Machine Learning Applications in Cancer Informatics

Abdel-Badeeh M. Salem

BioMedical Informatics and Knowledge Engineering Research Lab.,
Faculty of Computer and Information Sciences,
Ain Shams University, Abbasia, Cairo-Egypt
abmsalem@yahoo.com, asalem@eun.eg
http://elearn.shams.edu.eg/absalem/27.htm

Abstract. Cancer informatics is a multidisciplinary field of research. It includes oncology, pathology, radiology, computational biology, physical chemistry, computer science, information systems, biostatistics, machine learning, artificial intelligence (AI), data mining and many others. Machine learning (ML) offers potentially powerful tools, intelligent methods, and algorithms that can help in solving many medical and biological problems. The variety of ML algorithms enable the design of a robust techniques and new methodologies for managing, representing, accumulating, changing ,discovering ,and updating knowledge in cancer-based systems. Moreover it supports learning and understanding the mechanisms that will help oncologists, radiologists and pathologists to induce knowledge from cancer information databases. This paper presents the research results of the author and his colleagues that have been carried out in recent years on using machine learning in cancer informatics. In addition the talk discusses several directions for future research.

Keywords: Machine Learning, Cancer Informatics, Case-Based Reasoning (CBR), Ontological Engineering, Genetic Algorithms, Medical Knowledge Management.

1 Introduction

Cancer is a group of more than 200 different diseases; it occurs when cells become abnormal and keep dividing and forming either benign or malignant tumors. Cancer has initial signs or symptoms if any is observed, the patient should perform complete blood count and other clinical examinations. Then to specify cancer type, patient need to perform special lab-tests. Benign tumors can usually be removed and do not spread to other parts of the body. Malignant tumors, on the other hand, grow aggressively and invade other tissues of the body, allowing entry of tumor cells into the bloodstream or lymphatic system which spread the tumor to other sites in the body. This process of spread is termed metastasis; the areas of tumor growth at these distant sites are called metastasis.

From the informatics point of view, breast cancer classification, diagnosis and prediction techniques have been a widely researched area in the past decade in the world of medical informatics. Several articles have been published which tries to classify breast cancer data sets using various techniques such as fuzzy logic, support vector

R. Kountchev & B. Iantovics (Eds.): *Adv. in Intell. Anal. of Med. Data & Decis.*, SCI 473, pp. 1–14.
DOI: 10.1007/978-3-319-00029-9_1

machines, Bayesian classifiers, decision trees, neural networks, and case-based reasoning[1,2,3,4]. CBR researchers agree that the best way to satisfy the increasing demand of developing CBR applications is by development of CBR-based frameworks. CBR frameworks are provided by research groups to overcome the problem of disturbing the concentration of the researchers in different domains with programming of AI applications, and help researches focusing on building components that directly address key concepts on a provided infrastructure that facilitates application to large scale projects [4,5,6].

This paper discusses the application of machine learning techniques in cancer informatics. The paper describes the following applications; (a) Case-Based Reasoning for diagnosis of cancer diseases, (b) Ontological engineering for lung and breast cancer knowledge management, (c) data mining for assessing diagnosis of breast cancer, and(d) genetic algorithms based classifier for breast cancer disease.

2 Case-Based Reasoning Approach for Diagnosis of Cancer Diseases

This section presents a CBR-based expert system prototype for diagnosis of cancer diseases developed at Medical Informatics Group at Ain Shams University, Cairo [7, 8]. The main purpose of the system is to serve as doctor diagnostic assistant. The system provides recommendation for controlling pain and providing symptom relief in advanced cancer. It can be used as a tool to aid and hopefully improve the quality of care given for those suffering intractable pain. The system is very useful in the management of the problem, and its task to aid the young physicians to check their diagnosis.

2.1 Case-Based Reasoning Methodology

CBR receives increasing attention within the AI community [4]. CBR is an analogical reasoning method provides both a methodology for problem solving and a cognitive model of people. CBR means reasoning from experiences or "old cases" in an effort to solve problems, critique solutions, and explain anomalous situations. It is consistent with much that psychologist have observed in the natural problem solving that people do. People tend to be comfortable using CBR methodology for decision making, in dynamically changing situations and other situations were much is unknown and when solutions are not clear. CBR refers to a number of concepts and techniques that can be used to record and index cases and then search them to identify the ones that might be useful in solving new cases when they are presented. In addition, there are techniques that can be used to modify earlier cases to better match new cases and other techniques to synthesize new cases when they are needed.

