

Explanation-aware computing of the prognosis for breast cancer supported by IK-DCBRC: Technical innovation

Abdeldjalil Khelassi^{1,2}

- 1: D.Sc. in Computer Science, Associate Professor, Department of Informatics, Faculty of Sciences, Tlemcen University, Tlemcen, Algeria
- 2: Head of Knowledge and Information Engineering Research Team KIERT, Informatics Research Laboratory IRL, Tlemcen University, Tlemcen, Algeria

Corresponding author:

Associate Professor Dr. Abdeldjalil Khelassi, BP 230 Computer Science Department, Science Faculty, Tlemcen University, Tlemcen, Algeria, Telephone: +213550168666, Fax: +21343286308. Email: khelassi.a@gmail.com

Abstract:

Background: Active research is being conducted to determine the prognosis for breast cancer. However, the uncertainty is a major obstacle in this domain of medical research. In that context, explanation-aware computing has the potential for providing meaningful interactions between complex medical applications and users, which would ensure a significant reduction of uncertainty and risks. This paper presents an explanation-aware agent, supported by Intensive Knowledge-Distributed Case-Based Reasoning Classifier (IK-DCBRC), to reduce the uncertainty and risks associated with the diagnosis of breast cancer.

Methods: A meaningful explanation is generated by inferring from a rule-based system according to the level of abstraction and the reasoning traces. The computer-aided detection is conducted by IK-DCBRC, which is a multi-agent system that applies the case-based reasoning paradigm and a fuzzy similarity function for the automatic prognosis by the class of breast tumors, i.e. malignant or benign, from a pattern of cytological images.

Results: A meaningful interaction between the physician and the computer-aided diagnosis system, IK-DCBRC, is achieved via an intelligent agent. The physician can observe the trace of reasoning, terms, justifications, and the strategy to be used to decrease the risks and doubts associated with the automatic diagnosis. The capability of the system we have developed was proven by an example in which conflicts were clarified and transparency was ensured.

Conclusion: The explanation agent ensures the transparency of the automatic diagnosis of breast cancer supported by IK-DCBRC, which decreases uncertainty and risks and detects some conflicts.

Keywords: computer-aided diagnosis, case-based reasoning, multi-agent system, breast cancer prognosis, explanation-aware computing

Additional Information for citing this article:

Title of Journal: Electronic physician; Abbreviated title of journal: Electron. Physician

doi: 10.14661/2014.947-954

Editorial information:

Type of article: original

Received: June.20.2014

Revised: August.24.2014

Accepted: September.10.2014

Published: November.27.2014

© 2014 The Authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. Introduction

Cancer refers to a group of maladies characterized by uncontrolled cell proliferation in normal tissue of the body. Three proprieties can distinguish between malignant and benign tumors: 1) uncontrolled growth, 2) invasion, and 3) metastasis (1). According to previous research, the risk factors of cancer include the use of tobacco and alcohol (2, 3), obesity and inactivity (4, 5), nutritional insufficiencies, such as low intake of fruits and vegetables (6), pollution (including indoor smoke from household use of solid fuels) (7), and some sexually-transmitted infections, such as human papillomavirus (HPV) (8, 9). Research also has proven that the risk of cancer increases with age (10).

Cancer caused about 8.2 million human deaths in 2012, with breast cancer accounting for 521,000 of the deaths (11). The prognosis of breast cancer is a multi-disciplinary research domain that involves several specialities, including medicine, biology, psychology, and computer science. Also, some clinical techniques can be used for making diagnoses, e.g., microarray (12, 13), mammography (12, 14, and 15), and cytology (16, 17). Several computational techniques are used for computer-aided detection of breast cancer, e.g., artificial neural networks ANNs (14, 16, and 18), Bayes network (12, 18), genetic algorithms (19), fuzzy logic (14), linear programming (LP) (20), support vector machine (SVM) (18, 21), and k-nearest neighbors (KNN) (17). A promising result was generated by the intensive knowledge-distributed case-based reasoning classifier (IK-DCBRC) (22), which implements fuzzy sets theory (23) and the Case-Based Reasoning and Distributed approach for the automatic detection of the class of malignancies from cytological images. To date, uncertainty, which is caused by several natural and technical factors, remains an important obstacle in obtaining an accurate diagnosis. These factors include faults in the measures that are taken and the uncertainty of reasoning systems. In the interest of overcoming these issues, some computational techniques have been introduced, including fuzzy sets theory (23, 24), probability reasoning (12, 24), and explanation-aware computing (25).

