

Interactive Decision Support: A Framework to Improve Diagnostic Processes of Cancerous Diseases Using Bayesian Networks

Patrick Philipp, Jürgen Beyerer
Vision and Fusion Laboratory IES
Karlsruhe Institute of Technology KIT
Karlsruhe, Germany
Email: p.philipp@kit.edu

Sebastian Robert, Jürgen Beyerer
Fraunhofer Institute of Optronics, System
Technologies and Image Exploitation IOSB
Karlsruhe, Germany

Dirk Hempel
Steinbeis-Transfer-Institut
Klinische Hämatonkologie
Donauwörth, Germany

Abstract—Recommendations for actions included in a Clinical Practice Guideline (CPG) provide a reference framework for medical experts during diagnostic processes. To support the implementation of these recommendations, we propose an interactive decision support. In order to realize this, the diagnostic processes in the CPGs of Mantle Cell Lymphoma (MCL) and Multiple Myeloma (MM) are formalized using activities of the Unified Modeling Language (UML). Based on UML activities, a Bayesian Network is generated. The resulting models enable an assistance function allowing for patient specific CPG recommendations and subsequently for a suitable as well as personalized diagnosis embedded in an interactive Decision Support System (DSS).

Keywords: Diagnostic Processes, Bayesian Networks, Unified Modeling Language, Assistance, Planning, Decision Support

I. INTRODUCTION

Modern medicine offers very diverse approaches for diagnosis, which are rapidly evolving through intensive research. As a result, there is a huge amount of publications – e.g. a search in Pubmed using the term “multiple myeloma diagnosis” yields about 23.000 results [1]. For a decision maker a profound inquiry concerning a specific topic can therefore be extremely time consuming. Moreover, it is becoming increasingly difficult for an individual healthcare professional to keep his or her knowledge up-to-date and identify suitable diagnostic options for a patient by hand [2].

Clinical Practice Guidelines (CPGs) incorporate consolidated medical expertise which is condensed into general recommendations of actions. High quality CPGs are proposed by groups of experts considering the current state of the art [3]. As a result, for an individual medical practitioner, CPGs open up a reference framework for contemporary diagnostic actions [4]. Especially for physicians who practice outside major oncology centers, CPGs can be a valuable source of knowledge. A distribution of CPGs by print media (passive dissemination) has proven to have only little effect on actual

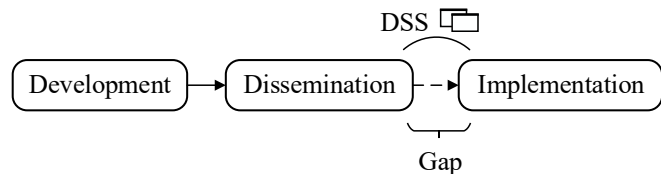


Fig. 1. The figure shows the stages of a Clinical Practice Guideline (CPG). First, the recommendations are developed by an expert team. In a second step they are disseminated. However, there is a gap between the theoretical knowledge (given by the CPG) and the practical actions performed by the medical practitioner. A Decision Support System (DSS) can help to bridge this gap and lower the barriers of CPG implementation. Modified from [5].

practitioners behavior [6] [7]. Therefore, we propose an interactive decision support (cf. Fig. 1) during the diagnostic process to bridge the gap between theoretical knowledge and practical solutions [5].

In this contribution we elaborate a formalization of CPGs concerning Mantle Cell Lymphoma (MCL) and Multiple Myeloma (MM) using activities of the Unified Modeling Language (UML). Based on this, a Bayesian Network (BN) is generated in order to provide actual decision support in context of a Decision Support System (DSS).

II. METHODOLOGY

There are two major challenges when exploiting CPGs for a DSS in context of diagnostic processes. Firstly, a CPG typically incorporates descriptions of diagnostic processes in form of continuous texts, tables and diagrams (i.e. unstructured data). Consequently, in order to be accessible for a technical system, the knowledge contained therein must be properly formalized. Secondly, based on the formalization, a decision support suitable for the current diagnostic process and the patient on hand has to be provided.

Regarding the first challenge, providing a proper formalization of a diagnostic process, we propose a framework which takes advantage of a dialog between experts of the

medical and the technical domain (cf. Fig. 2): together they develop a formalization based on a UML activity. Alternatively, because of the easy comprehensibility of a UML activity, a medical expert can formalize or modify a CPG by him- or herself. We believe that by the latter, possible barriers of CPG implementation, like the fear of paternalism or regimentation, can be reduced [5]. This is an important aspect regarding the acceptance and subsequently the actual use of a DSS in daily medical practice in order to bridge the gap between theoretical knowledge and practical solutions (cf. Fig. 1).

