Scheme for Unstructured Knowledge Representation in Medical Expert System for Low Back Pain Management

Debarpita Santra, Sounak Sadhukhan, S. K. Basu, Sayantika Das, Shreya Sinha and Subrata Goswami

Abstract A fundamental requirement to achieve effectiveness in a medical expert system is the proper representation of knowledge about a patient. Knowledge about a patient is stored in the expert system in terms of clinical records which should contain information about multiple visits of the patient. During each visit, several conversations are made between the patient and the consulting physician. These conversations, being unstructured in nature, cannot be stored in the computer using available structured knowledge representation schemes. So, we propose a recursive frame-based structure for representing clinical records. The frames related to a patient collectively form a frame system, where one frame may point to other frames. The proposed representation scheme is complete, consistent, and free from redundancy.

D. Santra · S. Sadhukhan (✉) · S. K. Basu
Department of Computer Science, Banaras Hindu University, Varanasi 221005, India
e-mail: sounaks.cse@gmail.com

D. Santra
e-mail: debarpita.cs@gmail.com

S. K. Basu
e-mail: swapankb@gmail.com

S. Das · S. Sinha · S. Goswami
ESI Institute of Pain Management, Kolkata 700009, India
e-mail: 93sayantika@gmail.com

S. Sinha
e-mail: shreyasinha256@gmail.com

S. Goswami
e-mail: drsgoswami@gmail.com

S. C. Satapathy et al. (eds.), *Smart Intelligent Computing and Applications*,
Smart Innovation, Systems and Technologies 105,
https://doi.org/10.1007/978-981-13-1927-3_4

1 Introduction

Medical diagnosis is made based on clinical tests and clinical examinations by the physician. There has been fast and tremendous growth of knowledge and advances in medical and other associated techniques. It is quite difficult and almost impossible for a treating physician to be at home with all these advances and pass on the benefits to the patients. A medical expert system is defined as a software using artificial intelligence techniques and up-to-date knowledge of a specific medical domain. Advantages of medical expert system can obviously be appreciated at the initial screening stages of diseases like low back pain (LBP).

Compared to any other diseases, LBP is responsible for causing more disability worldwide [1]. Though several methods are used in clinical practice to investigate and diagnose LBP, unfortunately, the exact cause of producing the LBP is not well identified sometimes if the patient is not subjected to elaborate clinical investigations. To cope with this issue, we are planning to design a reliable and efficient medical expert system to assess and manage LBP.

Like any expert system, the medical expert system for LBP should consist of four major modules: user interface, working memory, knowledge base, and inference engine. The patient information that are gathered through the user interface, should be represented and stored in the expert system in such a way that they can be interpreted and retrieved efficiently. This kind of information is highly unstructured in nature. So, capturing the semantics as well as the complex relationships among the data items is a challenging task. In this paper, we have proposed a frame-based representation scheme for the unstructured LBP patient information.

The rest of the paper is organized as: Sect. 2 gives an overview of related works. In Sect. 3, we discuss our proposed representation scheme and illustrate it using a simple example from the domain of LBP. Finally, Sect. 4 concludes the paper.

2 Related Study

Efforts have been made since 1970s to build up medical expert systems. MYCIN [2] which is the pioneering medical expert system for blood infections, aids physicians to recommend antibacterial medicines for the disease. Other renowned medical expert systems are CASNET [3] for glaucoma, INTERNIST [4] applicable for the domain of internal medicines, ONCOCIN [5] for cancer, etc. Some expert systems [6–8] are concerned with LBP. In case of MYCIN, a context tree is formed to hold the previous medical histories of patients and each instance of the tree basically represents individual clinical event [9]. This representation is not effective when a clinical record is visualized as a collection of subsequent events or visits. CASNET is a causal model for better understanding of disease processes. The findings about a patient are represented as a collection of individual nodes in the plane of observation, but the complex relationships among patient findings are not well captured. INTERNIST proposes a hierarchical organization for different disease categories that are associated with different organ systems in a human body. Each of the disease categories is further subdivided until leaf nodes are reached. A leaf node represents an individual disease. ONCOCIN proposes a temporal network to obtain the key information about the past chemotherapy cycle or previous visits of a patient. But this structure is also not efficient to capture the interrelatedness of patient data that appear in the domain of LBP [10–13] did not propose any unstructured knowledge representation scheme for capturing complex relationships among data items in individual patient records.