Following to the CBR methodology, the algorithm of interpreting and assimilating a new case can be summarized in the following processes [4]:

1. Assign Indexes: the features of the new case are assigned as indexes characterizing the event.

2. Retrieve: the indexes are used to retrieve a similar past case from the case memory.
3. Modify: the old solution is modified to conform to the new situation, resulting in a proposed solution.
4. Test: the proposed solution is tried out. It either succeeds or fails.
5. Assign and Store: If the solution succeeds, then assign indexes and stores a working solution. The successful plan is then incorporated into the case memory.
6. Explain, Repair and Test: If the solution fails, then explain the failure, repair the working solution, and test again. The explanation process identifies the source of the problem. The predictive features of the problem are incorporated into the indexing rules knowledge structure to anticipate this problem in the future. The failed plan is repaired to fix the problem, and the revised solution is then tested.

To perform the CBR process, the following knowledge structures (KSs) are very essential:

1. Indexing Rules KS: These rules identify the predictive features in the input that provides appropriate indexes into the case memory.
2. Case Memory KS: Case memory is the episodic memory, which comprises of database of experience.
3. Similarity Rules KS: If more than one case is retrieved form episodic memory, the similarity rules (or metrics) can be used to decide which case is more like the current situation.
4. Modification Rules KS: If no old case is going to be an exact match for a new situation, the old case must be modified to fit. We require knowledge about what kinds of factors can be changed and how to change them.
5. Repair Rules KS: Once we identify and explain an expectation failure, we must try to alter our plan to fit the new situation. Again we have rules for what kinds of changes are permissible.

2.2 Technical Aspects of the Case-Based Expert System

The system consists of three main modules; user interface, case base reasoning module and computational module all are interacted with the main environment of cancer diseases [7, 8]. The user is cancer expert doctor, the interaction is through menus and dialogues that simulate the patient text sheet contain symptoms and lab examinations. Computational module uses rule-based inference to give diagnostic decision and new case is stored in case library. Patient cases are retrieved in dialogue with similarity matches using the nearest neighbor matching technique. Frame knowledge representation technique is used for patient case indexing, storage and retrieval.

Fig. 1 shows the architecture of the CBR-based system. The system's knowledge base is diverse and linked through a number of indices, frames and relationships. The bulk of this knowledge consists of actual case histories and includes 70 cancer patient cases; some are real Egyptian cases and some from virtual hospitals on the internet. Fig.2 shows example of an Egyptian liver cancer case description of old women.

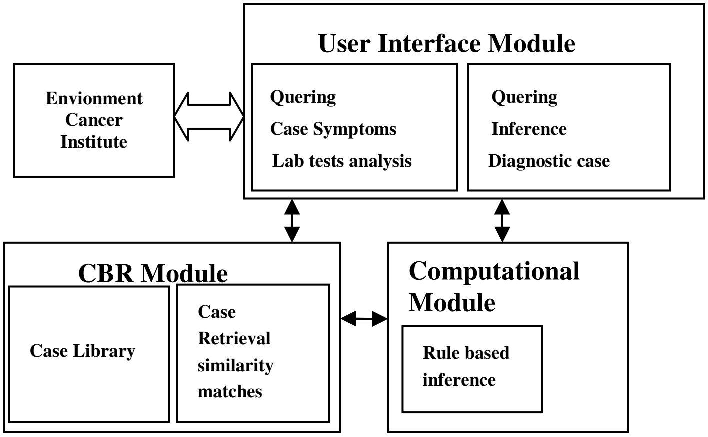


Fig. 1. Architecture of the CBR-based system for cancer diagnosis

Patient: 65-years old female not working, with nausea and vomiting.

Medical History: cancer head of pancreas

Physical Exam: tender hepatomgaly liver, large amount of inflammatory about 3 liters, multiple liver pyogenic abscesses and large pancreatic head mass.

Laboratory Findings: total bilrubin 1.3 mg/dl, direct bilrubin 0.4 mg/dl, sgot (ast) 28 IU/L, sgpt (alt) 26 IU/L.

Fig. 2. Egyptian liver cancer case

3 Ontological Engineering Approach for Cancer Knowledge Management

The term "ontology" is inherited from philosophy, in which it is a branch of metaphysics concerned with the nature of being. It began being used in Artificial Intelligence (AI) in the 1980s, and is now frequently used by computing and information science audiences. In AI ontology can be defined simply as a common vocabulary for describing a domain that can be used by humans as well as computer applications. Ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). Ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins. Ontologies have a range of potential benefits and applications in medical domain including the sharing of medical information across medical systems, enabling reuse of medical knowledge and providing intelligent and personalized researcher support [9, 10].