To address the second challenge, i.e. providing a suitable decision support with respect to the current boundary conditions (diagnostic process, patient, etc.), we propose a BN (cf. Fig. 2) which models the probability of the presence of a disease under consideration. Based on previously obtained examination values, next examination values are proposed regarding both: the reduction of uncertainty about the presence of the disease as well as the sequence of actions proposed by the guideline. This allows for natural deviations from a CPG, which are necessary for its proper implementation in daily practice [4].

III. RELATED WORK

The use of UML activities as starting point for the development of other models is, e.g., elaborated in [9], [10], [11]. In [12], [13] and [14] a transformation of a UML activity into a Petri Net is considered. Thereby the comprehensibility

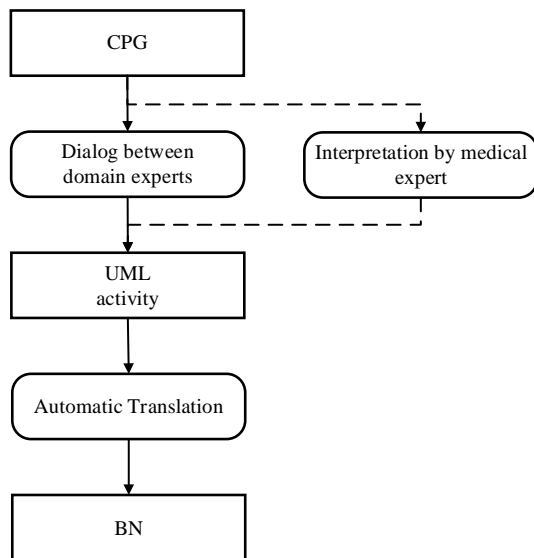


Fig. 2. A CPG can be transferred into a UML activity of the corresponding diagnostic algorithm. Formalization can be carried out by a dialog of experts or by the medical expert him- or herself. The resulting activity serves as an interface to a Bayesian Network (BN) which is used to carry out the actual decision support. In this work, we introduce translation rules, automatically transferring a given UML activity into a BN. Modified from [8].

of UML activities are linked to an abundance of analysis techniques provided by Petri Nets.

In [4], [5] we introduced a semi-automatic translation from UML activities of the diagnostic processes of Myelodysplastic Syndromes (MDS) and Chronic Myeloid Leukemia (CML) into a BN. In this paper, we extend our approach and focus on an automatic translation of UML activities into a BN on the application example of MCL, as well as MM. Diagnoses of these cancerous diseases are complex and have therefore, to the best of our knowledge, not been formalized and made accessible with the elaborated level of detail for an interactive DSS so far.

IV. FUNDAMENTALS ON UML ACTIVITIES

The use of UML is accepted in academia and industry worldwide [15], [16]. UML activities are among the behavioral diagrams – i.e. they are suitable to represent the proceeding of a particular algorithm or process [17]. They have been chosen as a basis for formalization because they are easy to understand for medical and technical experts [5]. Thereby, the comprehensibility of the workflow representation is a necessary precondition for the expert-based dialog and for the interpretation by a medical expert by him- or herself (cf. Sec. II).

A UML activity consists of different notation elements (shapes) which can be connected by the use of directed edges. I.e., this allows for the modeling of sequences, concurrency, decision and iteration [15]. Fig. 3 shows prototypical interconnections which appear in the CPGs of MM and MCL as well as other diseases with complex diagnostic processes [5]. Thereby a black dot symbolizes the start of an activity (initial node) whereas the double circle represents the end of an activity (activity final). Rounded rectangles represent the individual actions of the workflow

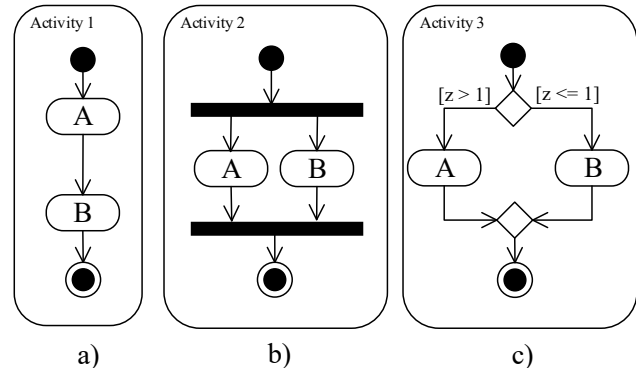


Fig. 3. Three typical routings of the CPG models of MM and MCL. In Subfigure a) the actions A and B are carried out sequentially whereas in Subfig. b) these actions can be performed in any arbitrary order. Subfig. c) depicts a decision depending on an expression (guard) which, if evaluated to true, enables the transition to the target of the directed edge (next action). I.e. if the value of a variable z is greater than 1, action A is performed, if not, action B.

whereas possible flows are given by arcs. The black bars used in Subfig. 3b) are concurrency nodes with different characteristics: the first black bar forks possible flows (fork node) whereas the subsequent black bar joins them (join node). In Subfig. 3c) the diamonds represent branch nodes which can be further divided into decision and merge nodes. Furthermore, Fig. 4 depicts the flow of objects, which are represented by rectangles. The abbreviated form of an object flow is a pin, cf. Activity 5 in Subfig. 4b).

