Visual Assistance in Development and Validation of Bayesian Networks for Clinical Decision Support

Juliane Müller-Sielaff, Seyed Behnam Beladi, Stephanie W. Vrede, Monique Meuschke, Peter J.F. Lucas, Johanna M.A. Pijnenborg, Steffen Oeltze-Jafra

Abstract—The development and validation of Clinical Decision Support Models (CDSM) based on Bayesian networks (BN) is commonly done in a collaborative work between medical researchers providing the domain expertise and computer scientists developing the decision support model. Although modern tools provide facilities for data-driven model generation, domain experts are required to validate the accuracy of the learned model and to provide expert knowledge for fine-tuning it while computer scientists are needed to integrate this knowledge in the learned model (hybrid modeling approach). This generally time-expensive procedure hampers CDSM generation and updating. To address this problem, we developed a novel interactive visual approach allowing medical researchers with less knowledge in CDSM to develop and validate BNs based on domain specific data mainly independently and thus, diminishing the need for an additional computer scientist. In this context, we abstracted and simplified the common workflow in BN development as well as adjusted the workflow to medical experts' needs. We demonstrate our visual approach with data of endometrial cancer patients and evaluated it with six medical researchers who are domain experts in the gynecological field.

Index Terms—Bayesian networks, Visual Analysis, Clinical Decision Support, Causal Model Development

1 Introduction

Clinical Decision Support Systems (CDSS) provide assistance to the medical doctor in finding the right diagnosis or the optimal patient-specific treatment. Crucial for the acceptance of and the trust in CDSSs are explainability and transparency [1]. Thus, the user needs to be able to understand the reasoning and decision-making process of a CDSS adequately so that the CDSSs support the justification of particular conclusions. Causal probabilistic network models, such as Bayesian networks (BN), have a clear semantics in terms of probability theory, conditional independence and dependence. They also support the explicit representation of causal structure. It is this semantics that allows explaining causal reasoning patterns and exploring "what-if" scenarios, supporting decision-making [2]. As a consequence, we firmly believe that causal probabilistic networks provide an appropriate foundation for developing CDSS.

To allow employing these models as a part of clinical routine assisting in decision-making, BNs have to be developed in medical research first and subsequently validated to obtain insight into their accuracy before their actual use. Development and validation are either done based on expert knowledge and manual modeling [3], or based on a large data pool and model learning using fully automated machine-learning methods [4], or a combination of both in a hybrid approach [5], [6]. Approaches that are fully based on expert knowledge and experience reflect expertise in a particular specialized area, but cannot cope with differences in views, inaccuracy in human judgment, and may include various forms of bias. Purely data-driven approaches, however, are only as accurate as the given dataset. In clinical medicine, datasets often belong to a small number of patients and their level of data quality rapidly decreases by increasing number of patients because of missing data records. Clearly, a hybrid approach where machine learning algorithms trained on data and integrated expert-based domain knowledge offers the best of both worlds. Our visual methodology described in this paper is based on the combination of data-driven development in combination with expert knowledge integration.

Standard tools for BN development and domain knowledge integration having graphical user interfaces, such as GeNIe [7], [8], Hugin [9], Netica [10], NeuroSuites [11], [12], and Bayes Server [13], are very complex and poorly adjusted to the medical researchers' needs [3], [14]. Therefore, such tools usually require collaboration between a computer scientist and a domain expert to incorporate domain knowledge, which is most likely highly time-consuming. For example the BN "TreLynCa" [3] in laryngeal cancer management was developed in close collaboration between

- *Juliane Müller-Sielaff is with Dept. of Neurology, Otto von Guericke University Magdeburg, Germany. E-mail: Juliane.Mueller@med.ovgu.de*
- *Seyed Behnam Beladi is with Department of Simulation and Graphics & Dept. of Neurology, Otto von Guericke University Magdeburg, Germany. E-mail: SeyedBehnam.Beladi@ovgu.de*
- *Monique Meuschke is with Department of Simulation and Graphics, Otto von Guericke University Magdeburg, Germany & Institute for Computer Science, University of Jena, Germany. E-mail: Meuschke@isg.cs.uni-magdeburg.de*
- *Stephanie W. Vrede is with Department of Obstetrics & Gynecology, Radboud university medical center, Nijmegen, the Netherlands. E-mail: Stephanie.Vrede@radboudumc.nl*
- *Peter J.F. Lucas is with the University of Twente, Enschede, & LIACS, Leiden University, the Netherlands. E-mail: Peter.Lucas@utwente.nl*
- *Johanna M.A. Pijnenborg is with Department of Obstetrics & Gynecology, Radboud university medical center, Nijmegen, the Netherlands. E-mail: Hanny.MA.Pijnenborg@radboudumc.nl*
- *Steffen Oeltze-Jafra is with Dept. of Neurology & the Center of Behavioral Brain Sciences, Otto von Guericke University Magdeburg, Germany. E-mail: Steffen.Oeltze-Jafra@med.ovgu.de*

Manuscript received June 04, 2021; revised June 04, 2021.

a computer scientist and a medical researcher over a three year period, whereby only the subnetwork for *TNM classification* (*T*umor size and type, lymph *N*ode in-filtration, and *M*etastasis spreading [15]) has been validated until present [3]. Because the computer scientist might make design decisions on the experts' domain to speed-up the development process, at some stage the expert may be confronted with a lack of understanding of the resulting BN model.

To address these problems and to reduce the time investments needed for BN development, we introduce an interactive visual approach assisting in the development and validation of BNs. Our tool is especially designed for a target audience of **medical researchers**. Thus, it requires medical knowledge but discounts prior knowledge in BNs. Due to the complexity of the topic itself, an application within clinical routine is not intended. The main question driving the development of our approach can be described as "Which visual design and interaction techniques allow medical researchers without experience in clinical decision support models (CDSM) to comprehensively develop and validate BNs mainly independently?". In our approach, we are taking the common process of BN development and validation into account, starting with (*i*) "data review & structural development", followed by (*ii*) "structural validation", (*iii*) "parametrization development", (*iv*) "parametrization validation", and (*v*) "model validation" (Fig. 1) [3]. A detailed description of these process steps is provided in Sec. 2. In our approach, however, we combine the (*iii*) "parametrization development", (*iv*) "parametrization validation", and (*v*) "model validation" steps to the "model parametrization and validation" step to reduce the complexity of BN development and validation. In addition, as our focus and application area is in medicine, we introduce *clinical workflow variable steps/groups* to assist users in comprehending and organizing the causal relations in the BN. For example, in making a diagnosis, the first step is a medical interview, followed by a physical examination, and laboratory tests. Variables can be assigned by the user to these steps in order to impose a meta structure on the BN.

In a perfect BN development and validation process paving the way for clinical use, collaborative modeling of multiple medical experts is required to ensure a broad applicability and reduce biases within the model. As a first crucial step towards this goal, we are focusing here on the visual and information requirements allowing single medical researchers to develop and validate BNs without the need for a BN modeling expert. This forms the basis for future research in collaborative BN creation.

Our main contributions are:

- user-centric interactive visual approach for BN development and validation adjusted to the medical researchers' needs
- a simplified pipeline of BN development and validation starting with (*i*) "data review & structural development", followed by (*ii*) "structural validation" and (*iii*) "model parametrization and validation" in an iterative way
- organization of the model variables into *clinical workflow steps / groups* inspired by decision-making in clinical routine to assist medical researchers in comprehending and justifying the causal relations in BNs.

2 Background on Bayesian Networks

A *Bayesian network* (BN) is a joint (or multivariate) probability distribution. It is defined by a family of conditional probability distributions, often called *Conditional Probability Tables* (CPTs), with one CPT per variable, and a directed acyclic graph (DAG) that reflects the dependencies of the sets of variables in the joint probability distribution (Fig. 2). A DAG consists of *nodes* and directed links (*arcs*) between nodes. By assuming that there exists an one-to-one correspondence between the nodes in the DAG and the random variables of the distribution, the DAG defines constraints on the distribution. The family of conditional probability distributions have the form

$$P(X_n \mid X_{\mathrm{pa}(n)}),$$

where n is a node in the DAG, X_n the corresponding variable in the probability distribution, $\mathrm{pa}(n)$ the set of nodes that link directly to node n, called the *parents*, and $X_{\mathrm{pa}(n)}$ is the family of variables associated with the parents. BNs present themselves as highly intuitive, which is mostly due to the fact that arcs (directed links between two nodes) can often be understood in terms of cause and effect relationships. For example an effect represented by the node "tumor recurrence" in endometrial cancer management is dependent on the presence of "lymph node metastases" and applied "therapy" (Fig. 2). Variables in a BN can be either discrete or continuous. In BN learning algorithms, often all variables are deemed as either discrete or continuous, but in principle these variables can also be mixed. However, dealing with arbitrary mixtures of discrete and continuous variables requires using sophisticated approximate methods, e.g., involving sampling from the distribution of the network. As a consequence, discrete BNs are more widely applied and are also the focus of our

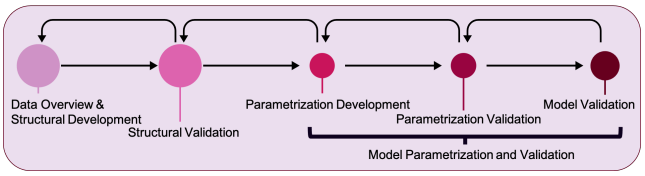


Fig. 1: The common process in BN development and validation taking into account data and user knowledge starts with (*i*) "data review & structural development", followed by (*ii*) "structural validation", (*iii*) "parametrization development", (*iv*) "parametrization validation", and (*v*) "model validation" [3]. At each step, the user is able to redo the previous step. During our iterative design development, we were able to combine the last three steps to the "model parametrization and validation" and thus, simplified the process significantly.

