(2):488–98. doi:10.1016/j.dss.2012.10.007.

[9] Druzdzel M. Explanation in probabilistic systems: is it feasible? will it work?. In: Proceedings of the 5th international workshop on intelligent information systems (WIS-96); 1996. p. 12–24. Deblin, Poland

[10] Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty. Cambridge: Cambridge University Press; 1982. p. 249–67. doi:10.1017/CBO9780511809477.019.

[11] Franklin A, Zhang J. Cognitive considerations for health information technology. In: Greenes RA, editor. Clinical decision support. Oxford: Academic Press; 2014. p. 619–40. doi:10.1016/B978-0-12-398476-0.00022-1.

[12] Cypko MA, Wojdziak J, Stoehr M, Kirchner B, Preim B, Dietz A, et al. Visual verification of cancer staging for therapy decision support. Comput Graph Forum 2017. doi:10.1111/cgf.13172.

[13] Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008;41(2):387–92. doi:10.1016/j.jbi.2007.09.003.

[14] Okumura T, Kajiyama T, Sonehara N. Differential diagnosis listing as relevance feedback: an essential user interface for clinical decision support systems. In: Proceedings of the IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2017. p. 53–4. doi:10.1109/CBMS.2017.92.

[15] Reijnen C, Gogou E, van der Putten L, Visser N, van de Vijver K, Santacana M, et al. Development and validation of an endometrial carcinoma preoperative Bayesian network using molecular and clinical biomarkers (ENDORISK): an ENITEC collaboration study. Int J Gynecol Cancer 2019;29(Suppl 4):A6–7. doi:10.1136/ijgc-2019-ESGO.7.

[16] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1988. ISBN 0-934613-73-7.

[17] Winkler RL. An introduction to Bayesian inference and decision. 2nd. Gainesville, FL: Probabilistic Pub; 2003. ISBN 978-0-9647938-4-2.

[18] Biran O, Cotton C. Explanation and justification in machine learning: a survey. In: Proceedings of the IJCAI-17 Workshop on Explainable AI (XAI); 2017. p. 8–13. Melbourne, Australia

[19] Druzdzel M, Henrion M. Qualitative verbal explanations in Bayesian belief networks. Artif Intell Simul Behav Q 1996;94:43–54.

[20] Elsaesser C, Henrion M. Verbal expressions for probability updates how much more probable is "much more probable"?. In: Proceedings of the 1990 machine intelligence and pattern recognition, 10. Elsevier; 1990. p. 319–28. doi:10.1016/B978-0-444-88738-2.50032-4.

[21] Haddawy P, Jacobson J, Kahn CE. BANTER: a Bayesian network tutoring shell. Artif Intell Med 1997;10(2):177–200. doi:10.1016/S0933-3657(96)00374-0.

[22] Yap G-E, Tan A-H, Pang H-H. Explaining inferences in Bayesian networks. Appl Intell 2008;29(3):263–78. doi:10.1007/s10489-007-0093-8.

[23] Suermondt H. Explanation in Bayesian belief networks. Stanford University; 1992. STAN-CS-92-1417

[24] Madigan D, Mosurski K, Almond RG. Graphical explanation in belief networks. J Comput Graph Stat 1997;6(2):160–81. doi:10.1080/10618600.1997.10474735.

[25] Lacave C, Luque M, Diez F. Explanation of Bayesian networks and influence diagrams in Elvira. IEEE Trans Syst Man Cybern Part B (Cybern) 2007;37(4):952–65. doi:10.1109/TSMCB.2007.896018.

[26] Champion C., Elkan C.. Visualizing the consequences of evidence in Bayesian networks. arXiv preprint arXiv:1707007912017.

[27] HUGIN. HUGIN EXPERT White Paper. 2020. URL http://download.hugin.com/webdocs/whitepaper/huginexpert-whitepaper.pdf.

[28] Woodberry O, Mascaro S. Programming Bayesian network solutions with Netica; 2014. ISBN 978-1-291-81267-1. OCLC: 940599924

[29] Beinlich I, Herskovits EH. Ergo: a graphical environment for constructing Bayesian. In: Proceedings of the 6th conference on uncertainty in artificial intelligence (UAI1990); 1990. p. 114–21. Corvallis, Oregon, USA

[30] Druzdzel M. SMILE: structural modeling, inference, and learning engine and GeNIe: a development environment for graphical decision-theoretic models. In: Proceedings of the 16th national conference on artificial intelligence and 11th conference on innovative applications of artificial intelligence; 1999. p. 902–23. Orlando, Florida, USA

[31] Garcia-Caballero H, Campos M, Juarez J, Palacios F. Visualization in clinical decision support system for antibiotic treatment. In: Proceedings of the XVI conferencia de la asociación española para la inteligencia artificial, 1; 2015. p. 71–80. Albacete, Spain

[32] Gil M, Pinto P, Simões A, Póvoa P, Da Silva M, Lapo L. Co-design of a computer-assisted medical decision support system to manage antibiotic prescription in an ICU ward. Stud Health Technol Inform 2016:499–503. doi:10.3233/978-1-61499-678-1-499.