3 Proposed Scheme

A patient record should comprise all the information that are collected during multiple visits of the patient. During these visits, lots of conversations take place between the patient and the consulting physician. These conversations are fully unstructured in nature and cannot be represented in computer using the available representation schemes for structured data. Here, we have proposed a suitable representation scheme for patient record to capture and analyze important relationships that exist among data items. We represent each visit as an individual frame that can capture different types of unstructured data with the help of attributes (slots) and values (fillers) associated with the visit. A slot may hold multiple values or no value at all. A frame allows recursive embedding, i.e., it can have other frames as slots. We visualize a patient record as a sequential collection of information related to visits. So, a patient record can be represented as a frame system which is a collection of connected frames. This representation scheme would facilitate achieving computational effectiveness of the medical expert system for LBP.

A frame associated with the tth visit ($1 \le t \le n$) is represented as V_t. So, a complete patient record R is a frame system with a frame V_t pointing to another frame V_{t+1}. We now use the concept of a class frame which can have many instances. We assume that there exist subclass frames of a class frame, where each subclass frame contains less or equal number of attributes of the class frame. A subclass frame may also have many instances. We define a class frame V as a collection of L slots ($L > 0$), where L_1 ($L_1 > 0$) slots contain only fixed values, L_2 (=2) slots contain two different frame instances of V, one for previous visit (if any) and another for the next visit (if any). The remaining slots L_3 ($= L - (L_1 + L_2)$) contain frames different from frames of type V. An individual slot from L_3 contains a particular frame F_i ($1 \le i \le L_3$), which is defined as a collection of M slots ($M > 0$), where M_1 ($M_1 > 0$) slots contain only fixed values and M_2 ($= M - M_1$) slots contain another F_j ($j > L_3$) frames. The values of M, M_1, and M_2 as well as the attributes will be different for each F_i and F_j frames. A class frame V may have many subclasses. A subclass frame V^s of V contains slots equal or less than that of V, but will always contain all the L_2 slots. The attributes that will be present in subclass frame V^s at tth visit would be decided dynamically based on the conversation on that visit. So, we say that V_t is an instance of V^s. We provide the generalized definition of the class frame V in Fig. 1.

Class Frame Name: V

1. **Previous Visit**: Frame $V_{t\text{-}1}$, if ($1 \le t \le n$) else empty
2. **Personal Information**: Frame F_1
3. **Chief Complaints**: Frame F_2
4. **Associated Symptoms**: $<t_{(4,1)}, t_{(4,2)}, \ldots, t_{(4,P)}>$ /* This multi-valued slot contains P ($P > 0$) fixed values each with suffix (f, j) where f is the slot no. and j is the index of value ($1 \le j \le P$) */
5. **Medical History**: $<t_{(5,1)}, t_{(5,2)}, \ldots, t_{(5,Q)}>$ /* This multi-valued slot contains Q ($Q > 0$)fixed values each with suffix (f, q) where q is the index of value ($1 \le q \le Q$) */
6. **Family Medical History**: $<t_{(6,1)}, t_{(6,2)}, \ldots, t_{(6,R)}>$ /* This multi-valued slot contains R ($R > 0$) fixed values each with suffix (f, r) where r is the index of value ($1 \le r \le R$) */
7. **General Examination**: Frame F_3
8. **Local Examination**: Frame F_4
9. **Blood Test**: Frame F_5
10. **Special Test**: Frame F_6
11. **Imaging**: Frame F_7
12. **Diagnosis**: Frame F_8
13. **Next Visit**: Frame V_{t+1}, if ($1 \le t \le n\text{-}1$) else empty

Fig. 1 Generalized structure of class frame V for the domain of LBP

For our current clinical consideration about the structure of V, $L = 13$, where $L_1 = 3$ and $L_3 = 8$. The attributes of frames $F_1, F_2, \ldots, F_8$ (Fig. 1) are different from the medical perspective. For example, while F_1 captures personal information like name, age, sex, occupation, etc., of an LBP patient as fixed values, the frame F_2 captures the chief (major) complaints of the patient like the duration and type of pain, radiation (if any), aggravating factors, relieving factors, diurnal variation, etc. Each of the complaints is captured using different slots of the frame. While some of the slots may contain fixed values (if not empty), other slots indicating radiating pain, aggravating factors, relieving factors, and diurnal variation point to other frames F_9 through F_{12}, respectively. These frames may also have recursive embedding. We illustrate our proposed scheme using an example as shown in Fig. 2.