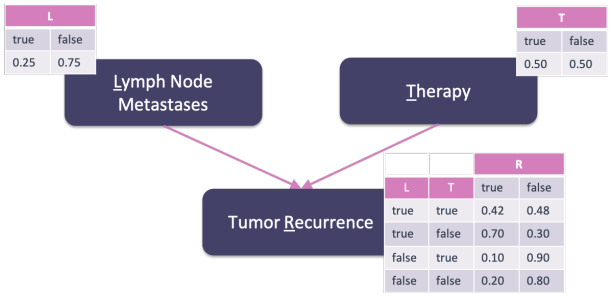


Fig. 2: Oversimplified BN for the differential purpose of the tumor recurrence of endometrial cancer patients. Lymph node metastases and therapy's influence on the tumor recurrence are given by the conditional probability table.

work [16]. BNs are typically used to compute the conditional probability distribution of each of the unobserved variables X_n conditioned on the variables with observed values of a patient, called the *evidence items*:

$$P(X_n \mid \text{evidence items})$$

Such distributions are highly meaningful as they reflect the updated, local effect of information concerning a patient on the other variables [17]. Thus, BNs can be used to compute the best patient-specific therapy or most-likely diagnosis for instance. Due to their structure and parametrization, BNs are especially suited for reasoning under uncertainty and deal with missing values in the evidence by incorporating prior probabilistic information from the network without making assumptions about these missing values [18].

The pipeline in BN development starts with structural development followed by parametrization development represented in the form of the CPTs and model validation (Fig. 1). Different learning techniques have been developed to identify the best fitting causal structure based on the data. Algorithms are roughly classified into "constraint-based" and "score-based" methods. Constraint-based methods, such as *grow-shrink* (gs) and *pc.stable*, build up the BN graph by carrying out conditional independent tests between sets of variables. All algorithms are related to the earlier *inductive causation algorithm* [19]. Score-based algorithms work differently: they apply a heuristic search algorithm, such as *hill climbing* or *tabu search*, to the DAG space, which is superexponentially in size, to find a graph structure that fits the data best expressed by a maximal value of a score function [20]. As the search is heuristic, the optimal graph structure found is by definition a local optimum. Nevertheless, many experiments with datasets have shown that results obtained are usually quite good. The scores used are based on log-likelihood with a penalty factor to prevent overfitting due to complexity of the network. The *Bayesian Information Criterion* (BIC) network score [21] is a typically used example of such a score taking into account the Markov equivalence of BNs (different structures represent the same probability distribution). To deal with missing values in data during learning, a data imputation step is required [22].

3 Related Work

In this section, we describe the related work on available visual tools in BN development (Sec. 3.1), review interactive methods in BN validation and modification (Sec. 3.2), and discuss visual explanation methods for BNs (Sec. 3.3).

3.1 Visual Assistance in BN Development

Xie et al. [23] proposed a visual approach to BN development and explanation. Their approach presents an overview about the data using histograms and includes a BN structure representation. A topological order is showing the causal flow within the network, node aggregation, and hidden cross-layer links are used to minimize edge crossings. We got especially inspired by the topological order and hidden cross-layer links, but neglected node aggregation, since our collaborators stated that every node should be shown in the structural representation.

In data-driven BN structure generation, visual approaches have been developed, such as *BayesPiles* [24] and *BayeSuites* [11], [25]. On the one hand, *BayesPiles* [24] allows for BN structure development and visual comparison in large collections of scored, directed networks to identify the best BN structure. Although this tool allows for comparison of many networks using matrix-based visualizations, it neglects the parametrization and requires much user effort to understand and justify the causal flow and reasoning within the network. *BayeSuites* [11], [25], on the other hand, allows users to choose the learning algorithms and represents the learned network in form of a force-directed node-link diagram. Although the tool provides basic information about the different learning algorithms, it requires preknowledge about BNs and the individual learning algorithms. Since our tool is developed and designed for medical researchers without knowledge in CDSS and in BNs, we prefer a data-driven selection of the best learning algorithm. Furthermore, both approaches neglect the integration of prior knowledge and user-centric validation.

3.2 Visualization in BN Validation and Modification

Especially the user-centric tasks of BN validation and modification require interaction facilities. Yi et al. [26] proposed seven general interaction categories organized around the users intent: *select*, *explore*, *reconfigure*, *encode*, *abstract/elaborate*, *filter*, and *connect*. In BN validation and modification, each category is required.

Cypko et al. [27] identified the process in BN validation and modification for expert-modeled BNs starting with (1) quantitative validation, followed by (2) qualitative validation, to (3) modification, and back. For the *quantitative validation* based on available patient cases, they propose statistical methods, such as accuracy, ROC, confusion matrix to compute the models precision. The *qualitative validation* includes the model's ability to predict values. This comprises the model behavior testing on patient records, followed by studying a single node or a sub-network in more detail. In the *modification* step, the causal dependencies and parametrizations are updated by the user. For the latter, Cypko et al. [28] developed a web tool simplifying probability integration for domain experts. In this context, they are using probability ranges and expert's sureness instead of precise probabilities, which got well accepted by domain experts. In our approach, we are focusing on the qualitative validation and modification step, since not all medical researchers are familiar with the statistic scores but qualitative validation of real patient data is similar to their work within clinical routine.

Tonda et al. [29] addressed the problem of the reasonable compromise between interaction and automated optimization of BN development. By integrating user knowledge in the model learning, biases or misclassifications can be avoided costing user fatigue. In their approach, simplistic visualizations, node-link diagrams, arc manipulation, and editable CPTs are integrated. Furthermore, they included a history of changes. In this tool, however, the user has to validate each node regarding the causes and CPTs, which is very time-consuming. For medical researchers, a more time-efficient validation adjusted to their needs is required.

3.3 Visual Explanation Methods for BNs

The explainability of decision support models in development and validation is equally important as in the application. Therefore, we also got inspired by visual approaches addressing the explainability of BNs for application. Müller et al. [14] gave a thorough overview of recent work in explanation methods for BNs in healthcare. They especially emphasized the representation of the causal flow within the network inspired by Wang et al. [30] and the

relevance of individual evidence items for probability distribution computation inspired by Champion et al. [31] as very important for comprehension by medical experts.

Since the visual representation of BNs' structure can be presumed as DAGs, visualization techniques applicable to directed graphs serve as valuable options in representing the structure of BNs. Commonly, DAGs are displayed using either node-link diagrams, tabular/matrix-based visualizations, or implicit tree layouts in less cases [32]. Different topology-driven layout techniques for node-link diagrams, such as force-directed, spectral layouts, or orthogonal layouts, further assist in node order identification and causal flow depiction [32]. For BNs, edge thickness is commonly used to represent the causal strength [33]. Bae et al. [34] investigated the best arrow representation for causal flow depiction resulting in tapered edges performing equally well as arrows and width is preferred over hue. Matrix-based visual approaches, such as DAGView [35], responsive matrix cells [36], and Node Quilts [37] for layered graphs represent clusters in the network in a concise manner but weaken in causal flow depiction. Graph depictions in combination with linguistic expressions allow for easier comprehension of the causal flow within the network, but are not well suited for large-scale models [38], [39], [40].

In common literature, CPTs are presented using matrix-/table-based visualizations, whereas either one table can be used to represent all [41] or multiple tables [42] are representing the individual conditions. The probabilities are either presented using numbers or bars indicating the likelihood of the individual states [42]. Although the representation using multiple tables is easier to read, it weakens in space expensiveness. Stacked bar charts can also represent CPTs [43]. Other depictions, such as 100 human beings with different highlighters and text serve well in probability comprehension but are worse in space utilization [44].