This section presents the developed two web-based ontologies for lung and breast cancers [11, 12]. The two ontologies; were built using the Protégé-OWL editing environment. The main goals behind building these ontologies are to allow finding and locating information about lung and breast cancers needed for interested users and domain experts, integrating information about lung and breast cancers to be accessed in an easy manner and providing the availability and accessibility of lung and breast cancers knowledge over the web.

3.1 Web-Based Lung Cancer Ontology

The lung cancer ontology was encoded in OWL-DL format using the Protégé-OWL editing environment [11]. The knowledge concerning breast cancer is collected from many sources including: Cancerbackup , American Cancer Society, American Lung Association. In this ontology (Figure 3) we have the following four main super classes ;

- **People**, which has the sub classes; male and female.
- **Medical_Interventions** which has sub classes; *Treatment, Staging* and *Diagnosis.*
- **Disease** which has sub class *cancer* which has sub class; *lung_cancers.*
- **Disease_attributes which has sub classes; Causes, Disease_stage, Pathological_category, Staging_systems and Symptoms**

The lung cancers are described in terms of its symptoms, causes, stages, pathological category, diagnosis and treatment. In this context, we described causes, stages, pathological category and symptoms as disease attributes. While diagnosis (including determining the stage of the lung cancer) and treatment are described as medical interventions.

3.2 Web-Based Breast Cancer Ontology

Breast cancer ontology was encoded in OWL-DL format using the Protégé-OWL editing environment [12]. The knowledge concerning breast cancer is collected from many sources including: MedicineNet ,The World Health Organization (WHO) , The breastcancer.org , The ehealthMD and The National Comprehensive Cancer Network (NCCN) .

In this ontology we have two main super classes

- MedicalThings which has sub classes Diseases, Medical_Interventions, Pathological_Category, References.
- People which has the sub classes; men and women.

The class Diseases has a subclass Cancers which has a subclass Breast_Cancer. The class Medical_Interventions has subclasses Diagnostic and Therapeutic. The class References has subclasses Causes, Disease_Stage, Staging, Symptoms and TNM_Stage. Some of the subclasses motioned above may has its own sub classes as shown in figure 4. These entire sub classes are related with is-a link. The breast cancers are described in terms of its symptoms, causes, stages, pathological category, diagnosis and treatment. In this context, we described causes, stages, and symptoms as references. While diagnosis and treatment are described as medical interventions as shown on figure 4.