In this work, we presented an intelligent agent, supported by IK-DCBRC, for explaining the prognosis of breast cancer. The agent that was developed uses a rule-based system that explains the reasoning trace, the strategy that was used, justification for the choice, and the terms that are used. The rest of this paper presents the IK-DCBRC system, explanation-aware computing, breast cancer prognosis and computer-aided detection, the explanation agent that was developed, and a discussion of the effectiveness of the proposed approach.

2. Material and Methods

2.1. IK-DCBRC System for computer-aided diagnosis

The case-based reasoning paradigm was implemented in several computer-aided medical diagnoses, some of which were reviewed in (26). Also, the system used in this survey, i.e., IK-DCBRC, was mentioned in (22, 27), and it has shown promising results for use in the diagnosis of cardiac arrhythmia (27) and breast cancer (22). IK-DCBRC uses Distributed, Case-Based Reasoning by integrating a set of agents and a set of case bases, each of which contains cases from the same class (disease), and just one agent, called the Similarity agent, can inferred from this case base. An adaptation agent is integrated for selecting the class from the responses of the other agents. A knowledge base that contains the physician’s rules also is inferred by the adaptation agent. Figure 1 shows the graphical user interfaces with the architecture of IK-DCBRC.

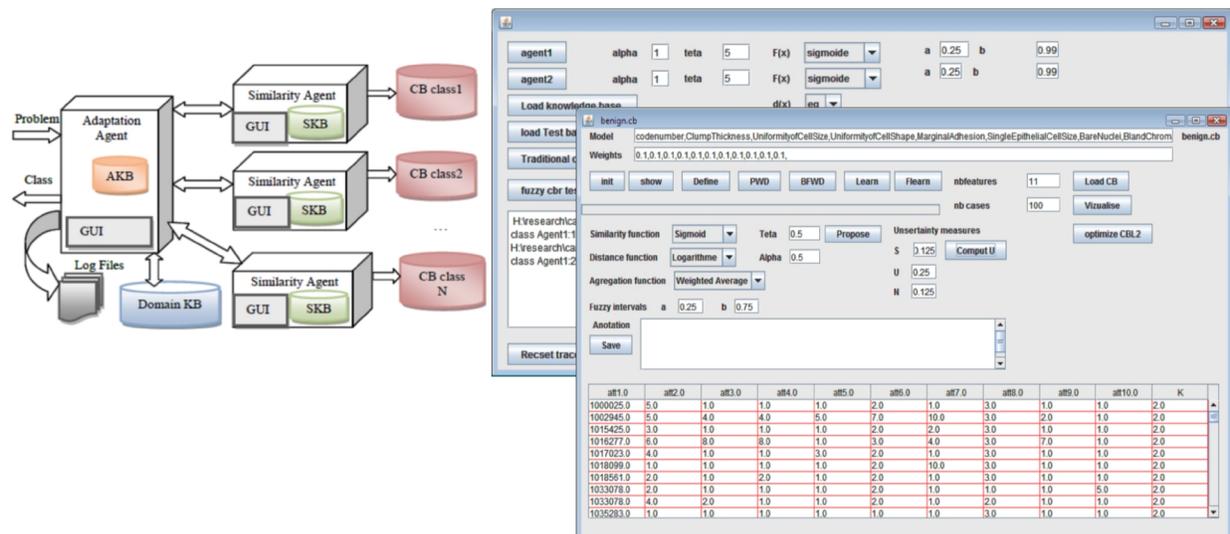


Figure 1. KI-DCBRC: The system used in computer-aided diagnosis

The log files generated by the system contains all needed information for explanations since the agents’ messages and the computed similarity measures of each similarity agent are included, as well as the uncertainty of the responses provided by adaptation rules and similarity parameters that were used. The breast cancer dataset used for machine learning and testing was collected and tested by O. L. Mangasarian and W. H. Wolberg (17, 20). They did

so by applying linear programming and the k-nearest neighbor algorithm and achieved rates of correct classification results of 91.5 and 93.5%, respectively. Using IK-DCBRC and several different strategies (26), the precision of the various approaches were determined, ranging from 98 to 100%.