4 Requirement Analysis

We identified the requirements of our approach on a regular basis in interdisciplinary meetings (with two visualization researchers, one expert in BN modeling, and three medical researchers) over six months and workshadowings over multiple years. Furthermore, we used our experience in BN development and validation in healthcare [3], a thorough literature review, as well as a large discussion at the European Network of Individual Treatment in Endometrial Cancer (ENITEC) annual meeting in 2020 [45]. In summary, we identified the following main requirements:

- **R1 Causal Flow and Clinical Workflow.** In contrast to the causal flow based on factors impacting the tumor (disease) behavior, the clinical workflow is represented by the stepwise approach from diagnosis, treatment, and outcome. For example, in reality a tumor has to be present to be identifiable in medical images whereas in clinical routine management, medical images are used to identify the presence of the tumor. This differing causal flow needs to be visually explained to the medical researchers.
- **R2 Data review.** To assess the dataset's representativeness, e.g., distribution, and identify biases in the data at an early stage, a data review including data distribution needs to be provided.
- **R3 BN Development Pipeline.** Each step in the common BN development pipeline including structural and parametrization development and validation, respectively, as well as model validation needs to be supported.
- **R4 Qualitative Model Validation Using Real Data.** Within medical research routine, experts commonly use real patient data to gather new insights and validate guidelines. This familiar process should be adopted to our approach. One medical collaborator stated: "Doctors are trained to think in terms of patients and it makes it easier for them to compare their own mental predictions with those by a model."
- **R5 Usability Over Power.** To reduce the time-expensiveness and complexity of our approach and to avoid long development periods like presented by Cypko et al. [3], a balance between required and optional/hidden development and validation steps need to be identified.
- **R6 Reduce Background Knowledge in BNs.** The tool should be usable without background knowledge in BNs. Thus, visual abstraction and features need to be developed to allow medical researchers to understand, justify, and use our interactive visual approach.

When investigating the available tools for BN development having a graphical user interface and considering the established requirements, no tool serves well enough for application by medical researchers. Although all tools provide key and further features in BN development and validation (**R3**), the complexity of approaches and required prior knowledge discourages medical researchers. A summary comparison provided by a technical author of this paper of common commercial tools against our approach is presented in Table 1. For this, our dataset related to endometrial cancer was used in all frameworks to develop a CDSM.

TABLE 1: Assessment of common BN development tools against the requirements elicited based on a three-staged scale of requirement fulfillment: *partly* ●○○, *large* ●●○, and *fully* ●●●.

Software	R1	R2	R3	R4	R5	R6
GeNIe [7], [8]	●○○	●○○	●●●	●●●	●○○	●○○
Hugin [9]	●○○	●○○	●●●	●○○	●○○	●○○
Netica [10]	●○○	●○○	●●●	●●●	●○○	●○○
NeuroSuites [11], [12]	●○○	●○○	●●●	●●○	●○○	●●○
Bayes Server [13]	●●○	●○○	●●●	●●●	●●●	●○○
our approach	●●●	●●●	●●●	●●●	●●●	●●●

5 Visualization and Interaction Design

In this section, we describe the developed visualization and interaction techniques for each step in BN development and validation (Sec. 5.1-5.3), and detail on the implementation and architecture design (Sec. 5.4). To demonstrate our approach, we are using a dataset focusing on endometrial cancer data and recorded within clinical routine [4]. It contains 18 variables recorded for 763 patients. Our medical collaborators performed a preprocessing step to discretize the variables and impute missing values [4].

We realized the front end of our interactive visual approach as a three page layout according to our simplified steps in the BN development and validation pipeline (Fig. 1) (**R3**). Furthermore, we included an additional introduction page allowing the user to either start with uploading a *.csv file* which contains the dataset for data-driven development or to select an already started session to edit and finish a prepared BN (Fig. 3).

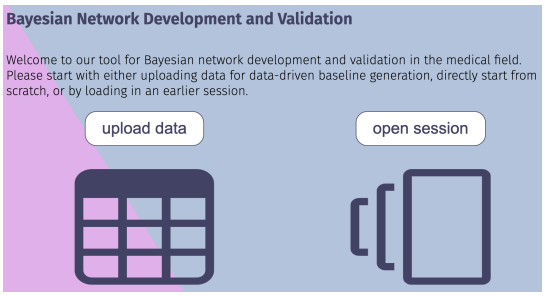


Fig. 3: First page in our interactive visual approach. The user can either upload a dataset for data-driven learning or open an already started session to continue working on it.

5.1 Data Review & Structural Development

In the *data review & structural development* step, medical researchers assess the dataset representativeness as well as the data-driven development of the BN structure. Although the data review and structural development tasks could be split into two domain problems, we decided to present them in one view since an adjustment of the dataset, e.g., omission of a variable, leads to direct changes in the learned BN structure. Furthermore, the data review presentation assists in investigating the data-driven BN structure by comparing the variables and distributions within the variables with each other. The visual representation is depicted in Fig. 4.

5.1.1 Data Review Presentation

To evaluate the dataset representativeness, medical researchers have to investigate and explore the variables included in the selected dataset, the percentage of available data per variable, and the distribution within the variables (**R2**). These information entities allow to identify biases in a concise manner. Furthermore, the total number of patients included in the dataset needs to be highlighted. In general, the data can be assumed as a table of discrete (ordinal and categorical) variables. In case the dataset contains quantitative data, the related variables are binned using quantile-based discretization methods.

In our visual approach, we present the variables within the dataset in a *small multiples* visualization to avoid overplotting issues and allow for concise comparison across variables (Fig. 4 (A)). Each small multiple consists of the variable's name, a histogram representing the distribution among the categories (Fig. 4 (A1)), and a vertical bar indicating the percentage of available data to assess the variable's quality level (Fig. 4 (A2)). Hovering methods assist in assessing the real frequencies for the individual distributions and the available data. The small multiples have a colored border and are ordered by the related *clinical workflow step / group*.

We integrate our novel concept of *clinical workflow steps / groups* to assist users in understanding the reasoning process within BNs as well as in semantically grouping the variables (**R1**) (Fig. 4 (B)). The editable order of steps provides some overarching meta structure that depicts the clinical workflow and assigning variables to workflow steps helps in relating/comparing their causal inter-relations to the clinical workflow. To emphasize the order, we use the standardized diverging color scale from bright pink to purple red. Hovering methods allow for emphasizing related variables through hiding other variables. Furthermore, the border color of each variable widget indicates the association with

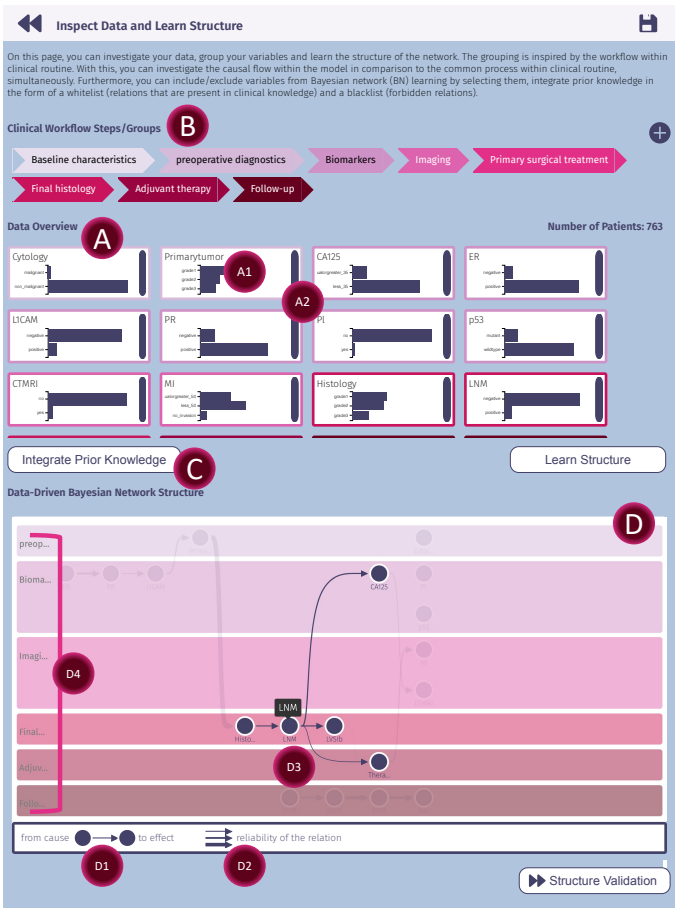


Fig. 4: Interactive Visual Approach to Data Review & Structural Development. We provide the user with a small multiples presentation of the variables contained in the selected dataset (A) including horizontal bars displaying the data distribution (A1) and a vertical bar showing the percentage of available data (A2). A reorderable set of clinical workflow steps / groups is displayed using arrows (B). The user can integrate her prior knowledge for data-driven model development (C) and can investigate the learnt model structure (D). In this context, the variables are positioned in horizontal direction along the causal flow within the network (D1) whereas the nested vertical direction displays the related clinical workflow step / group (D4). The reliability of relations is displayed via edge thickness (D2) and hovering methods assist in identifying related variables (D3).

a specific clinical workflow step / group. The regrouping feature is implemented via a pop-up view opening after clinical workflow step / group selection.