V. MODELING OF THE DIAGNOSTIC PROCESS

Fig. 5 shows the diagnostic process for MM. Due to the size of the model, the figure is simplified and important parts of the UML activity are emphasized by magnifiers.

Subfig. 5a) shows a sequence of actions that can be carried out right after the activity starts. The first action, Suspicion of MM, is an important precondition since the CPGs considered are diseases specific. I.e. a CPG of cancerous diseases generally starts with a suspicion for the disease under consideration [5]. The second action is H&P (History and Physical Examination). During H&P the medical practitioner, e.g. asks a patient if he feels bone pain. Therefore, the object Bone Pain is a result of the action H&P – whereby a pin notation [18] is used to specify the output of an action. Thereby an action can have several output or input pins. E.g., the third action, Blood Count, generates (just as H&P) several specific outputs.

Subfig. 5b) depicts the action Verify MM. This action includes an evaluation: The practitioner must decide whether or not the examination results confirm the disease. However, there are no fixed rules and the final decision lies with the medical expert [5]. Therefore, the decision as to whether a disease is present or not is not modeled deterministically (i.e., not by fixed rules). Instead, the medical practitioner is provided with the probability of the disease under consideration being present, given the present values. To facilitate the evaluation of this probability, the input pins of the action are extended by tuples consisting of two numbers. Each of them specifies how likely it is to observe an untypical examination value for patients having or not having the disease, respectively.

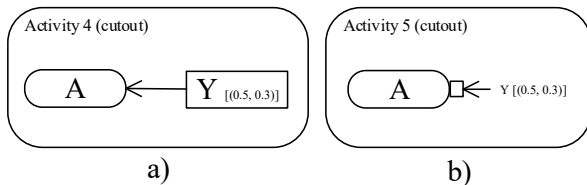


Fig. 4. Subfig. a) depicts an object flow that occurs before action A is carried out. The object itself is represented by a rectangle. Subfig. b) shows an alternative representation of the same fact: A small rectangle, a so-called pin, is drawn at the edge of the action icon.

Subfig. 5c) shows another action: Calculate MM Phase. In contrast to the action Verify MM, this action involves

