



# How to Improve the Adaptation Phase of the CBR in the Medical Domain

Ludmila Puzstová<sup>1</sup>(✉), František Babič<sup>1</sup>, Ján Paralič<sup>1</sup>,  
and Zuzana Paraličová<sup>2</sup>

<sup>1</sup> Department of Cybernetic and Artificial Intelligence,  
Faculty of Electrical Engineering and Informatics,  
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic  
{ludmila.puzstova.2, frantisek.babic,  
jan.paralic}@tuke.sk

<sup>2</sup> Faculty of Medicine, University of Pavol Jozef Šafárik,  
Trieda SNP 1, 040 11 Košice, Slovak Republic  
zuzana.paralicova@unlp.sk

**Abstract.** This paper reviews the Case-based reasoning (CBR) approach and its usability in the medicine and presents a new concept on how to improve its adaptation phase. We use the CBR as a supporting method for decision support like diseases diagnostics or therapy identification. We investigated existing approaches, studies, and research works to solve one of the most critical problems in the CBR cycle - adaptation, which is often done manually by the experts in the relevant field. Based on the findings and our experiences with medical diagnostics through suitable data analytical methods, we proposed a new solution to solve this challenge. This approach is based on a comparison of the stored decision rules with the new one related to the current case. This comparison can result in three alternative states: (1) case base contains a similar case, and relevant rule can be applied. (2) The new case is very different from the stored ones, so the input from participated experts is needed, and a new rule will be stored. (3) The new case is partially similar satisfying adaptability conditions, in such a situation we adopt related decision rule to the new conditions under the supervision of the expert. We plan to experimentally test and verify this concept within available medical samples from our previous experiments.

**Keywords:** Case-based reasoning · Medical diagnostics · Data analysis

## 1 Introduction

Information and communication technologies are increasingly being deployed in the medical domain to support various activities on all sides of the relevant processes. This situation is also related to the ever-increasing amount of data that needs to be processed and analyzed. For doctors, it's hard to consider a higher volume of data in the diagnosis procedure, or in determining the right treatment.

The evidence of the World Health Organization (WHO) from 2015 indicates that up to 5% of patients had an incorrect diagnosis. The diagnostic procedure represents a complicated process, in which it is essential to have the right information available for

the right people at the right time. If analytical support models are available for doctors, they will help them consider all the contexts and important hidden patterns in the data. They will also reduce the time needed to make a decision. The input and continuous participation of the experts is an essential part of the analytical projects. In some cases, we can capture and store the expert's knowledge in a suitable formal way.

Many researchers apply various intelligent techniques to create decision support systems or models to help the doctor determine the correct patient diagnosis and enable them to design the best treatment for their current health condition.

Case-based reasoning (CBR) approach was proposed by Schank in 1982 to solve identified problems related to the decision support systems like knowledge elicitation, adaptation, or maintenance [1]. Since medicine requires experts with a mixture of knowledge and experience, case-oriented methods should be very efficient, mainly because reasoning with cases corresponds with the typical decision-making process of physicians. Also, incorporating new cases means automatically updating parts of the changeable knowledge [2].

In our previous research works focusing on the diagnostics of the various diseases like metabolic syndrome, mild cognitive impairment, heart or brain attacks, we typically extracted different models and hidden knowledge through suitable analytical methods. In most cases, these were relatively small samples with records up to 500. It means that the adaptation of the generated decision rules to new examples was simple, e.g., we spent some time re-generate the new ones. However, in this paper, we want to focus on a different situation, when the knowledge base contains a large number of rules, and it is important to decide about possible updates effectively.

The paper is organized as follows: the first section introduces our motivation and the topic, the second one presents the Case-based reasoning in the medical domain, identified challenges and new proposed approach for adaptation phase. The conclusion summarizes the main points and outlines future work.

## 1.1 Case-Based Reasoning

The CBR methodology has attracted significant attention because the basic idea of reusing experience to solve previous problems is a powerful and often used way of addressing people's issues. In CBR terminology, a case usually means a problem situation that one needs to resolve.

Doyle et al. in 1998 [3] described the CBR as a problem-solving paradigm in many ways significantly different from other major artificial intelligence approaches. But the situation has changed during the last years. Computational analogy-making and CBR are closely related areas. Analogy-making involves at least several subprocesses like building representation, retrieval from a base for the analogy, mapping onto the target, validation, and learning from the experience [4]. Other approaches rely only on general knowledge of the problem area, but CBR can use specific knowledge about previous problematic situations [5]. CBR reasoning has an incremental character [6]: it means that whenever the problem is resolved, the new experience will be retained and immediately available for future use of problem-solving.