5.1.2 Structural Development Presentation

To generate the data-driven structure of BNs, medical researchers should be able to integrate prior knowledge in the structure learning process, such as proven and disproved causal relations reported in clinical literature (Fig. 4 (C)). Furthermore, the users need to identify the most suitable structure learning method for the selected dataset and need to be able to comprehend, justify, and modify the initially learned BN structure.

In contrast to common BN learning tools and in agreement with requirement **R6** (*Less Background Knowledge in BNs*), we compare the potential network candidates' score *BIC* of the BN

structure generated using six widely applied learning algorithms: “pc.stable”, “gs”, “hc”, “tabu”, “mmhc”, and “h2pc”, and only present the best structure. In agreement with our collaborators, we decided to not present the mentioned score to the user. The reason is that the values have no specific range and we wanted to avoid confusion and necessity of having vast background knowledge in BNs (**R5, R6**). This is a valid procedure since we integrate a user validation step (Sec. 5.2) in our approach later on. Prior knowledge can be integrated as input to the learning algorithm either in form of a “blacklist” and “whitelist” for forbidden and enforced constraints, respectively.

The BN structure can be assumed as a DAG model. Therefore, the presentation of the network structure (Fig. 4 (D)) is inspired by DAG presentations, more precisely, the causal flow depiction [30], as well as the Sugiyama layout in combination with cross-layer minimization to minimize edge crossings [23]. Thus, we present only those edges between two layers next to each other and not over multiple layers to avoid visual clutter. In a left-to-right manner, we represent the causal flow within the network (from cause to effect) (Fig. 4 (D1)). We use edge thickness to represent the reliability (arc strength) of the relation in a range of [0;1] (Fig. 4 (D2)). Hovering methods per node emphasize all related nodes (causes and effects), which might be hidden in static presentation due to the applied cross-layer minimization (Fig. 4 (D3)). In agreement with our collaborators, we decided to conceal node states to avoid overstraining the users (**R5**).

Inspired by the visualization of nested graphs [46], we integrate the clinical workflow steps / groups in a top-to-bottom manner (**R1**) (Fig. 4 (D4)). Nested graphs show the hierarchical structure and containment relationships among the graph nodes. The placement of nodes regarding the causal flow within the network **and** the clinical workflow steps / groups assist in justifying the reasoning within the BN structure.

5.2 Structural Validation

After the structural development, the data-driven structure needs to be validated by the medical researcher before developing the parametrization. Cypko et al. [47] defined the workflow for the expert-based structural validation starting with the data type verification, followed by variable identifier setting, node states validation, parent nodes setting, and examination methods selection. In this workflow, only the relationships to the parent nodes have to be validated since the CPTs take into account combinations of parent node states only. Inspired by the presented workflow, we developed a visual approach assisting in structural validation (Sec. 5.2.1). By means of iterative discussions with our collaborating medical experts we simplified and improved this structural validation step (Sec. 5.2.2). We will discuss the design development in the following.

5.2.1 Initial Structural Validation

Inspired by the workflow for BN development proposed by Cypko et al. [47], our initial approach integrated a presentation of the network structure and a validation of the stated information entities except for the examination method selection, since this is only serving well a purely expert-modeled BN (Fig. 5). Thus, the user has to validate each node before going forward with the parametrization learning based on the selected dataset.

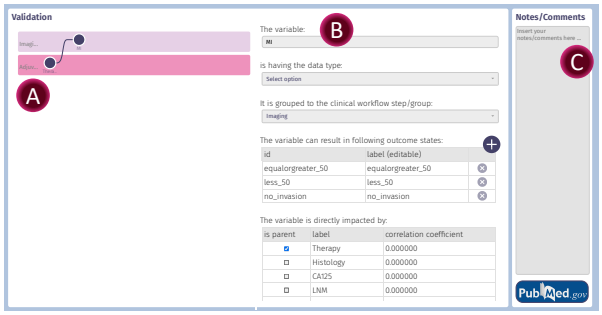


Fig. 5: Initial Visual Representation of the Structural Validation. The view is divided into three parts. The left part (A) displays the variable under investigation and its parent and child nodes on the left and right side, respectively. The central part (B) provides facilities for node validation inspired by Cypko et al. [47]. The right part allows for adding comments and literature search through a link to the medical search engine “PubMed”.

The validation view is divided into three parts. The left part (Fig. 5 (A)) presents the node under validation, the related parent and child nodes, as well as the clinical workflow-steps / groups. This network presentation is focusing on the sub-graph of the node under validation to assist in identifying related nodes while simplifying and reducing the complexity. To improve familiarity, this view adapts the whole network presentation from the structural development presentation (Fig. 4 (D)).

The middle part (Fig. 5 (B)) is based on the validation workflow proposed by Cypko et al. [47] and includes the editable categorical data: variable name, data type, as well as the related clinical workflow step / group. Furthermore, the medical researcher is able to modify the possible node states as well the potential parent nodes. To guide users in identifying potential parent nodes, we are ordering them by their correlation coefficient *Chi-Square* for qualitative data with the node under validation, since correlations are indicators for potential causal relations [48]. To assist users in finding related publications to add in the notes/comments field, a direct link to the medical search engine *PubMed* (https://pubmed.ncbi.nlm.nih.gov/) is included (Fig. 5 (C)).

Our iterative design meetings gave us more insights into our interactive tool design. Currently, the user has to validate each variable which results in repeating and tedious tasks. We recognized that this was too time-expensive and required complex user validation entries making the structural validation inconvenient for medical researchers, hence demanding an improved structural validation step.

5.2.2 Improved Structural Validation

In various design meetings, our project members and collaborators discussed the importance of each step in the validation workflow proposed by Cypko et al. [47] (**R5**). We identified that the integration of the data type is similar to the already implemented clinical workflow step / groups and therefore, can be omitted. Furthermore, the validation of the node states was deemed as unnecessary by our collaborating medical researchers given that it is included in the data preprocessing step. The parent relations, however, are still required and need to be validated. Furthermore, visual guidance in sub-graph identification assists in identifying missing relations.

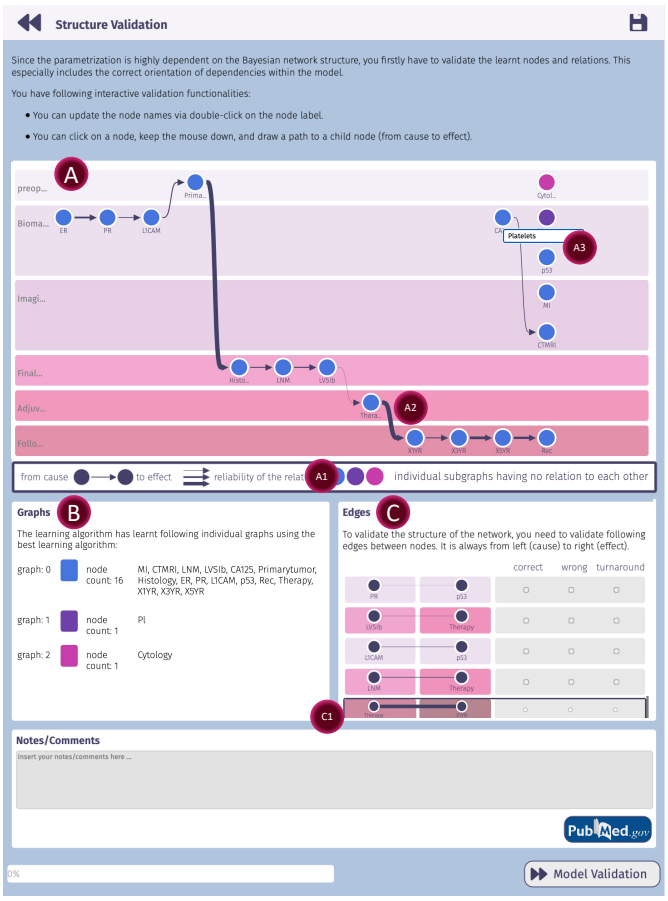


Fig. 6: Improved Interactive Structural Validation View. Our visualization consists of the network structure presentation (A) with node colors encoding independent sub-graphs (A1). Since we apply cross-layer minimization to minimize edge crossings, not all edges are presented in the static visualization but visible through hovering methods. An overview of the individual sub-graphs is given below (B). Interaction facilities allow for adding relations (A2) and relabeling variables (A3). The edge view (C) provides relationship validation facilities and emphasizes user-added relationships using a blue border (C1).