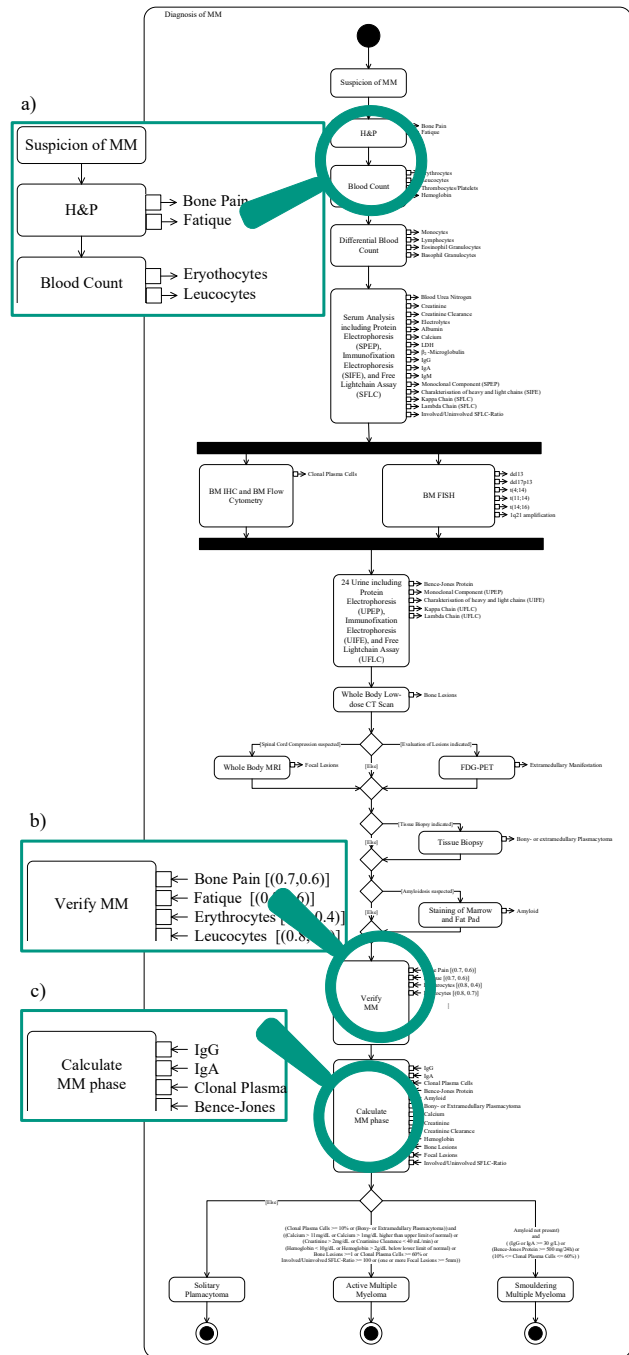


Fig. 5. This figure shows the diagnosis process of MM formalized as a UML activity. Essential parts of the diagram are emphasized (magnifiers). First, in Subfig. a), there is a sequential order of actions. E.g. H&P (History and Physical Examination) and Blood Count. An object flow is indicated by the output pins – e.g. Bone Pain is an object which results from H&P. In Subfig. b) input pins are shown, since the corresponding action (here: Verify MM) depends on the objects as input parameters – same applies to Subfig. c).

a deterministic decision. That means, this decision can be made on basis of fixed rules [5]. E.g., the type of MM can be derived from formulas. For simplification, the diagnostic process of MCL is not shown because it is 2-3 times the size of the depicted diagram. Nevertheless, the same assumptions apply to it.

VI. FUNDAMENTALS ON BAYESIAN NETWORKS

Medicine is one of many well researched fields for the application of BNs [19]. In terms of medical diagnostic processes, they are suitable because they can handle uncertainty and incomplete data. In particular they allow for calculating the probability of the presence of a diseases on basis of fragmentary (i.e. incomplete) examinations entailing uncertainty. A BN is a probabilistic graphical model. Therefore a BN over random variables $X^{0:N} = X^0, \dots, X^N$ is given by a tuple

$$B = (P, G).$$

Whereby P is a joint probability distribution

$$P(X^{0:N}) = \prod_{n=0}^N P(X^n | \text{Pa}(X^n)),$$

and

$$G = (V, E),$$

corresponds to a directed, acyclic graph (DAG) [20], [21], [22]. DAG G , also known as the structure of a BN, is used to define dependencies between random variables $X^{0:N}$. Thereby, the vertex set V represents the set of random variables. While a directed edge $V_k \rightarrow V_m$ within the set of edges E represents a direct dependency between two variables, the independence of two variables is symbolized by a missing edge [8].

Whereas BNs provide a good tradeoff between complexity and expressiveness to a knowledge engineer, the formal representation of a BN is not well suited for a joint dialog of a technical and a medical domain expert. Furthermore, the construction and modification of a BN by the medical expert him- or herself seems not to be feasible. Consequently, we propose rules, transferring a comprehensible UML activity into a BN.

VII. MODELING APPROACH

Given a UML activity represented as a graph $\mathcal{U} = (N, F)$, the set of nodes N can be divided into different sets given by [5]:

- \mathcal{A} : Set of action nodes,
- \mathcal{B} : Set of decision and merge nodes (branch nodes),
- \mathcal{C} : Set of fork and join nodes (concurrency nodes),
- \mathcal{I}, ω : Set of initial node, set of final nodes,
- \mathcal{O} : Set of object nodes.

Thereby, a node that is part of one of the sets $\mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{E}$ is called a control node. The set of object nodes \mathcal{O} is given by

the set of data pins. Furthermore, the set F of activity edges is given by

- $\mathcal{F}_{\text{control}}$: Control flow, i.e. activity edges linking control nodes and actions, as well as edges between themselves.
- $\mathcal{F}_{\text{object}}$: Object flow, i.e. activity edges linking object nodes and actions or linking object nodes and control nodes.

A. Structure

Formally, the translation $[[\mathcal{U}]]$ of a UML activity \mathcal{U} to the network structure G of the BN B is given by:

$$[[\mathcal{U}]] = (V, E),$$

using

$$V = v_r \cup \{e_i \mid (e_i, e_j) \in \mathcal{F}_{\text{object}}, e_i \in \mathcal{O}, e_j \in \mathcal{A} : e_j.\text{contains}(\text{"Verify"})\}, \quad (1)$$

$$E = \{(v_r, e_i) \mid (e_i, e_j) \in \mathcal{F}_{\text{object}}, e_i \in \mathcal{O}, e_j \in \mathcal{A} : e_j.\text{contains}(\text{"Verify"})\}. \quad (2)$$

In (1) a root node v_r , representing the presence of a disease is added to the set of vertices V of DAG G . Furthermore, input pins of actions which include a kind of assessment, represented by the keyword Verify, are added to the set of vertices as well. Please note, that an object with an identical name (e.g. as input of different actions) is correctly added to the set only once, due to the fact that $a \cup \{a, b\} = \{a, b\}$. Same applies to actions, therefore we assume that actions of the UML activity have a unique name.

Since a specific disease causes typical examination results, there is a directed edge from the root node to the relevant examination values (2). This causal interpretation of an edge normally reflects the expert's understanding and is therefore advantageous for an expert-based parametrization.