For the sake of illustration of our proposed scheme using Fig. 2, we have considered a patient named XX who is a jute mill worker. The patient has visited the doctor for the first time on December 26, 2016 with complaints of LBP. The patient is a female of age 56 years. She feels continuous pain below L5 of the spine and in buttocks since last 2 months. Most of the time she becomes depressed with her work and her discomfort level is 6 on the Visual Analog Scale (VAS). So, the severity of her pain can be interpreted as “high”. The pain starts slowly and is aggravated when she bends or lifts some moderately heavy object. Also, her pain is aggravated when she is sitting or standing for long time. Her pain is aggravated in morning and night also. Her pain radiates to the right lower leg at every night. Her sleep is disturbed and bowel/bladder habit is normal. Associated with this, she also has symptoms of headache and nausea. She has allergy and suffers from diabetes. Her mother had died of cancer recently and her father is diabetic.

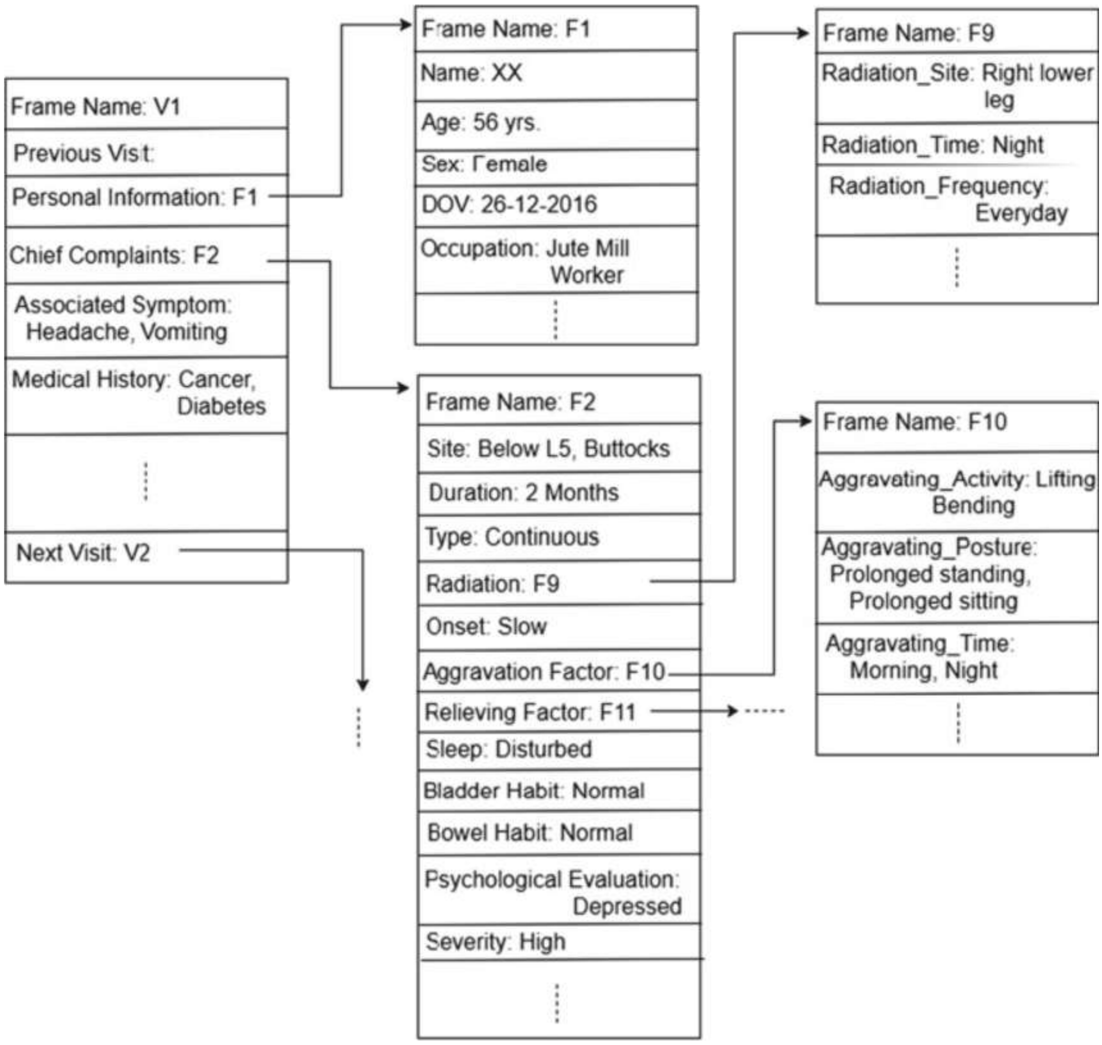

Fig. 2 Representation of our example using the proposed scheme

3.1 *Formal Representation of a Patient Record*

We have formally represented a record of a patient P using six tuples $<N_p, T_p, R_p, P_p, C_p, D_p>$, where