Based on the updated requirements, our visual approach starts with the interactive network structure presentation (Fig. 6 (A)). In this context, we use the rainbow color scheme for unordered values to distinguish independent sub-graphs (Fig. 6 (A1)). This highlights missing connections or variables independent for and from other variables within the dataset. For example, the variable *Platelets* is not dependent on or has a dependency to another variable within our dataset which is highlighted by a different node color – purple (Fig. 6 (A3)). A detailed description of the sub-graphs including the related variables is given below the network presentation (Fig. 6 (B)). Interaction facilities in the network view, such as node name modification using double click (Fig. 6 (A3)) or drawing edges between nodes to add relations (Fig. 6 (A2)) further supports the interactive structural validation. In this context, newly added relations always have a reliability of 100% since the user integration is regarded as reliable. Additionally, the time-consuming and tedious repeating validation per node (Sec. 5.2.1) is replaced by a faster and clearer edge validation using radio buttons (**R5**) (Fig. 6 (C)). The user validates by choosing whether

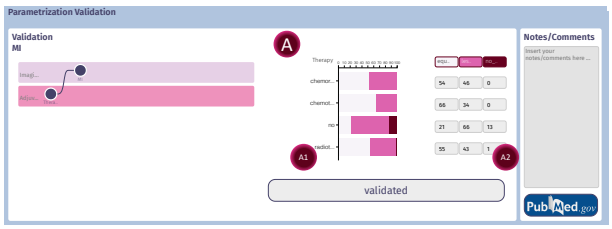


Fig. 7: Initial Visual Representation of the Parametrization Validation. Similar to the initial structural validation (Fig. 5), the view is divided into three parts with parametrization validation in the center (A). Unlike in structural validation, we apply stacked bar charts (A1) to represent the CPTs and editable text boxes to allow for user-specfic parametrization validation (A2).

a relation is correct, wrong, or the direction is reversed. In this view, user-added relations are emphasized with a dark blue border (Fig. 6 (C1)).

5.3 Parametrization Development, Parametrization Validation, and Model Validation

After completing the structural validation step, the parametrization is computed based on the available data and needs to be validated for correctness focusing on the node under validation and its parent nodes. We integrated the *parametrization development* without user adjustable parameters to reduce the complexity of our approach (**R5**) and to omit the demand for background knowledge in BNs (**R6**). For parameter learning, we use the default algorithm implemented in the *bnlearn* package in R – the "Maximum Likelihood parameter estimation" [49].

5.3.1 Initial Parametrization Validation

To investigate and validate the CPTs it is common that users get an overview of the parent node state probabilities and modify them using interaction. Similar to the initial structural validation (Sec. 5.2.1), the view is divided into three parts having the parametrization validation in the center (Fig. 7 (A)). Inspired by the visual approach developed by Hassall et al. [43], we display the CPTs using space-efficient stacked bar charts (Fig. 7 (A1)). Each stacked bar chart represents the probability distribution among the validated node states for a combination of parent node states. Other potential depiction forms, such as multiple tables [42] or 100 human beings with different highlighted regions [44] hamper efficient space utilization and were rejected by our medical collaborators. Editable text boxes allow for probability distribution updates (Fig. 7 (A2)). In validating the parametrization, the user has to check each node in the network by repeating interactions.

Similar to the iterative improvements for the structural validation step (Sec. 5.2.1), we identified that the common and presented procedure in parametrization validation is very time-consuming, abstract and tedious for the user. Furthermore, through discussions and by the means of iterative design meetings, we determined that the "parametrization development", "parametrization validation", and "model validation" can be combined to the "model parametrization and validation" step. Therefore, we developed an improved model parametrization and validation view.

5.3.2 Improved Model Parametrization and Validation

To significantly reduce the expense of time and complexity of our approach, we developed a novel approach allowing for simultaneous model parametrization and validation using real patient

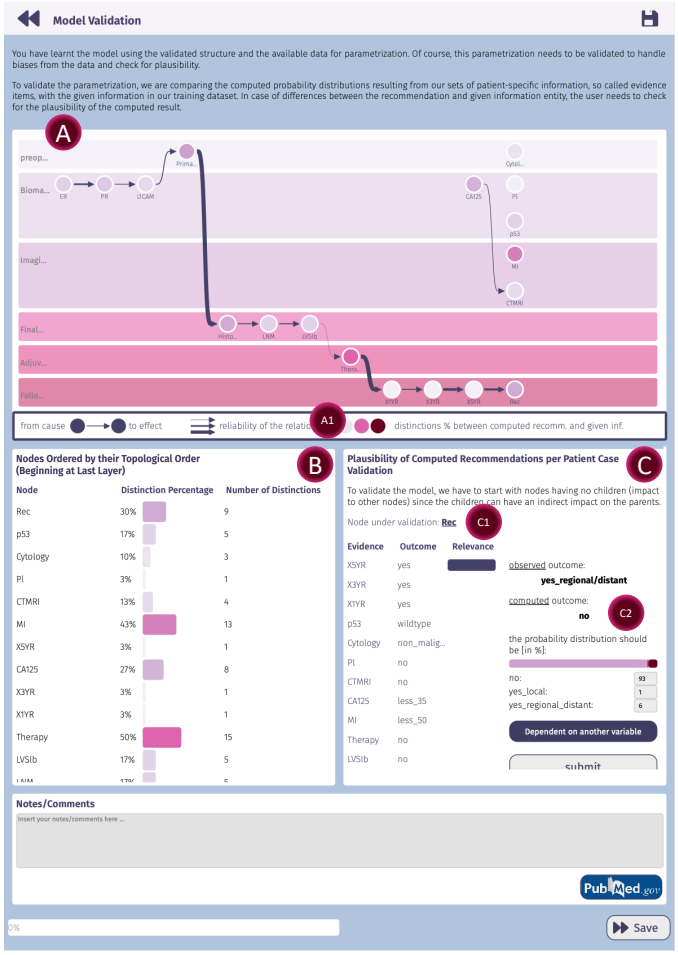


Fig. 8: Interactive Model Parametrization and Validation View. This view displays the network structure (A) and uses gradient color from bright pink (low) to dark purple (high) to encode the percentage of discrepancies between computed recommendation and given information (A1). A bar chart displays an overview of the related discrepancies per node (B). We use real patient data for parametrization validation (C). In this context, we present a patient's available information entities ordered by their relevance of influence for the computed outcome (C1) and visually compare the given information and the computed recommendation for this patient (C2). The user is able to modify the probability distribution resulting from this set of available information in editable text boxes (C3). In case the probabilities sum up to 100%, the updated probability distribution can be submitted.

cases (**R4**). Our approach is inspired by the qualitative validation of BNs described by Cypko et al. [47].

In contrast to their initial parametrization validation, which was purely based on clinical expertise and literature, we also consider medical data. To perform a qualitative validation, we use 25% of all cases in the available dataset, which is commonly used in machine learning for testing, and compute the probability distribution for each node inspired by the "leave-one-out crossvalidation". In this context, we are consulting all information available for each patient, compute the probability distribution for each variable using all other information entities, and compare the result with the given information in the dataset. In case the computed recommendation and the given information are equal, we assume that the model is working correctly and the data is accurate. Otherwise, the user has to validate whether the model needs to be adjusted or if the dataset is biased in some way.

For example, assume the oversimplified BN presented in Fig. 2 and imagine a patient who has no "lymph node metastases" and has received no "therapy". Based on the presented BN, the recommendation for *tumor recurrence* for this patient is *no recurrence*. For this patient, however, a tumor recurrence was observed. Now, the user has to validate whether the parametrization is wrong and needs to be updated, the data was wrongly collected, or if the tumor recurrence also depends on other variables. Our medical collaborators preferred this qualitative validation procedure since it is familiar to clinical decision-making (**R1**).

We aggregate the cases with discrepancies between computed recommendation and available information per node and represent the related relative frequencies in a gradient color scale from bright pink to dark purple representing low and high relative frequencies of discrepancies, respectively (Fig. 8 (A1)). Furthermore, we use a bar chart to emphasize variables having many discrepancies below the network view (Fig. 8 (B)). In our visualization, the bar chart is ordered by the layers in the BN structure, starting with the last layer. Reason for that is that the last-layered variables need to be validated first to avoid indirect impacts from child nodes to parent nodes. For example, looking at our oversimplified BN (Fig. 2) again, there is a effect to cause dependency from "tumor recurrence" to "therapy". Since CPTs just represent the cause to effect relationships, however, we have to validate the cause to effect relationships beforehand and thus, start with validating the last layers. After the variables from the last layer are validated, the variables from the penultimate layer can be validated and so on.