Therefore, solving each problem in the CBR cycle consists of 4 phases such as retrieve, reuse, revise, and retain. We identified several models in existing literature like the Hunt model, the Allen model, the model by Kolodner and Leake [7], and the R4 model proposed by Aamodt and Plaza [8]. The R4 model is one of the most used and defines the CBR cycle with the following four primary steps.

Finding and retrieving the most similar cases is done at the RETRIEVE step. According to several authors [9, 10], this phase is one of the most important. It includes a case-finding process based on their similarity. For this purpose, we can typically use the nearest neighbor search, inductive approaches, knowledge approaches, Bayesian network, clustering Euclidian distance, or other similarity measures. In many practical applications, it is often difficult to distinguish the REUSE and REVISE steps because many researchers associate them into one phase called ADAPT (adaptation) [11]. In this step is the case used again, and the proposed solution is checked. In the last step - RETAIN - is performed preservation (storage) of the learned case for future use. There are several approaches to achieving this goal like retaining only the solution of the previous problem or the new one. In many cases, this retention process leads to uncontrolled growth of the case's base, which consequently causes the system performance to deteriorate in terms of speed [12].

## 2 Case-Based Reasoning in the Medical Domain

CBR reasoning process is medically accepted and getting increasing attention from the medical domain. In 1988, 1989 and 1991 were organized three CBR seminars by the American Defense Research Agency (DARPA), which officially marked the beginning of CBR discipline.

The authors in [13] present a summary of 21 studies that dealt with medical CBR systems. They described a list of methods used in each CBR step and the success rate in system verification. In the RETRIEVE phase, the most used methods are Euclidian distance, nearest neighborhood, similarity function, and weight set ranked by a decision tree. In the REUSE phase, authors used neural nets, fuzzy rules, stepwise regression, manual reuse, but most systems do not use any technique. The REVISE phase performed either manually, or do not use any specific method. In RETAIN phase are cases stored manually or not at all. The most used evaluation methods are k-fold cross-validation, leave-one-out strategy, conditional probability, AUC curve, statistical frequency, and correlation.

Many studies [11, 14] have attempted to investigate existing medical CBR systems since 1987. The most systems were developed to solve a specific disease; most systems perform as prototypes and not as the final product. Another visible trend was the successful hybridization of CBR with various computational methods. According to [11], in 32 systems out of 76, CBR was used in combination with other techniques. Also, out of 76 systems, in 51 systems, automatic adaptation is completely avoided, so they only work as retrieval systems.

The use of CBR in the medical field is currently reviving. The knowledge base of medical knowledge is continually changing; sometimes, there is more than one solution; doctors have different approaches and medicines. The fact that the CBR system

methodology very much resembles the doctor's thinking process suggests the successful use of CBR in medicine [15]. The main advantage of CBR in this field is the possibility of adapting the knowledge base [16], which is a significant aspect of decision making in the medical field.

## 2.1 Existing Limitations

Although the use of the CBR method appears to be successful, there are some limitations. In the medical domain, the number of similar cases is often extremely high, and this fact causes a complex generalization [14]. High memory/storage requirements and time-consuming retrieval accompany CBR systems utilizing large case bases and can take significant processing time to find similar cases in case-base. CBR systems have problems with handling noisy data. Unsuccessful assessment of such noise may result in the same problem being unnecessarily stored numerous times. In turn, this implies inefficient storage and retrieval of cases. The number of systems using the full CBR cycle (retrieve, adaptation, retain) is still very low. However, the most critical problem in the successful implementation of CBR techniques in medical systems is the problem of the adaptation step. D'Aquin et al. [16] note that this step is a relatively complex process because it has to address the lack of relevant patient information, the usability, consequences of the decision, the proximity of decision thresholds and the need to consider patients in different ways. Schmidt et al. indicate that introducing the adaptation step in the CBR system was a challenging step in medicine [17]. Most CBR systems which don't apply the adaptation step, can't solve some new problems, and thus their accuracy is unconvincing in critical areas [5]. The adaptation phase is, therefore limited to planning tasks [18].