2.2. Explanation-aware computing for medical applications

The concept of explanation is a common area of interest in several related domains, such as didactics, physics, cognitive systems, philosophy. Intelligent systems, as an applied science of cognitive systems, are complex systems that are used by a wide variety of critical applications in which uncertainty represents an important risk. Under these circumstances, the explanation of the behaviors of intelligent systems becomes a necessity for a trusted interaction between the users and these complex systems. Several researchers have introduced the explanation concept in their intelligent systems, which are used to generate a variety of extensions for ensuring pragmatism in the human-machine interaction process.

The uncertainty in medical applications is an important issue that has been considered by several of the systems that have been developed. The meaningful interaction proposed by explanation systems impacts the uncertainty and can reduce the risks in the medical applications. Explanation-aware computing in computer-aided diagnosis is a complex task in which machine learning can be used as discussed in (28). In the existing systems that support the explanation, there are four forms:

- Reasoning traces as in (29) in which the explainer generates the trace of reasoning processes
- Justifications as in (30) in which the explainer provides justifications for a reasoning to be reused in future judgments
- Reasoning strategy as in the use of neomycin (30)
- Terminology identified in (31) for explaining the terms and concepts of the domain

Explanation-aware computing becomes a necessity for all complex medical applications in which uncertainty and risks are the first preoccupation, and they are seriously considered because human life is at stake. Several researchers contribute the explanation in their medical applications as well as, e.g., MYCIN in bacterial diagnosis and therapy (29), Öztürk and Aamodt (32) in hypothesis and test strategies for diagnosis in Electronic Medical Record, and processing MR Images (33).

2.3. The Breast Cancer Prognosis

Breast cancer prognosis research refers to many techniques, including microarray (25), cytology (18-13) and mammography (19). Cytology analysis is an indispensable test before any surgical intervention for distinguishing between malignant and benign tumors. Figure 2 shows two samples of tumors cells with benign on the left and malignant on the right.

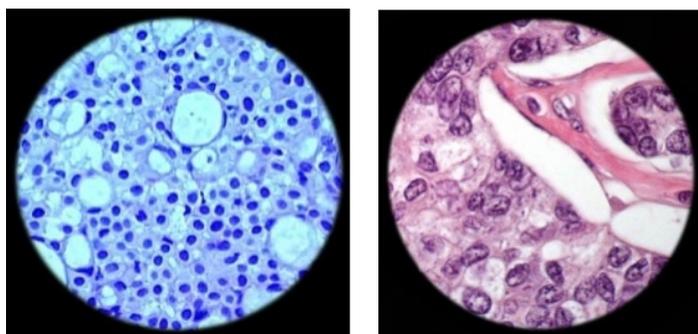


Figure 2. Sample of cytological images of breast cancer: left benign and right malignant (34, 35)

In the literature, there are several computational techniques that have been used for the computer-aided detection and diagnosis of breast cancer. These techniques include artificial neural networks ANNs (14, 16, and 18), bayes network (12, 18), genetic algorithms (19), fuzzy logic (14), linear programming (20), SVM (18, 21), and KNN (17). And the different sources of information include microarray (12, 13), mammography (12, 14, and 15), and cytology (16, 17, and 20). Figure 3 summarizes some of the published accuracies of these techniques. A promising result was generated by IK-CBRC cited in (22) and described in section 2.1, which used the hybrid and distributed approach in the automatic detection of malignant cancer class from the cytological images. The results are well described in (22).