B. Parametrization

Besides the structure, the parameters of the BN have to be specified. Let X^n be the random variables associated to nodes $v_n \in V$. We define

$$\text{Val}(X^n) = \{0, 1\},$$

using $n = 0, \dots, N$ with $N = |V|$. Thereby $\text{Val}(\cdot)$ denotes the set of values an associated random variable can take. I.e., for all variables a Bernoulli distribution is used and the associated nodes v_n are binary-valued. The a-priori probability of a disease being present is set to $\pi = (0.5, 0.5)$ to represent a maximum of uncertainty of the diagnosis.

Let $\mathcal{O}_{\text{name}}$ be the set of unique input pin names associated with an action containing the keyword Verify. We define a surjective function $g : \mathcal{O} \rightarrow \mathcal{O}_{\text{name}}$. This function returns the name of a given object node o . We define another, bijective function $h : \mathcal{O}_{\text{name}} \rightarrow \{1, \dots, N\}$, with $N = |\mathcal{O}_{\text{name}}|$. By function h , a unique natural number is assigned to each element of $\mathcal{O}_{\text{name}}$.

To specify the underlying conditional probability distribution for each examination, we define a matrix \mathbf{B}^k for each element $o_{\text{name}} \in \mathcal{O}_{\text{name}}$:

$$\exists! \mathbf{B}^k \forall o_{\text{name}} \in \mathcal{O}_{\text{name}} : k = h(o_{\text{name}}).$$

The elements $b_{i,j+1}^k$ of each matrix \mathbf{B}^k are given by

$$b_{i+1,j+1}^k = \begin{cases} P(X^k = 1|X^0 = i), & \text{if } j = 1 \\ 1 - P(X^k = 1|X^0 = i), & \text{otherwise,} \end{cases}$$

where X^k with $k = 1, \dots, N$ are random variables associated with examination value k , and X^0 is the random variable representing the (non) presence of a disease. The indices $i + 1, j + 1$ of the matrix \mathbf{B}^k are given such that $i \in \text{Val}(X^0) = \{0, 1\}$, and $j \in \text{Val}(X^k) = \{0, 1\}$. The conditional probabilities $P(X^k = j|X^0 = i)$ are derived from the state of the object nodes. Each node $o \in \mathcal{O}$ is associated with a tuple of probability values $(o.\text{prob}_m)_{m=1}^2$ and

$$P(X^k = 1|X^0 = i) = o.\text{prob}_{i+1}$$

$$\forall(o, a) \in \mathcal{F}_{\text{object}} \text{ with } a \in \mathcal{A}, o \in \mathcal{O}, (h \circ g)(o) = k.$$

Given the presented annotations and translation rules, a UML activity can be automatically translated into a BN.

C. Decision Support

In [4] we introduced a decision support approach considering the current boundary conditions (diagnostic process, patient, etc.). On the basis of previously obtained examination values, next examination values are proposed, both in terms of reducing uncertainty about the presence of the disease and in terms of the sequence of actions proposed in the guideline. This allows for natural deviations from CPG recommendations, which is a necessary precondition for its proper implementation in daily practice [4]. For example, if the guideline states that a CT scan must be performed, but the patient is pregnant, the doctor must act accordingly, even if this is not explicitly stated in the guideline.

For the reduction of uncertainty about the (non-) presence of a disease under consideration, the diagnostic node plays an important role: the more the probability mass scatters over the states ‘‘present’’ or ‘‘not present’’ of the corresponding random variable, the more uncertainty exists about the presence or absence of the disease.

A measure of the uncertainty of a random variable is given by its entropy [5], [23]. The more the probability mass scatters over the states of the random variable the higher the entropy is. Let X^0 be a discrete random variable with n states x_1^0, \dots, x_n^0 and $P(X^0)$ the probability distribution. The entropy (in bit) is then given by [23]:

$$H(X^0) = - \sum_{i=1}^n P(x_i^0) \log_2 P(x_i^0), \quad (3)$$

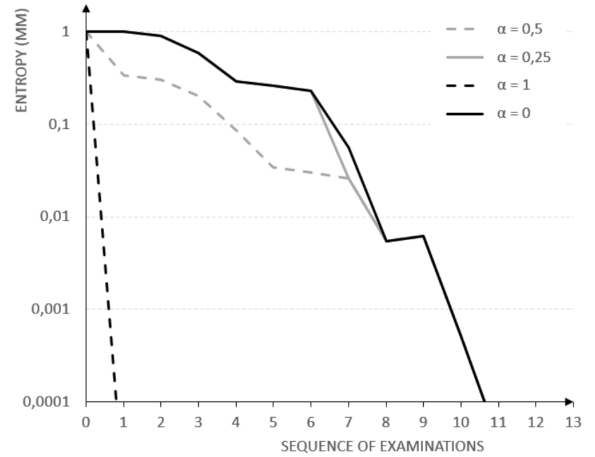


Fig. 6. Figure shows the entropy of the diagnostic node of MM versus the order examinations for different values of α . For simplification, different values are grouped.