## 2.2 Adaptation Step Problem

The study [13] mentioned that medical CBR systems solve the problem of adaptation in two ways. Most systems avoid an adaptation problem by applying only the RETRIEVE step in the CBR cycle, while others are trying to resolve it. One of the first medical expert systems CASEY [19] attempted to solve the adapting problem through rule-based domain theory. Knowledge acquisition is a barrier to the development of rule-based systems; therefore, the development of adaptation rules has never become a successful technique in medical CBR systems [17]. Some of the newer systems successfully used adaptation using computational techniques, e.g., eXiT \* CBR.v2 [20] revises and reuses cases using genetic algorithms; EquiVox developed by Henriet et al. [21] performs adaptation using artificial neural networks.

The studies [5, 6] solved the adaptation problem by the creation of a hybrid CBR system integrating CBR (case) and RBR (rule) reasoning. This system automatically applied the adaptation process using adaptive rules.

In the study [5], after the resolution of the new case, the knowledge base was expanded, and the adaptation and reasoning rules were updated. To achieve integration into REUSE step was added a new process called REASON, which applied the reasoning rules to get a solution if the REUSE and ADAPT process failed to find a solution. They first applied the CBR and after that, RBR to the available data. The authors used

multiple cross-validations to evaluate accuracy. The developed prototype achieved an average accuracy of 99.53% on the diagnosis of thyroid disease and 99.33% on breast cancer diagnosis (accuracy by other systems ranged from 80% to 97%).

The study [6] provided a hybrid system to help healthcare professionals in early diagnosis on cancer patients. In the proposed approach, CBR was used as the primary reasoning process, and RBR was used to improve part of the process. For this research, they gathered real data about patients with gastrointestinal cancer. To evaluate accuracy, they also used multiple cross-validations. The results showed increased diagnosis accuracy by 22.92% compared to the use of a single CBR method.

Salem and El Bagouras [22] have proposed a hybrid adaptation model that combines transformational and hierarchical adaptation techniques with artificial neural networks and factors for the diagnosis of thyroid cancer. Zubi and Saad [23] used combined data mining techniques with neural networks for early diagnosis of lung cancer. For the diagnosis of breast cancer, Keles, Keles, and Yavuz [24] used neuro-fuzzy rules, while Sharaf-el-Deen et al. [25] introduced a hybrid approach that also combined CBR and RBR reasoning.

We can see that authors tried to solve the adaption problem in three ways: avoiding the adaptation problem by using CBR systems only for RETRIEVE step; the use of computational techniques such as genetic algorithms and artificial neural networks; creating a hybrid CBR system that integrates CBR and RBR reasoning.

### 2.3 New Proposal How to Support the Adaptation

Figure 1 presents graphically our approach on how to support the adaptation phase in the CBR cycle. The assumption is a list of decision rules generated by suitable machine learning algorithms stored in case base: IF conditions THEN consequences (target value, expected diagnosis).

The CBR cycle starts with the RETRIEVE step as a response to a new example without target diagnosis. The new case is compared with existing ones from the case base by an inference mechanism. We will calculate the distances between cases with similarity metrics like Euclidean, Manhattan, or Hamming distance. The result of the comparison can be one of the three alternatives:

1. The mechanism will find an identical case to the new one. The target diagnosis will be the same as for the existing one.
2. The mechanism will not find a match; all stored cases are significantly different. This situation requires re-generate the current rules based on the original set of records extended by the latest case classified by the expert.
3. The mechanism will find partly similar cases with different target values. Therefore, the CBR cycle will continue with other steps like REUSE, REVISE, and RETAIN.