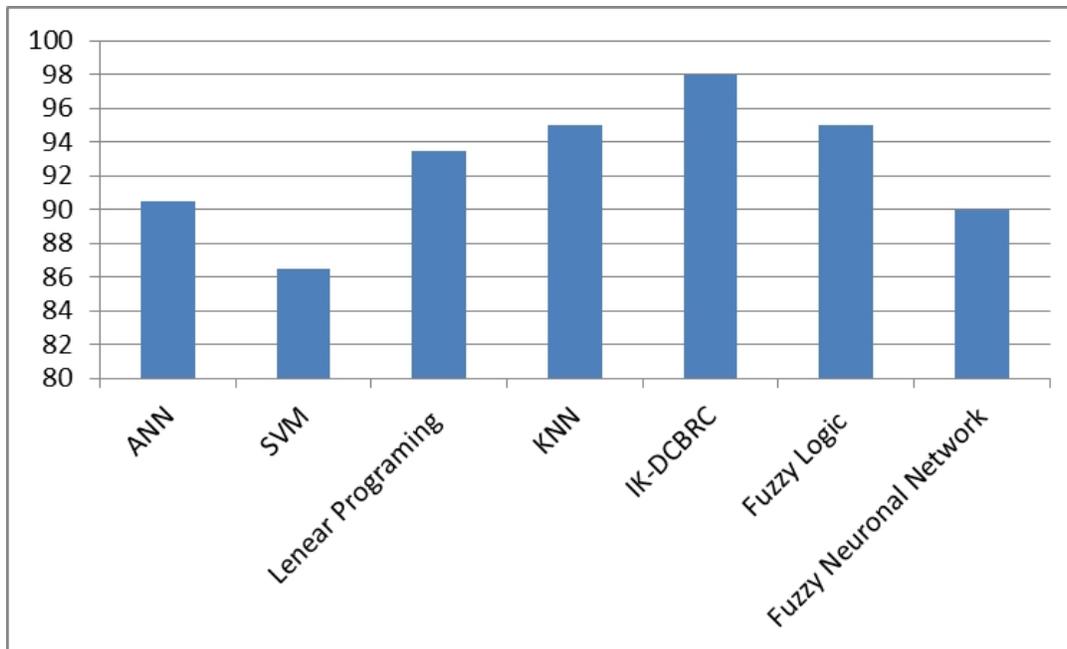


Figure 3. Accuracy of some computational techniques in breast cancer diagnosis from the literature (16, 18, 20, 17, 22, 14, and 14, respectively)

3. Results

3.1 Explanation process

Figure 4 describes the architecture of the explanation agent and its environment. First, the cytological images should be analyzed by Xcyt (34), which is a remote cytological image analyzer, to create a query. The IK-DCBRC classifies the generated query to distinguish between the malignant and benign classes. The explanation agent explains the results of the previous diagnosis according to the rules of the explanation knowledge base and the logged information.

Via the explanation agent's user interface, shown in Figure 5, the user selects the demanded level of abstraction for the explanation. As in (29-31), the explanation agent visualizes the trace of reasoning, justification, the terminology, and the applied strategy according to the selected level of abstraction.

The cognitive agent for explanation defines the following three levels of abstraction:

- **Level 1: Ordinary explanations.** In this level, the agent presents just the information that is needed about the class of the query, the disease, associated with the terminological explanations, the indexed links of useful wiki documents, and the cytological image.
- **Level 2: Deep explanations.** In this level the agent adds the justification, the reasoning trace, and the strategy to the explanation presented in the prior level.
- **Level 3: Maintaining explanations.** This level contains all kinds of explanations plus the professional information that can help developers understand the abnormality of the reasoning system in the case of errors.

3.2. Explanation example

Figure 5 shows an example of explanation in which the user selects Level 3 and the log file of the breast cancer diagnosis case. After visualizing the explanation, in the reasoning trace, the first agent associated with the benign class generates that the diagnosed case is not similar with a rate of 17% and 83% unknown; the response of the second agent associated with the malignant class is 1% similar and 99% unknown. The response of the adaptation agent is malignant by using rule number 2, and no rule was used by the rule-based system. The terminological explanation also is presented with the wiki doc and the localization of the abnormality in the breast cancer pattern. The strategic explanation is ensured by visualizing the similarity function that was used and the weights of the features for each class.