[33] Umer A, Mattila J, Liedes H, Koikkalainen J, Lotjonen J, Katila A, et al. A decision support system for diagnostics and treatment planning in traumatic brain injury. IEEE J Biomed Health Inform 2018. doi:10.1109/JBHI.2018.2842717. 1–1

[34] Jin Z., Yang J., Cui S., Gotz D., Sun J., Cao N.. Carepre: an intelligent clinical decision assistance system. arXiv:181102218 [cs]2018;ArXiv: 1811.02218.

[35] Wang J, Mueller K. Visual causality analysis made practical. In: Proceedings of the 2017 IEEE conference on visual analytics science and technology (VAST); 2017. p. 151–61.

[36] Kohlhammer J, May T, Hoffmann M. Visual analytics for the strategic decision making process. In: Amicis RD, Stojanovic R, Conti G, editors. GeoSpatial visual analytics. Dordrecht: Springer Netherlands; 2009. p. 299–310. ISBN 978-90-481-2899-0.
[37] Daradkeh M, Abul-Huda B. Incorporating uncertainty into decision-making: An information visualisation approach. In: Linden I, Liu S, Colot C, editors. Decision support systems VII. Data, information and knowledge visualization in decision support systems. Cham: Springer International Publishing; 2017. p. 74–87. ISBN 978-3-319-57487-5.
[38] Cypko MA, Stoehr M, Kozniewski M, Druzdzel MJ, Dietz A, Berliner L, et al. Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment; 2017. p. 1–12. doi:10.1007/s11548-017-1531-7.
[39] Cypko MA. Therapy decision support system using Bayesian networks for multidisciplinary treatment decisions. University of Leipzig; 2017.
[40] Clemen RT, Reilly T. Making hard decisions with decisiontools. Cengage Learning; 2013.
[41] Tufte ER, McKay SR, Christian W, Matey JM. Visual Explanations: Images and Quantities, Evidence and Narrative. Computers in Physics 1998;12(2):146–8. doi:10.1063/1.168637.
[42] Endres D, Schindelin J. A new metric for probability distributions. IEEE Trans Inf Theory 2003;49(7):1858–60. doi:10.1109/TIT.2003.813506.
[43] Mackinlay J. Automating the design of graphical presentations of relational information. ACM Trans Graph (ToG) 1986;5(2):110–41.
[44] Flannery JJ. The relative effectiveness of some common graduated point symbols in the presentation of quantitive data. Cartograph: Int J Geograph Inf Geovisual 1971;8(2):96–109. doi:10.3138/J647-1776-745H-3667.
[45] Healey C, Enns J. Attention and visual memory in visualization and computer graphics. IEEE Trans Vis Comput Graph 2012;18(7):1170–88. doi:10.1109/TVCG.2011.127.
[46] Cai X, Efstathiou K, Xie X, Wu Y, Shi Y, Yu L. A study of the effect of doughnut chart parameters on proportion estimation accuracy. Comput Graph Forum 2018. doi:10.1111/cgf.13325.
[47] Elzen Sv d, Wijk JJv. Small multiples, large singles: a new approach for visual data exploration. Comput Graph Forum 2013. doi:10.1111/cgf.12106.
[48] Gaebel J, Wu H-G, Oeser A, Cypko MA, Stoehr M, Dietz A, et al. Modeling and processing up-to-dateness of patient information in probabilistic therapy decision support. Artif Intell Med 2020;104:101842. doi:10.1016/j.artmed.2020.101842.
[49] Bostock M, Ogievetsky V, Heer J. D3 data-driven documents. IEEE Trans Vis Comput Graph 2011;17(12):2301–9. doi:10.1109/TVCG.2011.185.
[50] BayesFusion. SMILE: structural modeling, inference, and learning engine. 2020. URL https://www.bayesfusion.com/smile/.
[51] Stoehr M, Cypko MA, Denecke K, Lemke HU, Dietz A. A model of the decision–making process: therapy of laryngeal cancer. Int J CARS 2014;9(Suppl 1).
[52] Pfister DG, Spencer S, Brizel DM, Burtness B, Busse PM, Caudell JJ, et al. Head and neck cancers, version 2.2014. J Natl Comprehen Cancer Netw 2014;12(10):1454–87.
[53] Isenberg T, Isenberg P, Chen J, Sedlmair M, Mller T. A systematic review on the practice of evaluating visualization. IEEE Trans Vis Comput Graph 2013;19(12):2818–27. doi:10.1109/TVCG.2013.126.
[54] Gunning D. Explainable artificial intelligence (XAI), 2. Defense Advanced Research Projects Agency (DARPA); 2017.
[55] Sarkar M, Brown MH. Graphical fisheye views of graphs. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '92. New York, NY, USA: ACM; 1992. p. 83–91. doi:10.1145/142750.142763.