where $H(X^0) \in [0, \log_2(n)]$. Furthermore, the mutual information

$$I(X^0, X^1) = H(X^0) - H(X^0|X^1) \quad (4)$$

represents the reduction in the uncertainty of X^0 due to the knowledge of X^1 . If there are several possible examinations X^i with $i = 1 \dots N$ that can be performed, one would prefer the one with the highest mutual information (i.e. highest reduction of uncertainty about X^0). Given a BN with a node representing the random variable X^0 , the entropy $H(X^0)$ can be calculated by formula (3).

The conditional entropy $H(X^0|X^i)$ is defined as [23]:

$$H(X^0|X^i) := \sum_{j=1}^n P(X^i = x_j^i) H(X^0|X^i = x_j^i), \quad (5)$$

where $j = 1 \dots n$ are the states of random variable X^i . In daily practice, the choice of an examination based solely on Eq. (4) is not possible. E.g., using that would mean that an examination that is very specific for a disease would be suggested first, regardless of whether this examination is highly invasive or expensive which may outweigh its diagnostic benefit [5]. To take the costs (e.g. invasiveness) into account, the order of the examinations listed in the recommendations of the CPG is used: For example, at the beginning of the diagnostic process, the proposed exams are typically less invasive than at the end of the diagnostic process, when the physician is more confident that a disease is present or not.

Therefore, in [5] we propose a weighted reduction of uncertainty which integrates cost-benefit considerations during the diagnostic process:

$$\text{Recommendation}_i = (1 - \alpha)(1 - d_i/m) + \alpha I(X^0, X^i), \quad (6)$$

where d_i is the depth of examination in the CPG associated with X^i , m is the overall depth of the CPG, and α is

Table I
PATIENT WITH SUSPECTED MM ($\alpha = 0$).

No.	Depth	Examination	Entropy
0	0	-	1
1	1	History & Physical Positive	0.995727452
2	2	CBC, differential, platelet count Positive	0.913661787
3	3	Serum BUN Positive	0.585156990
4	3	Serum LDH, beta-2 microglobulin Positive	0.293209343
5	3	Serum quant., SPEP, SIFE Positive	0.259543321
6	3	Serum free light chain (FLC) assay Positive	0.229109565
7	4	Plasma cell FISH Positive	0.055779575
8	4	Plasma cell proliferation Positive	0.005413689
9	5	24-h urine protein, UPEP, UIFE Positive	0.006144079
10	6	Whole body low-dose CT scan Positive	0.000507863
11	7	Whole body MRI/PET-CT Positive	0.000037310
12	8	Tissue biopsy Positive	0
13	9	Staining of marrow for amyloid Negative	0

the weight factor. For $\alpha = 0$ the recommendations follow exactly the order given by the CPG (considering only cost) – for $\alpha = 1$ only the reduction of uncertainty is considered (considering only diagnostic benefit).

VIII. RESULTS

For verification, typical patients regarding cancer statistics of MM and MCL are generated [24]. Fig. 6 shows the progress of the entropy for the diagnostic node of MM versus the sequence of examinations. Please note that for simplification the examination values are grouped. While more examination values are assessed, the entropy drops slowly by following the CPG recommendation (solid black line). At the last examinations, the entropy reaches the value of 0. That is, because a tissue biopsy, which is one of the late (and highly invasive) examinations, is regarded as confirming the disease for sure (cf. Tab. I).

The black dotted line depicts the progress of entropy while the DSS is maximizing the reduction of uncertainty in each step (i.e. $\alpha = 1$). Consequently, invasive and risky examinations are proposed first, because they can reduce uncertainty the most [5]. In our case, tissue biopsy is proposed first, which reduces the entropy to 0 at once. Instead of considering either diagnostic benefit ($\alpha = 1$) or costs ($\alpha = 0$), the DSS can be parametrized with $\alpha \in (0, 1)$. This allows for balancing both aspects. E.g. for $\alpha = 0.5$, Plasma Cell Proliferation is assessed first, allowing for a moderate reduction of uncertainty with respect to invasiveness (dashed gray line). At the same time, an early and natural deviation from the CPG is possible. For $\alpha = 0.25$, the recommendations are stronger bound to the CPG, but also allowing for some deviations in step 7 and 8. This is because the assessment of Plasma Cells in step 7 and 8 is a strong indicator for the (non) presence of MM.