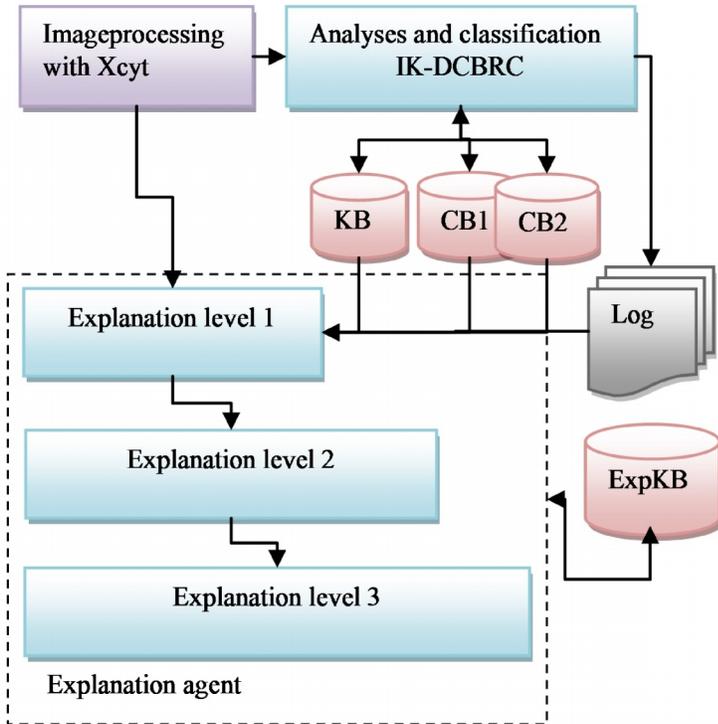


Figure 4. Explanation agent and its environment

```

Explanation
Query information:
ID:1352300020367
patterncode number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epitheli
features1017122,8,10,10,8,7,10,9,7,1,4,
Reasoning Trace:
RBS:
RBS:unknwonrule Num:_
Agent1:
similarity func:1 weight ,0,0,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.0
response: sim: 0.0ukn: 0.83nsim: 0.17
Agent2:
similarity func:1 weight ,0,0,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.11,0.0
response: sim: 0.01ukn: 0.99nsim: 0.0
Adaptation agent descision:
RBSunknwonrule Num:_
AKB Malignant rule Num:2
terms:Cytology , Breast cancer, Malignant
terminological doc:
http://en.wikipedia.org/wiki/Breast_cancer
    
```

Figure 5. Explanation Agent's explanation of the breast cancer diagnosis

4. Discussion

As described in the comparative diagram in Figure 3, in which the accuracy was between 86.5 and 98.5% and error rate was between 1.5 and 13.5%, the existence of risks in the automatic detection of breast cancer was proven. This risk changes from one computational technique to another. But researchers should detect and prevent doubts about decreasing these risks, and this is the aim of this contribution. In this paper, a cognitive agent was described for medical explanation by reusing the trace of reasoning, the case bases, and the adaptation knowledge base of IK-DCBRC. The interaction between the agent and the users is personalized and made meaningful via a graphical interface with the user. Many doubts can be detected by visualizing the explanations after the diagnosis. The multi-level of abstraction in the explanation process ensures an appropriate assignment of explanation according to the context of the uses of the application.

In Figure 5, the explained case is well classified, but the decision was deduced with just 1% of membership in the malignant case base, which increases the doubts, but the explanation rule explains that the decision was deduced also by considering the membership of the case in the non-similar set of the benign class, i.e., 17%, which decreases the doubts about this diagnosis. Although the proposed solution decreases the risks of breast cancer prognosis, there are other sources of risks that are not considered as faults of measures, conflicts in the knowledge bases, and others. But the visualization of the reasoning trace can guide the detection of conflicts and faults.

5. Conclusions

This study makes an original contribution for the resolution of the problem of uncertainty in breast cancer prognosis supported by IK-CBRC. This contribution consists of introducing an explanation-aware computing process after the automatic detection of the class of breast cancer. The proposed solution impacts breast cancer diagnosis by sensitizing physicians to the entire context of the computer-aided diagnosis. Many of the advantages and benefits of this proposition are discussed and approved.

Acknowledgments:

The authors thank so much every person who shared useful information with us or provided valuable guidance. We also give special thanks for Professors: Thomas Roth-Bergopher, Mohammed Amine Chikh, and Klaus-Dieter Althoff for the advice they provided.