To allow for a variable validation per patient case, we present all patient information in a list inspired view on the right side below the network view (Fig. 8 (C1)) with the exception of variable information under validation, e.g., "Rec" (tumor recurrence) in this example. The information entities are ordered by their relevance for the computed recommendation inspired by the relevance computation proposed by Müller et al. [14]. Bars further represent the individual relevancies. Thus, users can identify the influential information entities on the node under validation in a concise fashion and justify the reasoning process. Furthermore, we display the observed and most probable computed outcome in a comparative manner (Fig. 8 (C2)). Next, we decided to only present the sub-part of the whole CPT dependent on the patient-specific parent node states instead of the whole CPT. For instance, we display the sub-part of the CPT for "Rec" for the parent node state "X5YR – yes" only and hide the sub-part for "X5YR – no" do avoid overstraining the user (Fig. 8 (C2)). We use a stacked bar chart to save space and supplement it by editable text inputs for user validation (Fig. 8 (below C2)). Before being able to submit the updated probability distribution, we check if the probabilities sum up to 100%. In our visual approach, we only show the relevant parts and thus, reduce the complexity and understanding effort, as well as simplify the user input. If a node is dependent on other variables, the user has to go back to the structural validation page (Sec. 5.2). In this case, however, the user-specific modifications in parametrization are lost since they highly depend on the network structure.

5.4 System Architecture and Implementation

We implemented our tool as an interactive client-server web environment utilizing JavaScript, D3.js [50], Python, R, and Flask

framework [51]. We use the *bnlearn* library [49] in R to learn the structure and parametrization of the BNs. For the internal BN representation in python, integration of user modifications, and inference computation, we are using the SMILE engine for python (pySMILE) [52]. We utilize the standardized *dsc* format as suggested in the Bayesian Network Interchange Format proposal [53] to transfer the BN between python and R.

6 Evaluation Study

In this section, we describe the process of evaluating our visual assistance tool and present the evaluation study results. In this context, we focus on the assessment of the usability, comprehensibility, and medical relevance of our visual approach to identify potential weaknesses and improvements. This is essential before asking medical researchers to perform the time-consuming process of developing a CDSM.

6.1 Study Design

We performed the evaluation study as an one-on-one *expert review* [54] since our tool is designed for medical researchers and our available data focuses on the medical domain of endometrial cancer management [4]. Thus, our target users have to have background knowledge about the data to use the tool. Each evaluation study took approximately one hour via remote screen sharing and screen control. We started each evaluation session with a short introduction of our motivation and our goals, followed by a brief tutorial explaining different aspects of the tool. This took approximately 25 minutes. Then, we asked the participants to complete a list of tasks exploring different parts of the tool and BN development and validation on their own. Furthermore, the participants were asked to answer a questionnaire about the usability and medical relevance of our visual approach.

Participants. In total, we asked six participants (3 male, 3 female) with medical background in endometrial cancer to assess our visual approach and fill out the evaluation survey. Five of the participants were not involved in the design and implementation process and are no co-authors of the paper. One participant was younger than 30 years, one between 31 to 40 years, two between 41 to 50 years, and two participants were 61 to 70 years old. Five medical researchers have a PhD or higher while one is currently in the process of doing her PhD. All participants consider themselves to have a good or very good medical knowledge in endometrial cancer. The experience and familiarity with interactive visualizations was rated more controversial (from no experience to good experience). All participants rated their own experience from no to medium experience. One participant said that she has no experience with CDSS, three have knowledge about the concept but without practical experience, and two have knowledge about the concept with practical experience in application. No participant was familiar with the development and validation of CDSS. The experience with interactive visualizations to investigate CDSS was also rated very low. Although one participant stated a medium experience, three rated their experience as poor.

Tasks. The tasks that the participants were asked to solve are inspired by the the "low-level components of analytic activity in information visualization" by Amar et al. [55] and the "task taxonomy for graph visualization" by Lee et al. [56]. In this context, we considered for example tasks related to *value retrieval* (investigating data distributions in the data review presentation), *filtering* (which variables are grouped to a clinical workflow step?), *adjacency* (Which nodes are directly related? What is a causal reason of node x?), and *modification* (Please, add the missing relations.). The participants fulfilled the tasks mainly without the help of the interviewer and their task results were recorded.

Questionnaire. Afterwards, we asked the participants to fill out a questionnaire consisting of eight demographic questions, 12 general questions about the tool, seven questions related to the visualization design and interaction techniques, and six development and validation step-specific questions, followed by some feedback and suggestions. In total, we asked 35 questions including eight multiple choice questions, 24 questions answered on a five-point-likert scale, and three open-ended questions. The questionnaire is inspired by the *ICE-T survey form* based on *Value-Driven Evaluation of Visualizations* [57] and the *Heuristic Approach to Value-Driven Evaluation of Visualizations* [58]. In contrast to the ICE-T survey form, we shortened the questionnaire and tailored the questions to our visual approach since our tool integrates too many features making it difficult to assess them all at once. For example, instead of asking the participant to assess questions like *"The visualization presents the data by providing a meaningful visual schema."*, we specify the sentence to explicit parts of our visual encodings like: *"The application of colors assists in sub-graph identification."*. A detailed description of the tasks, questions, and user-specific answers can be found in the supplementary material.

Validity and Limitations. The results of our expert feedback study are based on the participants' subjective opinions about our tool and their subjective knowledge about the data. Due to the long evaluation study design, we recognized some fatigue. An iterative evaluation study design over multiple sessions could have improved this issue. However, our evaluation study still provides valuable insights into the usefulness and medical relevance of our visual approach. Furthermore, no statistically significant results can be obtained based on data from six participants only. Therefore, we are focusing on a qualitative evaluation of the results.

6.2 Findings and Discussion

Based on the recorded answers and findings, we identified that the participants' background knowledge in CDSS and interactive visualization plays a crucial role in the expert feedback. For example, one technical advanced participant stated: "*This is fun to do, I can use the tool all day long!*" whereas a technical inexperienced participant claimed: "*I find the causality not easy to understand and can often think in both directions*".

Usability. In general, participants rated the learning curve from moderate to fast and were able to use the tool mainly independently. One participant stated: "At first, the tool was eyeballing and very complex. But after the introduction, I was able to use it. A longer familiarization phase is needed in order to use it within medical research." All participants indicated that our tool allows for visual exploration with limited or no background knowledge in computer science. Thus, we conclude that background information about the data and related domain expertise is sufficient in order to develop and validate decision support models using our visual approach. The split up of our approach on three pages was rated either as excellent or as moderate in order to avoid overwhelming the user with too many information.

Medical Relevance. All participants stated that our tool can be used to develop CDSM, and to generate new hypothesis about the data. These topics were rated very high since there is a high

demand for clinical decision support. One participant indicated that our approach could not only assist in development and validation of CDSMs but also in their hospital-specific adjustment. For example, she explained that there are hospital-specific differences between applied therapy forms due to available specialists and/or technical support. Furthermore, the participants said that our approach helps them in justifying the reasoning within decision-support models and assists in building trust.

Sufficient Structure Presentation. In general, our participants rated that the concept of BNs is hard to understand. Using our approach, however, the participants were able to use, validate, and justify the causal flow within the network. The application of colors and the initially hidden edges over multiple layers were rated as comprehensible and good, respectively.

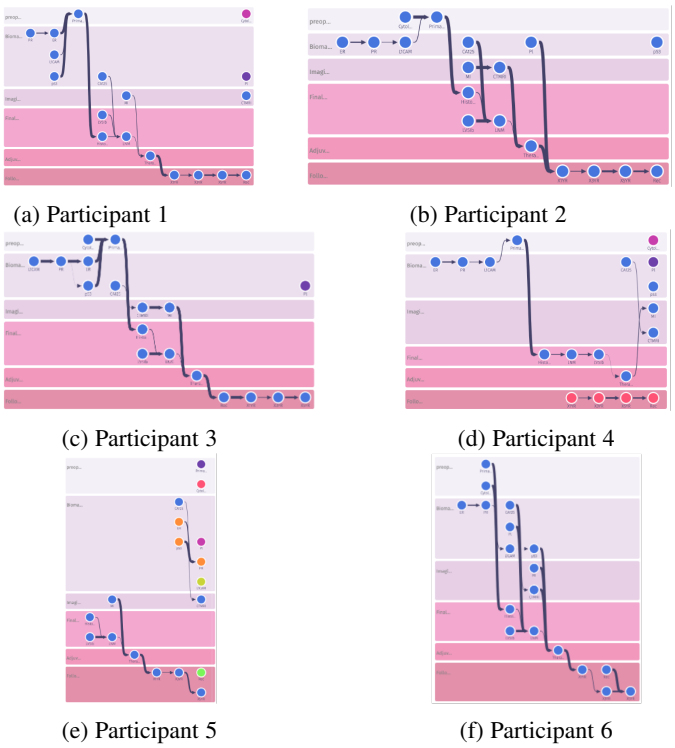
Improvable Parametrization Validation. Although our participants liked the integration of real patient data within the parametrization validation step, the concept of real probability distributions was rated as hard to assess. A participant suggested the application of percentage ranges (unsure, sure, very sure) to improve both, CPT comprehension and validation.