- N_p is a set of variables and $N_p = \boldsymbol{V} \cup \boldsymbol{F} \cup \boldsymbol{A}$, where $\boldsymbol{V} = \{V_t \mid V_t$ is a frame variable corresponding to the frame V_t $(1 \le t \le n)\}$, $\boldsymbol{F} = \{F_{(t,m)} \mid F_{(t,m)}$ is a frame variable corresponding to the frame F_m at tth visit and total no. of F_m variables is greater than or equal to $L_3\}$, $\boldsymbol{A}$ is the set of temporary variables required for ease of internal computations.
- T_p is a set of fixed values in the frame system, where $T_p = T_{\boldsymbol{V}} \cup T_{\boldsymbol{F}}$, where $T_{\boldsymbol{V}}=\{\boldsymbol{t}_{(a,b),t} \mid \boldsymbol{t}_{(a,b),t}$ is the bth $(b>0)$ fixed value of the slot a $(1 \le a \le L_1)$ for frame $V_t\}$, $\boldsymbol{T_F} = \{\boldsymbol{t}_{k,(t,i)} \mid \boldsymbol{t}_{k,(t,i)}$ $(1 \le k \le K)$, K being the total number of fixed values that frame F_i can contain at tth visit$\}$
- R_p is called the starting point where $R_p \in \boldsymbol{V}$

– P_p is a set of replacement rules for retrieving the record of a patient in terms of fixed values finally. A replacement rule is of the form $\alpha \rightarrow \beta$, where $\alpha \in N_p$ and β represents strings on $(N_p \cup T_p)^*$. The set of replacement rules is given below:

(i) $R_p \rightarrow V_t$
(ii) $V_t \rightarrow V_{t-1}\, A_1 \mid A_1$/* $A_1 \in \boldsymbol{A}$ */
(iii) $A_1 \rightarrow (\wedge\, F_{(t,i1)})\, (\wedge\, \boldsymbol{t}_{(a,b),t})\, (\wedge\, F_{(t,i2)}) \mid (\wedge\, \boldsymbol{t}_{(a,b),t}) \mid F_{(t,1)}$
/* (for all a, b) and with $1 \le i1 \le 2$ and $i2 > i1$, $(i1 + i2) = L_3$, and ($\wedge$ $F_{(t,i)}$), ($\wedge$ $\boldsymbol{t}_{(a,b),t}$) are concatenation of subsequent frames and concatenation of subsequent fixed values, respectively */
(iv) $F_{(t,i)} \rightarrow TA_2T \mid TA_2 \mid A_2\, T \mid T \mid \lambda$
/* λ means Null value and $i1$, $i2$ from rule (iii) and $i3$ from rule (v) will be replaced by i. Here, T, $A_2 \in \boldsymbol{A}$ */
(v) $A_2 \rightarrow F_{(t,i3)}\, A_2 \mid \lambda$
/* $i3$ represents the subsequent indices of pointed frames of $F_{(t,i)}$ and is different from $i1$ and $i2$*/
(vi) $T \rightarrow \boldsymbol{t}_{k,(t,i)}\mathrm{T} \mid \boldsymbol{t}_{k,(t,i)}$

– C_p is the set of constraints imposed on invoking of the replacement rules. Constraints define on which conditions or contexts a replacement rule should be invoked. Constraints also define when to terminate in case of recursive calls. Also, constraints impose restrictions on sequence of invocation of replacement rules to derive a final string of fixed values.
– D_p is the set of dependencies among a set of frame variables and a set of fixed values. An element (either variable or fixed value) may depend on another element or set of elements to aid in deducing the final string that represents a partial or complete patient record. For example, the value of the frame slot "Diagnosis" in V_t would depend on the values of all previous frames and fixed values in that frame.