Conflict of Interest:

There is no conflict of interest to be declared.

References

1. Hanahan D, Robert A. Weinberg, « Hallmarks of cancer: the next generation », Cell, 2011, 144(5): 646-74. PMID : 21376230, DOI: 10.1016/j.cell.2011.02.013
2. Avraham Z, Baron-Epel O, Boker LK. [The relationship between passive smoking, breast cancer risk and n-acetyltransferase 2 (NAT2)]. Harefuah. 2014;153(3-4):171-5, 238. Epub 2014/05/06. PMID: 24791559.
3. Hyun Ja Kim, Seungyoun Jung, Wendy Chen, Walter C. Willett, and Eunyoung Cho Alcohol consumption and breast cancer risk by family history of breast cancer and folate intake: A prospective cohort study, The FASEB Journal. 2013;27:lb387
4. Travier N, Fonseca-Nunes A, Javierre C, Guillamo E, Arribas L, Peiro I, et al. Effect of a diet and physical activity intervention on body weight and nutritional patterns in overweight and obese breast cancer survivors. Med Oncol. 2014;31(1):783. Epub 2013/12/07. Doi: 10.1007/s12032-013-0783-5. PMID: 24310809.
5. Yili Wu, Dongfeng Zhang, Shan Kang 'Physical activity and risk of breast cancer: a meta-analysis of prospective studies', Breast Cancer Res Treat, February 2013, 137(3): 869-82, PMID: 23274845, doi: 10.1007/s10549-012-2396-7.
6. D. Aune, D. S. M. Chan, A. R. Vieira, D. A. Navarro Rosenblatt, R. Vieira, D. C. Greenwood, T. Norat Fruits, vegetables and breast cancer risk: a systematic review and meta-analysis of prospective studies, Breast Cancer Res Treat, July 2012, 134(2): 479-93, PMID: 22706630, doi: 10.1007/s10549-012-2118-1.
7. Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors, The Lancet, November 2005, 366(9499): 1784-93. PMID: 16298215