Figure 7 shows the progress of entropy versus examination values for the diagnosis of MCL. Please note that for simplification only 19 characteristic examination values in

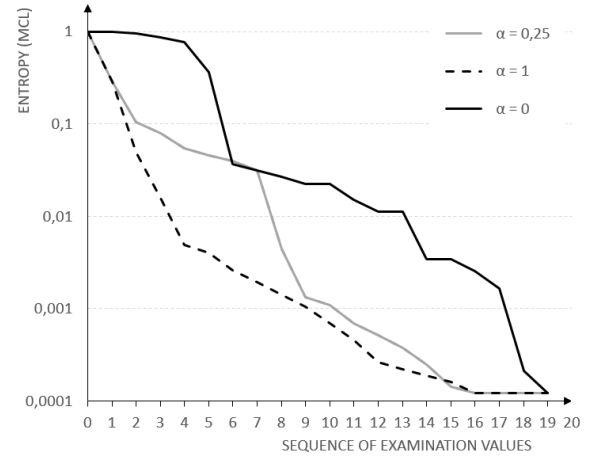


Fig. 7. Entropy of the diagnostic node of MCL versus the order of examination values. For simplification only 19 characteristic examination values in two consecutive depths are shown.

two consecutive depths are shown (cf. Tab. II). The solid black lines depicts how the entropy decreases while more examination values are assessed in the order proposed by the CPG ($\alpha = 0$). The examinations are listed in Tab. II. The dashed black line shows the progress of entropy while in each step only the reduction of uncertainty is considered ($\alpha = 1$). Consequently, highly specific examinations like immunohistochemical staining of cyclin D1 antibodies are proposed first. Using $\alpha = 0.25$, a natural deviation from the CPG model is possible. E.g. the examination of cyclin D1 is proposed in an earlier step while other examinations are proposed later according to the CPG.

Table II
PATIENT WITH SUSPECTED MCL ($\alpha = 0$).

No.	Depth	Examination value	Entropy
0	0	-	1
1	5	Hematop. Review of Lymphnode Positive	0.985228136
2	5	Immunphenotyping IHC: CD20 Positive	0.962900415
3	5	Immunphenotyping IHC: CD5 Positive	0.862371528
4	5	Immunphenotyping IHC: CD10 Negative	0.763337690
5	5	Immunphenotyping IHC: CD23 Negative	0.364989518
6	5	Immunphenotyping IHC: CycD1 Positive	0.037070557
7	5	Immunphenotyping IHC: BCL6 Negative	0.031222027
8	6	Immunphenotyping FLW: CD20 Positive	0.026975342
9	6	Immunphenotyping FLW: CD19 Positive	0.022288263
10	6	Immunphenotyping FLW: CD200 Positive	0.022288263
11	6	Immunphenotyping FLW: CD5 Positive	0.014994722
12	6	Immunphenotyping FLW: CD10 Negative	0.011189228
13	6	Immunphenotyping FLW: kappa/lambda Positive	0.011189228
14	6	Immunphenotyping FLW: CD23 Negative	0.003442625
15	6	t(11;14) Negative	0.003442625
16	6	t(14;18) Negative	0.002548508
17	6	Immunohistochem.: LEF1 Negative	0.001640160
18	6	Immunohistochem.: SOX Positive	0.000211186
19	6	Immunohistochem.: IGHV-Mut Negative	0.000122914

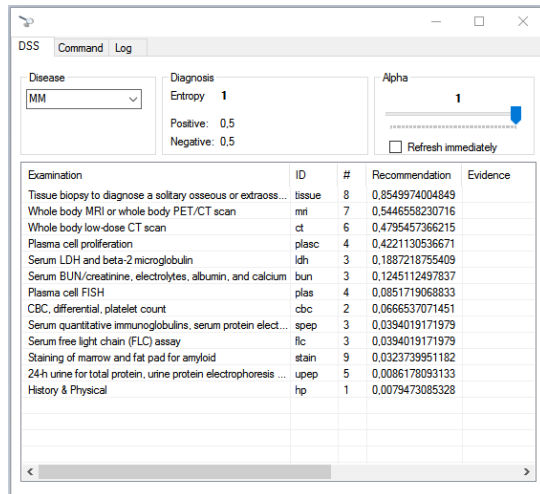


Fig. 8. User interface of the DSS.

Fig. 8 shows an exemplary graphical user interface of the proposed DSS. In the upper part of the application, the entropy and the probability of the (non) presence of the disease under consideration is shown. The parameter α can be adjusted using a slider – the proposed examination value are sorted accordingly. Regarding the acceptance of this concept by the end-user, further investigations will be carried using a currently developed software system for the treatment of cancerous diseases.

IX. SUMMARY AND CONCLUSION

In this work, we presented a framework for a decision support system (DSS) for the diagnostics of the cancerous diseases Mantle Cell Lymphoma (MCL) and Multiple Myeloma (MM). Given previous examination results, our DSS is able to propose next examinations tailored to the current boundary conditions (diagnostic process, patient, etc.). The medical practitioner can choose between examinations to balance invasiveness as well as the reduction of uncertainty about the (non) presence of a disease.