3.2 Reasoning Mechanism

Our proposed representation scheme does not allow random or direct access of a fixed value of a frame. That is, we have to follow a systematic pathway to get the information connected with the tth visit. After all the internal processing of frame V_1 is complete, we would proceed to the next visit (frame V_2) and in this way, we will step forward sequentially until we reach frame V_t. We can construct a derivation tree using the replacement rules we have provided in Sect. 3.1. The root of the tree is R_p. Every node v of the tree is either a variable or a fixed value or λ. If we have a replacement rule such as $X \rightarrow X_1\, X_2\, X_3$, where all three at R.H.S. of the rule are variables, we say that X will be placed at a parent node v in the tree and X_1, X_2, X_3 will be placed at the child nodes v_1, v_2, v_3 of v, respectively. Now, suppose we have another rule $X_1 \rightarrow Y_1\, Y_2$, where Y_1 and Y_2 are fixed values. So, node v_1 will have two children $v_{1.1}$ and $v_{1.2}$ which are leaves. We also consider another replacement rule $X_2 \rightarrow Y_3\ Y_4\ Y_5$, where in R.H.S., first two are variables, and the last one is a fixed value and so on. The access/reasoning sequence will be X- X_1- Y_1- Y_2- X_2- Y_3- Y_4- Y_5- X_3… So, always the leftmost variable present on the R.H.S. of a replacement rule $\boldsymbol{H}$ will be explored first until it reaches the leaf nodes, then the next variable of $\boldsymbol{H}$ will be explored. This process will continue until we visit all the leaves of the tree. The patient record that is the yield of the derivation tree is the concatenation of the leaf nodes that appear in the order from left to right with no repetition. If we want to retrieve the information (INFO) corresponding to a slot Slot$_{k,(t,i)}$ in an individual frame $F_{(t,i)}$, we simply use syntax <Slot$_{k,(t,i)}$.INFO$_{k,(t,i)}$>. Let us illustrate the reasoning mechanism with our example.

R$_p$ => V_1 // using rule (i): $R_p \rightarrow V_t$

=> A_1 // using rule (ii): $V_t \rightarrow A_1$

=> $F_{(1,1)}\ F_{(1,2)}$<Associated Symptom.(Headache, Vomiting)>**…**

// using rule (iii): $A1 \rightarrow (\wedge F_{(t,i1)})(\wedge \boldsymbol{t}_{(a,b),t})(\wedge F_{(t,i2)})$

=> $T\ F_{(1,2)}$ < Associated Symptom.(Headache, Vomiting)>**…**

// using rule (iv): $F_{(t,i)} \rightarrow T$

=> <Name.XX><Age.56 yrs><Sex.Female>**…**<DOV.26-12-2016>**…** <Occupation.Jute Mill Worker>**…** F$_{(1,2)}$<Associated Symptom.(Headache, Vomiting)>**…**

// using rule (vi): $T \rightarrow \boldsymbol{t}_{k,(t,i)}T \mid \boldsymbol{t}_{k,(t,i)}$

=> <Name.XX><Age.56 yrs><Sex.Female>**…**<DOV.26-12-2016>**…** <Occupation.Jute Mill Worker>**…** TA_2T <Associated Symptom.(Headache, Vomiting)>**…**

// using rule (iv): $F_{(t,i)} \rightarrow TA_2T$

=> <Name.XX><Age.56 yrs><Sex.Female>**…**<DOV.26-12-2016> **…** <Occupation.Jute Mill Worker>**…**<Site.(Below L5, Buttock)><Duration.2 months><Type.Continuous>F$_{(1,9)}$$A_2T$ <Associated Symptom.(Headache, Vomiting)>**…**

// using rule (vi): $T \rightarrow \boldsymbol{t}_{k,(t,i)}T$ multiple times

=> <Name.XX><Age.56 yrs><Sex.Female>**…**<DOV.26-12-2016>**…** <Occupation.Jute Mill Worker>**…**<Site.(Below L5, Buttock)><Duration.2 months><Type.Continuous> <Radiation_Site.Right lower leg> <Radiation_Time.Night><Radiation_Frequency.Everyday>A_2T <Associated Symptom.(Headache, Vomiting)>**…**

// using rule (iv): $F_{t,i} \rightarrow T$ and rule (vi): $T \rightarrow \boldsymbol{t}_{k,(t,i)}\ \boldsymbol{T}$ multiple times

=> **…….**

=> <Name.XX><Age.56 yrs><Sex.Female>**…**<DOV.26-12-2016>**…** <Occupation.Jute Mill Worker>**…** <Site.(Below L5, Buttock)><Duration.2 months><Type.Continuous><Radiation_Site.Right lower leg><Radiation_Time.Night><Radiation_Frequency.Everyday> <Aggravating_Activity.(Lifting, Bending)> <Aggravating_Posture.(Prolonged Sitting, Prolonged Standing)> <Aggravating_Time.(Morning, Night)>**…**<Associated Symptom.(Headache, Vomiting)>**…**

// using rule (vi): $T \rightarrow \boldsymbol{t}_{k,(t,i)}T \mid \boldsymbol{t}_{k,(t,i)}$ multiple times

The final string is the concatenation of all the fixed values that represents the overall information of the patient during her visit.