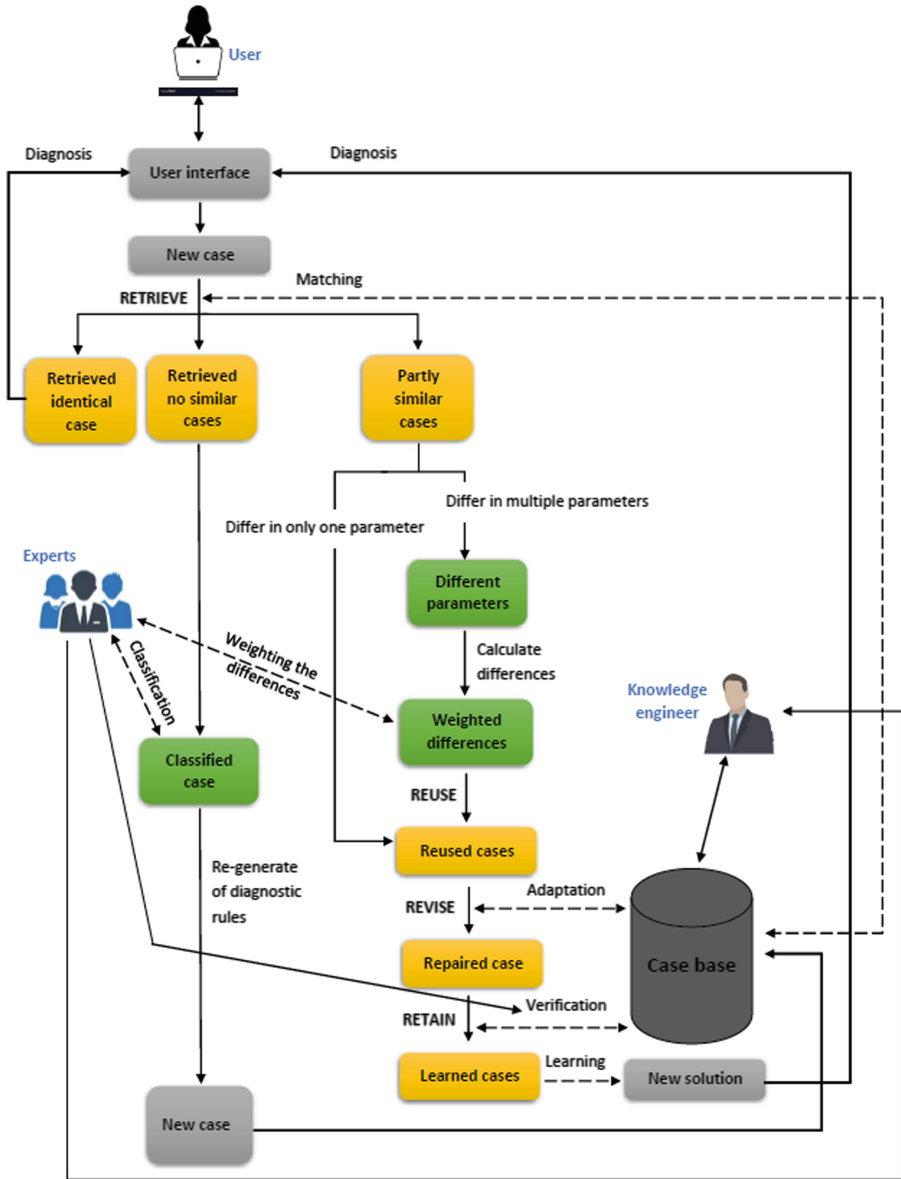


Fig. 1. The new approach for adaptation

Before the cycle continues, we will investigate the differences:

- If the cases differ only in one condition (parameter) on the left side of the rules, the expert will consider possible adjusting of it. After several iterations, we will be able to create a separate knowledge base with the knowledge from the experts and will be able to do this step in a semi-automatic way. An example:

**Formal scenario:**

New case: IF  $parameter_1 = X$  AND  $parameter_2 = Y$   
 Decision rule: IF  $parameter_1 \in \langle Z, V \rangle$  AND  $parameter_2 \in \langle K, L \rangle$   
 THEN  $target\ value = 1$   
 Comparison:  $X \in \langle Z, V \rangle$  AND  $(Y < K$  OR  $Y > L)$

The cases differ in one parameter (**parameter<sub>2</sub>**). Therefore, the expert considers the following adaptation of the stored rule, and the new case will be classified as a positive diagnosis (REUSE-REVISE-RETAIN).

Adapted rule 1: IF  $parameter_1 \in \langle Z, V \rangle$  AND  $parameter_2 \in \langle K, Y \rangle$   
 THEN  $target\ value = 1$   
 Adapted rule 2: IF  $parameter_1 \in \langle Z, V \rangle$  AND  $parameter_2 \in \langle Y, L \rangle$   
 THEN  $target\ value = 1$

**Specific scenario:**

New case: IF LDL = 1.8 AND HDL = 4.6  
 Decision rule: IF LDL  $\in \langle 1.5, 3.1 \rangle$  AND HDL  $\in \langle 3.1, 4.5 \rangle$  THEN MCI = 1  
 Adapted rule: IF LDL  $\in \langle 1.5, 3.1 \rangle$  AND HDL  $\in \langle 3.1, 4.6 \rangle$  THEN MCI = 1

If the system will find several partially similar cases with different decision rules, it is possible to assign the weights by the experts expressing their suitability. This part of the concept will be an objective of further research.