Comparison of Validated Models. During the evaluation study, we asked the participants to develop and validate a primitive BN based on the same dataset focusing on endometrial cancer patients. We collected the resulting BN after each evaluation study and compared them (Fig. 9). Ideally, all derived networks should be equal. However, we recognized that the created networks differ both in structure and in parametrization due to different medical background knowledge and experience. In this context, however, we have to emphasize that the users were not asked to develop and validate a BN directly serving as a decision-support tool due to time reasons but were asked to integrate their knowledge without an extensive literature review. Furthermore, reasons for differing networks were based in the participants' fatigue after the long evaluation study as well as in their varying technical background knowledge. The differences between the resulting BNs emphasize the need of a collaborative development and validation approach inspired by our identified requirements and visualizations.

Additional Features. During the evaluation study and in collaboration with our study participants, we recognized a number of missing additional features. For example, users asked for a history and emphasis of changes in BN validation. Currently, the modifications are not stored and we did not provide visual comparison between the initial data-driven BN and the network after validation. This feature, however, would further assist in validation and improve usability. Furthermore, the participants indicated that the emphasis of the shortest path between two nodes of interest would assist in network justification and understanding the reasoning process.

7 Conclusion and Future Work

In this work, we presented an interactive visual tool assisting medical researchers in developing and validating CDSMs based on BNs. In contrast to common tools in BN development and validation, our approach considers the workflow within clinical routine as well as an abstracted and shortened pipeline in BN generation adjusted to medical researchers' needs. In an evaluation study with 6 experienced medical researchers having domain-specific background in endometrial cancer, we assessed the usefulness and medical relevance of our approach as very useful.



(a) Participant 1 (b) Participant 2

(c) Participant 3 (d) Participant 4

(e) Participant 5 (f) Participant 6

Fig. 9: Structures of Bayesian networks in endometrial cancer management resulted from the evaluation studies with different participants.

Although our tool is especially designed for and in collaboration with medical researchers, we believe that parts of our project can also be applied in other domains, such as biology, finance, and climate. Having the related background knowledge about the data and domain, we estimate that our approach allows users to generate and validate BNs mainly independently. However, we have to investigate our hypothesis in a future study with experts and datasets from various domains. Further future work includes the integration of structural and parametrizational comparative views allowing users to track their modifications. Additionally, a long-time evaluation study with multiple medical researchers developing a real CDSM using our interactive visual approach could emphasize interesting and relevant features as well as lead to useful BNs for treatment decision support.

Acknowledgments

The research leading to this work was supported by the Federal State of Saxony-Anhalt, Germany (FKZ: I 88).

References

[1] I. Vourgidis, S. J. Mafuma, P. Wilson, J. Carter, and G. Cosma, "Medical expert systems – a study of trust and acceptance by healthcare stakeholders," in *Advances in Computational Intelligence Systems*, A. Lotfi, H. Bouchachia, A. Gegov, C. Langensiepen, and M. McGinnity, Eds. Cham: Springer International Publishing, 2019, pp. 108–119.

[2] P. Arora, D. Boyne, J. J. Slater, A. Gupta, D. R. Brenner, and M. J. Druzdzel, "Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine," *Value in Health*, vol. 22, no. 4, pp. 439–445, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1098301519300579

[3] M. A. Cypko, *TreLynCa: A Tumor Board Decision Model for Laryngeal Cancer*. Wiesbaden: Springer Fachmedien Wiesbaden, 2020, pp. 57–68. [Online]. Available: https://doi.org/10.1007/978-3-658-32594-7_6

[4] C. Reijnen, E. Gogou, L. van der Putten, N. Visser, K. van de Vijver, M. Santacana, J. Bulten, E. Colas, A. Gil-Moreno, A. Reques, G. Mancebo, C. Krakstad, J. Trovik, I. Haldorsen, H. Engerud, J. Huvila, M. Koskas, V. Weinberger, L. Minar, E. Jandakova, A. van der Wurff, X. Matias-Guiu, F. Amant, H. Küsters-Vandevelde, J. Ramjith, L. Massuger, M. Snijders, P. Lucas, and J. Pijnenborg, "Development and validation of an endometrial carcinoma preoperative bayesian network using molecular and clinical biomarkers (ENDORISK): an ENITEC collaboration study," *International Journal of Gynecologic Cancer*, vol. 29, no. Suppl 4, pp. A6–A7, 2019.

[5] H. Amirkhani, M. Rahmati, P. J. Lucas, and A. Hommersom, "Exploiting experts' knowledge for structure learning of Bayesian networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 11, pp. 2154–2170, 2016.

[6] M. Velikova, P. J. Lucas, M. Samulski, and N. Karssemeijer, "On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks," *Artificial intelligence in medicine*, vol. 57, no. 1, pp. 73–86, 2013.

[7] M. J. Druzdzel, "GeNIe: A development environment for graphical decision-analytic models," in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 1999, p. 1206.

[8] L. BayesFusion, "Genie modeler," *BayesFusion, LLC, Pittsburgh, PA, accessed Nov*, vol. 30, p. 2017, 2017.

[9] S. K. Andersen, K. G. Olesen, F. V. Jensen, and F. Jensen, "HUGIN-A Shell for Building Bayesian Belief Universes for Expert Systems," in *IJCAI*, vol. 89, no. August, 1989, pp. 1080–1085.

[10] N. S. Corp., "Netica application," https://www.norsys.com/netica.html, accessed: 2021-04-21.

[11] M. Michiels, P. Larranaga, and C. Bielza, "BayeSuites: An open web framework for massive Bayesian networks focused on neuroscience," 02 2020.

[12] S. P. Blanco, "Visualization and interpretation in large Bayesian networks," Master's thesis, Technical University of Madrid, July 2019.

[13] Bayes Server Ltd, *Bayes Server*, 2020. [Online]. Available: https://www.bayesserver.com(accessedOctober19,2020)

[14] J. Müller, M. Stoehr, A. Oeser, J. Gaebel, M. Streit, A. Dietz, and S. Oeltze-Jafra, "A visual approach to explainable computerized clinical decision support," *Computers & Graphics*, vol. 91, pp. 1 – 11, 2020.

[15] D. G. Pfister, S. Spencer, D. M. Brizel, B. Burtness, P. M. Busse, J. J. Caudell, A. J. Cmelak, A. D. Colevas, F. Dunphy, D. W. Eisele *et al.*, "Head and neck cancers, version 2.2014," *Journal of the National Comprehensive Cancer Network*, vol. 12, no. 10, pp. 1454–1487, 2014.

[16] Y.-C. Chen, T. A. Wheeler, and M. J. Kochenderfer, "Learning Discrete Bayesian Networks from Continuous Data," *J. Artif. Int. Res.*, vol. 59, no. 1, p. 103–132, May 2017.

[17] R. L. Winkler, *An introduction to Bayesian inference and decision*, 2nd ed. Gainesville, FL: Probabilistic Pub, 2003.

[18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[19] T. Verma and J. Pearl, *Equivalence and synthesis of causal models*. UCLA, Computer Science Department, 1991.

[20] M. Scutari, "Learning Bayesian Networks with the bnlearn R Package," *Journal of Statistical Software, Articles*, vol. 35, no. 3, pp. 1–22, 2010. [Online]. Available: https://www.jstatsoft.org/v035/i03

[21] G. Schwarz, "Estimating the Dimension of a Model," vol. 6, no. 2. Institute of Mathematical Statistics, 1978, pp. 461 – 464.

[22] A. Franzin, F. Sambo, and B. Di Camillo, "bnstruct: an R package for Bayesian Network structure learning in the presence of missing data," *Bioinformatics*, vol. 33, no. 8, pp. 1250–1252, 12 2016. [Online]. Available: https://doi.org/10.1093/bioinformatics/btw807

[23] X. Xie, F. Du, and Y. Wu, "A visual analytics approach for exploratory causal analysis: Exploration, validation, and applications," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020.