3.3 Discussion of Properties

Our proposed scheme has a number of characteristics as follows. It is complete from two aspects. Firstly, we are able to capture the entire record of a patient visitwise, and secondly, we can reach at fixed values starting from R_{p} without entering into infinite loop. The scheme is consistent, as we have defined range of all the variables and also defined what kind of value a variable may be assigned. We also keep provision of storing null values. Every variable is different at every visit and if a variable accepts two different fixed values at two different visits, it will not be a consistency violation, as diagnosis is dependent on the latest value.

Moreover, the structure works in an integrated manner, which is ensured by recursive embedding of frames in an individual frame. Also, there are dependencies among a set of variables and a set of fixed values during a particular visit. This also ensures the integrity of our scheme. The proposed structure does not support redundant information, as we have partitioned our total structure into many frames, where each frame contains a particular type of information. Also, when a particular frame points to other frames, we distinguish the other frames using different indices. So, a frame does not point to itself, which ensures that our design is free from redundancy. For easy access to the partial or complete patient record, we differentiate multiple visits using variable t and use simple syntax <V_t.Record> to grab the information till the tth visit.

4 Conclusion

In this paper, we have indicated how frame-based knowledge representation scheme provides the foundation for describing a patient record, which contains unstructured data items also. To make the representation clinically intuitive and computationally efficient, we have used the concept of recursive embedding in a single frame. We have also explained how to reason using this representation. In the context of diagnosing LBP, the design has the properties like completeness, consistency, integrity, and non-redundancy. The design facilitates ease of access of the stored data items. We plan to use this scheme for implementation of medical expert system for LBP management in the near future.

References

1. WebMD: https://www.webmd.com
2. Shortliffe, E.H.: Computer-Based Medical Consultation: MYCIN (1976)
3. Weiss, S.M., Kulikowski, C.A., Safir, A.: A model-based consultation system for the long term management of glaucoma. In: 5th International Joint Conference on Artificial Intelligence, pp. 826–832 (1977)
4. Pople Jr., H.E.: The Formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning. In: 5th International Joint Conference Artificial Intelligence, pp. 1030–1037 (1977)
5. Kahn, M.G., Ferguson, J.C., Shortliffe, E.H., Fagan, L.M.: Representation and use of temporal information in ONCOCIN. In: Annual Symposium on Computer Application in Medical Care, pp. 172–176 (1985)
6. Toth-Tascau, M., Stoia, D.I., Andrei, D.: Integrated methodology for a future expert system used in low back pain management. In: 7th IEEE International Symposium on Applied Computational Intelligence and Informatics, pp. 315–320 (2012)
7. Abu Naser, S., AlDahdooh, R.: Lower back pain expert system diagnosis and treatment. J. Multi. Eng. Sci. Stud. **2**, 441–446 (2016)
8. Landi, A., Davis, R., Hendler, N., Tailor, A.A.: Diagnoses from an on-line expert system for chronic pain confirmed by intra- operative findings. J. Anesth. Pain Med. **1**, 1–7 (2016)
9. Shortliffe, E.: Computer-Based Medical Consultations: MYCIN (2012)
10. Kong, G., Xu, D., Yang, J.: Clinical decision support systems: a review on knowledge representation and inference under uncertainties. Int. J. Comput. Intell. Syst. **1**, 15–167 (2008)
11. Sundari, A.M., Balasundaram, R.: Computerisation in obstetrics and gynaecology—An expert approach. In: IEEE Engineering in Medicine and Biology Society and 14th Conference of the Biomedical Engineering Society of India, pp. 8–9 (1995)
12. Klar, R., Zaiss, A.: Medical expert systems: design and applications in pulmonary medicine. Lung **168**(Suppl), 1201–1209 (1990)
13. Barnett, G.O., Hoffer, E.P., Packer, M.S., Famiglietti, K.T., Kim, R.J., Cimino, C., Feldman, M.J., Oliver, D.E., Kahn, J.A., Jenders, R.A., Gnassi, J.A.: DXplain—Demonstration and discussion of a diagnostic decision support system. In: Annual Symposium on Computer Application in Medical Care, pp. 822 (1993)