8. Catherine de Martel, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, et. al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncol* 2012;13(6),607-15 doi:10.1016/S1470-2045(12)70137-7
9. Bani Hashemi S.H, Karimi S., Mahboobi H. Lifestyle changes for prevention of breast cancer. *Electron. Physician*, 2014; 6 (3): 894-905. Doi: 10.14661/2014.894-905
10. Cancer Research UK (January 2007). "UK cancer incidence statistics by age". Available from: <http://info.cancerresearchuk.org/cancerstats/incidence/age/>.
11. GLOBOCAN 2012, IARC. Available from: <http://globocan.iarc.fr/>
12. Nahar J, Imam T, Kevin S. Tickle, A. B. M. Shawkat Ali, Phoebe Chen Y. Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer. *Expert system with applications* 2012, 39(16):12371-7. doi: 10.1016/j.eswa.2012.04.045
13. Bechindaritz I, Anest A. "Case based reasoning with Bayesian Model Averaging: An improved method for survival analysis on Microarray Data" ICCBR10, 18th Int Conf in Case Based Reasoning, Alexandria, Italy, LNAI6176 Springer (2010): 346-59,
14. Moghbel M, Mashohor S. A review of computer aided detection diagnosis (CAD) in Breast thermography for breast cancer detection. *Artificial Intelligence Review*. 2013, 39(4): 305-13, doi: 10.1007/s10462-011-9274-2.
15. Pisano E. D, Gatsonis C, Hendrick E, Yaffe M, Baum J. K, Acharyya S, et. Al. Diagnostic performance of digital versus film mammography for breast-cancer screening, *N Engl J Med*. 2005, 353(17):1773-83, DOI: 10.1056/NEJMoa052911.
16. Saritas I. Prediction of breast cancer using artificial neural network. *J. of Medical System*, 36(5): 2901-7, doi: 10.1007/s10916-011-9768-0
17. William H. Wolberg, Mangasarian O. L. "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", *Proceedings of the National Academy of Sciences, U.S.A.*, December 1990, 87:9193-6.
18. Maglogiannis I, Zafiroopoulos E, Anagnostopoulos I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based Classifiers. *Applied Intelligence*, 2009, 30(1): 24-36. Doi: 10.1007/s10489-007-0073-z
19. Cheng-Hong Y, Yu-Da L, Li-Yeh C, Hsueh-Wei C. Evaluation of breast cancer Susceptibility using Improved Genetic Algorithms to generate genotype SNP Barcodes, 2013, *IEEE\ACM Transactions on Computational Biology and Bioinformatics*. 2013, 10(2): 361-71. Doi:10.1109/TCBB.2013.27
20. Mangasarian O. L, Wolberg W. H. "Cancer diagnosis via linear programming", *SIAM News*, September 1990, 23(5): 1-18.
21. Ruey-Feng C, Wen-Jie W, Woo Kyung M, Yi-Hong C, Dar-Ren C. Support Vector Machines for Diagnosis of Breast Tumors on US Images, *Academic Radiology*. February 2003, 10(2): 189-97, doi: 10.1016/S1076-6332(03)80044-2
22. Khelassi A. "Data mining application with case based reasoning classifier for breast cancer decision support". *MASAUM International Conference on Information Technology 2012 MICIT'12*; 07/2012.
23. Zadeh, L. Fuzzy Sets, *Information and Control*, 1965, 8:338-53, doi:10.1016/S0019-9958(65)90241-X
24. Tatari F, Mohammad R. Akbarzadeh T, Sabahi A. Fuzzy probabilistic multi-agent system for breast cancer risk assessment and insurance premium assignment. *J. of Biomedical informatics*, December 2012, 45(6): 1021-34, DOI: 10.1016/j.jbi.2012.05.004
25. Aamodt A. A Knowledge-Intensive, Integrated Approach to Problem Solving and Sustained Learning. PhD thesis, Norwegian Institute of Technology, Department of Computer Science, Trondheim, May 1991.
26. I. Bichindaritz, C Marling, Case-based reasoning in the health sciences: What's next? *Artificial Intelligence in Medicine*, 36 (2), 127-35. PMID: 16459064
27. Abdeldjalil KHELASSI, Mohamed Amin Chick. Fuzzy knowledge-intensive case based classification for the detection of abnormal cardiac beats. *Electron. Physician*, 2012;4(2): 565-571. Doi: 10.14661/2012.565-71
28. McSherry D. A lazy learning approach to explaining Case BAED Reasoning Solutions. *Proceedings of the 20th International Conference on Case-Based Reasoning (ICCBR 2012)*, volume 7466 of Lecture Notes in Computer Science, Lyon, France, Springer , September 2012: 241-254
29. William J. Clancey. The epistemology of a rule-based expert system: A framework for explanation. *Artificial Intelligence*, 20(3): 215-251, 1983. Doi: 10.1016/0004-3702(83)90008-5
30. Swartout W. R. XPLAIN: A system for creating and explaining expert consulting programs. *Artificial Intelligence*, 21(3):285-325, 1983. doi:10.1016/S0004-3702(83)80014-9

31. Swartout W. R, Stephen W. Smoliar. On making expert systems more like experts. *Expert Systems*, 1987, 4(3):196–207. DOI: 10.1111/j.1468-0394.1987.tb00143.x
32. Öztürk P, Aamodt A. A context model for knowledge intensive case-based reasoning. *International Journal of Human Computer Studies*, 1998, 48(3): 331–55. doi:10.1006/ijhc.1997.0174
33. Kofod-Petersen A, Aamodt A. Contextualised ambient intelligence through case-based reasoning. *Proceedings of the Eighth European Conference on Case-Based Reasoning (ECCBR 2006)*, volume 4106 of *Lecture Notes in Computer Science*, Berlin, Springer, September 2006: 211–225.
34. Street N, Kyoung-Mi L. web site of XCYT project and breast cancer cytology images: the project available from: <http://dollar.biz.uiowa.edu/xcyt/> the Images available from: <http://dollar.biz.uiowa.edu/~street/xcyt/images> accessed on September.21.2014
35. Jeremy Thomas, web site : Breast Pathology on the Web - a guide to breast pathology for trainees and others available from: <http://www.breastpathology.info> accessed on September.21.2014