[24] A. Vogogias, J. Kennedy, D. Archambault, B. Bach, V. A. Smith, and H. Currant, "BayesPiles: Visualisation Support for Bayesian Network Structure Learning," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 1, Nov. 2018.

[25] S. P. Blanco, "Visualization and interpretation in large Bayesian networks," Master's thesis, Technical University of Madrid, Madrid, Spain, Jul. 2019.

[26] J. S. Yi, Y. a. Kang, J. Stasko, and J. A. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1224–1231, 2007.

[27] M. A. Cypko, M. Stoehr, M. Kozniewski, M. J. Druzdzel, A. Dietz, L. Berliner, and H. U. Lemke, "Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment," *International journal of computer assisted radiology and surgery*, vol. 12, no. 11, pp. 1959–1970, 2017.

[28] M. A. Cypko, D. Hirsch, L. Koch, M. Stoehr, G. Strauss, and K. Denecke, "Web-tool to Support Medical Experts in Probabilistic Modelling Using Large Bayesian Networks With an Example of Hinosinusitis," *Studies in Health Technology and Informatics*, vol. 216, pp. 259–263, 2015.

[29] A. Tonda, A. Spritzer, and E. Lutton, "Balancing user interaction and control in Bayesian network structure learning," in *11th Biannual International Conference on Artificial Evolution (EA-2013), Bordeaux, France*, 2013, pp. 21–23.

[30] J. Wang and K. Mueller, "Visual Causality Analysis Made Practical," in *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Phoenix, AZ: IEEE, Oct. 2017, pp. 151–161.

[31] C. Champion and C. Elkan, "Visualizing the consequences of evidence in Bayesian networks," *arXiv preprint arXiv:1707.00791*, 2017.

[32] C. Nobre, M. Meyer, M. Streit, and A. Lex, "The State of the Art in Visualizing Multivariate Networks," *Computer Graphics Forum*, vol. 38, no. 3, pp. 807–832, 2019.

[33] J. Bae, T. Helldin, and M. Riveiro, "Understanding Indirect Causal Relationships in Node-Link Graphs," *Computer Graphics Forum*, vol. 36, no. 3, pp. 411–421, Jun. 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13198

[34] J. Bae, E. Ventocilla, M. Riveiro, T. Helldin, and G. Falkman, "Evaluating Multi-attributes on Cause and Effect Relationship Visualization:," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 64–74. [Online]. Available: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006102300640074

[35] E. M. Kornaropoulos and I. G. Tollis, "Dagview: An approach for visualizing large graphs," in *Graph Drawing*, W. Didimo and M. Patrignani, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 499–510.

[36] T. Horak, P. Berger, H. Schumann, R. Dachselt, and C. Tominski, "Responsive Matrix Cells: A Focus+Context Approach for Exploring and Editing Multivariate Graphs," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020.

[37] J. Bae and B. Watson, "Developing and evaluating quilts for the depiction of large layered graphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2268–2275, 2011.

[38] M. Druzdzel and M. Henrion, "Qualitative Verbal Explanations in Bayesian Belief Networks," *Artificial Intelligence and Simulation of Behavious Quarterly*, vol. 94, pp. 43–54, 1996.

[39] C. Elsaesser and M. Henrion, "Verbal expressions for probability updates How much more probable is "much more probable"?" in *Machine Intelligence and Pattern Recognition*. Elsevier, 1990, vol. 10, pp. 319–328.

[40] S. Latif, K. Su, and F. Beck, "Authoring combined textual and visual descriptions of graph data." in *EuroVis (Short Papers)*, 2019, pp. 115–119.

[41] C.-H. Chiang, P. Shaughnessy, G. Livingston, and G. Grinstein, "Visualizing graphical probabilistic models," *UML CS*, 2005.

[42] E. Chuprikova, A. M. MacEachren, J. Cron, and L. Meng, "Visual analytics-enabled Bayesian network approach to reasoning about public camera data," in *EuroVis workshop on reproducibility, verification, and validation in visualization (EuroRV3). Prague: The Eurographics Association*, 2018.

[43] K. L. Hassall, G. Dailey, J. Zawadzka, A. E. Milne, J. A. Harris, R. Corstanje, and A. P. Whitmore, "Facilitating the elicitation of beliefs for use in Bayesian Belief modelling," *Environmental Modelling & Software*, vol. 122, p. 104539, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S136481521930221X

[44] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. J. Han, and R. Chang, "Improving Bayesian Reasoning: The Effects of Phrasing, Visualization, and Spatial Ability," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 529–538, 2016.

[45] "European network of individual treatment in endometrial cancer (enitec)," https://www.esgo.org/network/enitec/, accessed: 2021-04-21.

[46] D. Harel, "On visual formalisms," *Communications of the ACM*, vol. 31, no. 5, pp. 514–530, 1988.

[47] M. A. Cypko, "Therapy decision support system using Bayesian networks for multidisciplinary treatment decisions," Ph.D. dissertation, University of Leipzig, Leipzig, Germany, 2017.

[48] H. A. Simon, "Spurious correlation: A causal interpretation," *Journal of the American statistical Association*, vol. 49, no. 267, pp. 467–479, 1954.

[49] M. Scutari and R. Ness, "Package 'bnlearn'," *Bayesian Network Structure Leaning, parameter learning and inference and inference, R package version*, vol. 4, 2020.

[50] M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.

[51] M. Grinberg, *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2018.

[52] BayesFusion, "SMILE: Structural Modeling, Inference, and Learning Engine," 2020. [Online]. Available: https://www.bayesfusion.com/smile/

[53] R. Dodier, "A Revision of the Bayesian Network Interchange Format," in *1998 Conference on Uncertainty in Artificial Intelligence*, 1998.

[54] N. Elmqvist and J. S. Yi, "Patterns for visualization evaluation," *Information Visualization*, vol. 14, no. 3, pp. 250–269, 2015.

[55] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 2005, pp. 111–117.

[56] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp. 1–5.

[57] J. Stasko, "Value-driven evaluation of visualizations," in *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, 2014, pp. 46–53.

[58] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko, "A heuristic approach to value-driven evaluation of visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 491–500, 2018.

Juliane Müller-Sielaff joined the *Medicine and Digitalization* (MedDigit) group at the Department of Neurology, Univ. of Magdeburg, Germany as a doctoral researcher in 2018. She received her M.Sc. in Computer Science in 2016 from TU Braunschweig. Her research interests include information visualization, the visual analysis of medical data, and visual explainability of model-based clinical decision support.

Seyed Behnam Beladi started collaborating with the *Medicine and Digitalization* (MedDigit) group at the Department of Neurology, Univ. of Magdeburg, Germany in 2020. He is a Digital Engineering master student at Otto von Guericke University Magdeburg. His research interests are visual analytics and data science specially in medical applications' domain.

Stephanie W. Vrede is a PhD-student at the department of Obstetrics and Gynecology, Radboud university medical center, Nijmegen, the Netherlands. In 2018, she received her Medical Degree (M.D.) from the university of Utrecht, the Netherlands. Her research interest are in endometrial cancer and gynecology oncology.

Monique Meuschke is a Post Doctoral Research Fellow at the University of Magdeburg (UoM), and the University of Jena, Germany. In 2015, she received a master in Computational Visualistics from the UoM. In 2019, she received a Ph.D. in Computer Science from the UoM. Her research interests are the visual analysis and exploration of medical blood flow data as well as information visualization and visual analytics for medical applications.

Peter J.F. Lucas received the PhD degree in Mathematics and Computer Science from the Free University, Amsterdam, the Netherlands in 1996, and has in addition an MD degree from Leiden University. Currently, he is a full professor in Artificial Intelligence at the Digital Society Institute of the University of Twente, Enschede, The Netherlands. His research interests include probabilistic graphical models, statistical machine learning, decision-support systems, and medical informatics.

Johanna M.A. Pijnenborg received her PhD degree in Medicine from the University Hospital Maastricht, the Netherlands in 2005. Currently, she is working as a gynecological oncologist at the Radboud university medical center Nijmegen, The Netherlands and member of well-known societies and networks focusing on gynecological diseases, such as the chair of the ENITEC network. Her research interests are gynecologic oncology, minimal invasive surgery, pathology, and translational research.

Steffen Oeltze-Jafra heads the working group *Medicine and Digitalization* at the Department of Neurology, Univ. of Magdeburg, Germany. In 2016, he received a habilitation in Computer Science, in 2010 a Ph.D. in Computer Science, and in 2004, a diploma in Computational Visualistics from Univ. of Magdeburg. His research interests are in the quantitative analysis of clinical routine data, the visual analysis of biomedical data and in model-based clinical decision support.