- If the cases differ in multiple parameters:
  1. We identify a list of different parameters.

**Formal scenario:**

New case: IF  $parameter_1 = X$  AND  $parameter_2 = Y$  AND  $parameter_3 = Z$   
 THEN  $target\ value = 1$   
 Decision rule: IF  $parameter_1 \in \langle A, B \rangle$  AND  $parameter_2 \in \langle C, D \rangle$  AND  
 $parameter_3 \in \langle E, F \rangle$   
 THEN  $target\ value = 1$   
 Comparison:  $X \in \langle A, B \rangle$  AND  $(Y < C$  OR  $Y > D)$  AND  $(Z < E$  OR  $F > D)$

These cases are different in **parameter<sub>2</sub>** and **parameter<sub>3</sub>**.

2. For each of these parameters, we calculate a difference with existing cases with suitable similarity metric. Next, the expert will help us to allocate weights by importance for particular differences.
3. The parameters with high weights will be adapted to the most similar case, and the target class will be determined (REUSE-REVISE-RETAIN).

**Specific scenario:**

New case: IF LDL = 1.8 AND HDL = 4.6 AND BMI = 34 THEN MCI = 1  
 Decision rule: IF LDL  $\in$  <1.5, 3.1> AND HDL  $\in$  <3.1, 4.5> AND BMI  $\in$  <25.1, 29.3> THEN MCI = 1  
 Adapted rule 1: IF LDL  $\in$  <1.5, 3.1> AND HDL  $\in$  <3.1, 4.6> AND BMI  $\in$  <25.1, 34> THEN MCI = 1  
 etc.

### 3 Conclusion

The CBR methodology has attracted significant attention because the basic idea of reusing experience to solve previous problems looks very attractive. It can use specific knowledge about past problematic situations solving. The number of medical systems using the full CBR cycle (retrieve, adaptation, retain) is still very low. The most critical issue is the successful adaptation step. We propose a new concept to solve this issue. We found the inspiration in the research of professor Holzinger research group called interactive machine learning (iML) with a human-in-the-loop. This approach leads to algorithms that can interact with both computational agents and human agents and can optimize their learning behavior through these interactions [26, 27].

For this purpose, we use a combination of data analysis methods and CBR extending by communication with an expert, which helps us determine the importance of the parameters, their settings and the determination of the suitable adaptation.

In future work, we will focus on experimentally testing and verification of the proposed approach on the available medical data samples.

**Acknowledgment.** The work was partially supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant no. 1/0493/16 and The Slovak Research and Development Agency under grants no. APVV-16-0213 and APVV-17-0550.

### References

1. López-Fernández, H., Fdez-Riverola, F., Reboiro-Jato, M., et al.: Using CBR as design methodology for developing adaptable decision support systems. In: Jao, C.S. (ed.) *Efficient Decision Support Systems - Practice and Challenges From Current to Future*. IntechOpen, Croatia (2011)
2. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., Gierl, L.: Cased based reasoning for medical knowledge-based systems. *Int. J. Med. Informatics* **64**(2), 355–367 (2001)
3. Doyle, M., Hayes, C., Cunningham, P., Smith, B.: *CBR Net: smart technology over a network*. Department of Computer Science, Trinity College Dublin (1998)
4. Kokinov, B., French, R.M.: Computational models of analogy-making. In: Nadel, L. (ed.) *Encyclopedia of Cognitive Science*, vol. 1, pp. 113–118. Nature Publishing Group, London (2003)