Since Clinical Practice Guidelines (CPGs) contain condensed and consolidated medical knowledge about the diagnostics of the complex cancerous diseases of interest, our approach integrates this knowledge by formalizing it using UML activity diagrams. In a first step, the UML activities serve as a basis for the dialogue between medical and technical domain experts being easily understandable for both. In a second step, the UML activities are automatically translated into Bayesian Networks which are able to model the probability of the (non) presence of the respective cancerous disease. We showed the translation rules and validated the decision support based on the generated Bayesian Networks.

Given the comprehensibility and easy adaptability of UML activities, our framework could serve as a starting point for formalizing not only the diagnostics of MCL and

MM, but also other complex diseases. With our automatic translation into probabilistic models, decision support during diagnostics of other cancerous diseases is feasible.

REFERENCES

- [1] National Center for Biotechnology Information, U.S. National Library of Medicine, "Pubmed," March 2018, Accessed: 2018/03/12. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed>
- [2] C. D. Mulrow, "Rationale for systematic reviews." *BMJ: British Medical Journal*, vol. 309, no. 6954, p. 597, 1994.
- [3] A. d. W. M. F. AWMF, "Das Leitlinien-Manual von AWMF und ZQ," *ZaeFQ: Zentrum ärztliche Qualitätssicherung*, vol. 95 (Suppl I), 2001.
- [4] P. Philipp, Y. Fischer, D. Hempel, and J. Beyerer, "Framework for an interactive assistance in diagnostic processes based on probabilistic modeling of clinical practice guidelines," in *Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology*, Elsevier, pp. 371–390, 2016.
- [5] P. Philipp, Y. Fischer, D. Hempel, and J. Beyerer, "Modeling of clinical practice guidelines for interactive assistance in diagnostic processes," in *HIMS 2015, International Conference on Health Informatics and Medical Systems*, July 27–30, Las Vegas, USA, 2015, pp. 3–9.
- [6] E. Field and K. Lohr, *Guidelines for Clinical Practice: From Development to Use*. National Academies Press, 1992.
- [7] E. Steinberg, S. Greenfield, M. Mancher, et al., *Clinical Practice Guidelines We Can Trust*. National Academies Press, 2011.
- [8] P. Philipp, Y. Fischer, and J. Beyerer, "Expert-based probabilistic modeling of workflows in context of surgical interventions," *CogSima 2017, IEEE Conference on Cognitive and Computational Aspects of Situation Management*, 2017.
- [9] Z. Han, L. Zhang, and J. Ling, "Transformation of uml activity diagram to yawl," *Enterprise Interoperability*, Springer, pp. 289–299, 2010.
- [10] W. van der Aalst and K. M. Van Hee, *Workflow management: models, methods, and systems*. MIT press, 2000.
- [11] A. Hofstede, N. Russell, W. Aalst, and M. Adams, *Modern Business Process Automation: YAWL and its Support Environment*. Springer, 2010.
- [12] H. Störrle, "Semantics of control-flow in UML 2.0 activities," in *Visual Languages and Human Centric Computing, 2004 IEEE Symposium on Visual Languages and Human-Centric Computing*, 2004, pp. 235–242.
- [13] H. Störrle, "Semantics and verification of data flow in UML 2.0 activities," *Electronic Notes in Theoretical Computer Science*, vol. 127, no. 4, pp. 35–52, 2005.
- [14] B. Agarwal, "Transformation of uml activity diagrams into petri nets for verification purposes," *International Journal Of Engineering And Computer Science*, vol. 2, no. 3, pp. 798–805, 2013.
- [15] OMG - Object Management Group. (2014.08.01) *Unified Modeling Language Superstructure Version 2.4.1*. [Online]. Available: <http://www.omg.org/spec/UML/2.4.1/Superstructure/PDF/>
- [16] C. Kecher, *UML 2: das umfassende Handbuch*. Bonn: Galileo Press, 2011.
- [17] C. Rupp, S. Queins, et al., *UML 2 glasklar*. Carl Hanser Verlag GmbH Co KG, 2012.
- [18] H. Störrle, *UML 2 für Studenten*. Pearson Studium, 2005, vol. 320.
- [19] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*. cRc Press, 2003.
- [20] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
- [21] R. E. Neapolitan, *Probabilistic reasoning in expert systems: theory and algorithms*. Wiley, 189.
- [22] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [23] T. M. Cover and J. A. Thomas, *Elements of information theory*. Hoboken, New Jersey: John Wiley & Sons, 2006.
- [24] Robert Koch-Institut, "Zentrum für Krebsregisterdaten – Krebs in Deutschland," October 2015, Accessed: 2017/11/23. [Online]. Available: <http://www.krebsdaten.de/Krebs/DE/Content/Publikationen/>