5. Sharaf-El-Deen, D.A., Moawad, A.F., Khalifa, M.E.: A new hybrid case-based reasoning approach for medical diagnosis systems. *J. Med. Syst.* **38**, 9 (2014). <https://doi.org/10.1007/s10916-014-0009-1>
6. Saraiva, R.M., Bezerra, J., Perkusich, M., Almeida, H., Siebra, C.: A Hybrid Approach Using Case-Based Reasoning and Rule-Based Reasoning to Support Cancer Diagnosis: A Pilot Study. *MedInfo* (2015). <https://doi.org/10.3233/978-1-61499-564-7-862>
7. Leake, D.B.: *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. MIT Press, Cambridge (1996)
8. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational Issues, Methodological Variations, and System Approaches. In: *AI Communications*, vol. 7, pp. 39–59. OSI Press (1994)
9. De Mantaras, R.L., et al.: Retrieval, reuse, revision, and retention in case-based reasoning. *Knowl. Eng. Rev.* **20**(03), 215–240 (2005)
10. Pal, S.K., Shiu, S.C.: *Foundations of Soft Case-Based Reasoning*. Wiley, Chichester (2004)
11. Holt, A., Bichindaritz, I., Schmidt, R., Perner, P.: Medical applications in case-based reasoning, September 2005. <https://doi.org/10.1017/s0269888906000622>
12. Lawanna, A., Daengdej, J.: Methods for case maintenance in case-based reasoning. *Int. J. Comput. Inform. Eng.* **4**, 10–18 (2010)
13. Blanco, X., Rodríguez, S., Corchado, J.M., Zato, C.: Case-based reasoning applied to medical diagnosis and treatment. In: Omatu, S., Neves, J., Rodriguez, J.M.C., Paz Santana, J. F., Gonzalez, S.R. (eds.) *Distributed Computing and Artificial Intelligence*. AISC, vol. 217, pp. 137–146. Springer, Cham (2013). [https://doi.org/10.1007/978-3-319-00551-5\\_17](https://doi.org/10.1007/978-3-319-00551-5_17)
14. Choudhury, N.: A survey on case-based reasoning in medicine. October 2018. <https://doi.org/10.14569/ijacsa.2016.070820>
15. Macura, R.T., Macura, K.: Case-based reasoning: opportunities and applications in health care. *Artif. Intell. Med.* **9**(1), 1–4 (1997)
16. D’Aquin, M., Lieber, J., Napoli, A.: Adaptation knowledge acquisition: a case study for case-based decision support in oncology. *Comput. Intell.* **22**(3–4), 161–176 (2006)
17. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., Gierl, L.: Cased-based reasoning for medical knowledge-based systems. *Int. J. Med. Informatics* **64**(2), 355–367 (2001)
18. Pous, C., et al.: Modeling reuse on case-based reasoning with application to breast cancer diagnosis. In: Dochev, D., Pistore, M., Traverso, P. (eds.) *AIMSA 2008. LNCS (LNAI)*, vol. 5253, pp. 322–332. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-85776-1\\_27](https://doi.org/10.1007/978-3-540-85776-1_27)
19. Koton, P.: A medical reasoning program that improves with experience. *Comput. Methods Programs Biomed.* **30**(2), 177–184 (1989)
20. Pla, A., López, B., Gay, P., Pous, C.: Distributed case-based reasoning tool for medical prognosis. *Decis. Support Syst.* **54**(3), 1499–1510 (2013)
21. Henriet, J., Leni, P.E., Laurent, R., Salomon, M.: Case-based reasoning adaptation of numerical representations of human organs by interpolation. *Expert Syst. Appl.* **41**(2), 260–266 (2014)
22. Salem, A.B.M. Bagoury, B.M.E.L.: A case-based adaptation model for thyroid cancer diagnosis using neural networks. In: *Society, Florida Artificial Intelligence Research*, Cairo, Egypt, pp. 155–159 (2003)
23. Zubi, Z.S., Saad, R.A.: Using some data mining techniques for early diagnosis of lung cancer. In: *Recent Researches in Artificial Intelligence, Knowledge Engineering, and Data Base*, Tripoli, Libya, pp. 32–37 (2011)
24. Keles, A., Keles, A., Yavuz, U.: Expert system based on neuro-fuzzy rules for diagnosis breast cancer. *Expert Syst. Appl.* **38**(5), 5719–5726 (2011)

25. Sharaf-elDeen, D.A., Moawad, I.F.: A breast cancer diagnosis system using hybrid case-based approach. *Int. J. Comput. Appl.* **72**, 9–14 (2013)
26. Holzinger, A.: Interactive machine learning (iML). *Informatik Spektrum* **39**(1), 64–68 (2016)
27. Holzinger, A.: Interactive machine learning for health informatics when do we need the human in the loop. *Brain Inf.* **3**(2), 119–131