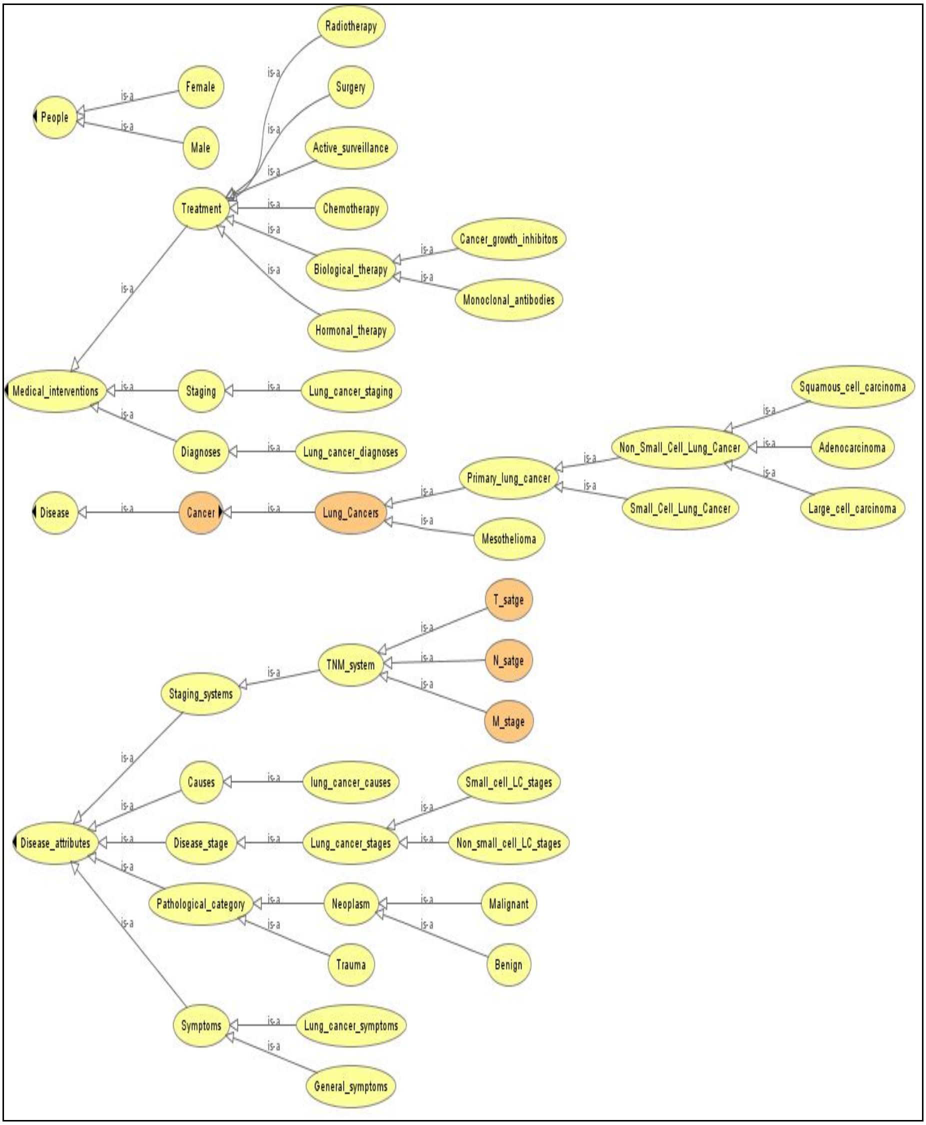


Fig. 3. The Lung Cancer Class Hierarchy

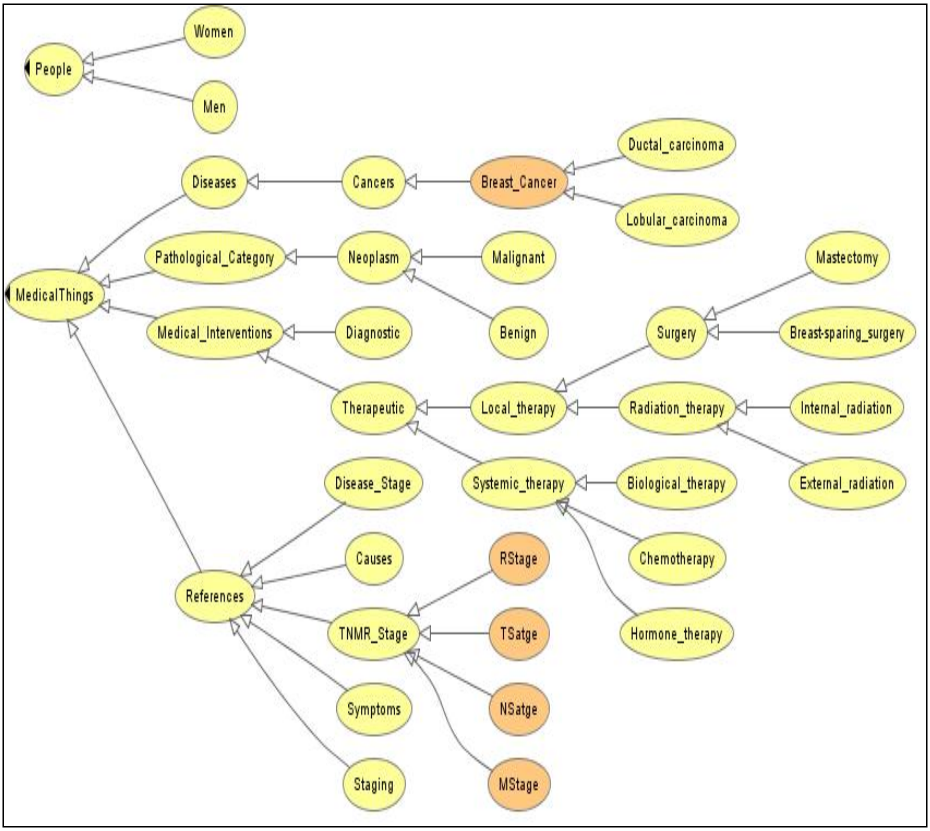


Fig. 4. The Developed Breast Cancer Ontology

In the breast cancer ontology, we described the diagnosis of the breast cancer as instances of the class Diagnostic. Also causes, stages, staging (how to determine the stage of the cancer) and symptoms of the breast cancer are described as instances. In the breast cancer ontology, the classes M_stage, N_stage and T_stage are defined as enumerated classes. Each one of these classes is described in terms of its instances.

4 Breast Cancer Classification Using a Combination of Ontology and Case-Based Reasoning

In our research [13, 14], we developed a new technique in the field of breast cancer classification. It uses a combination of ontology and case-based reasoning methodologies. Two recent frameworks are examined building the classifier. One is the open source jCOLIBRI system developed by GAIA group and provides a framework for building CBR systems. The other is the novel open source CBR tool, myCBR, developed at the German Research Center for Artificial Intelligence (DFKI). The objective of this classifier is to classify the patient based on his/her electronic record whether he/she is benign or malignant.

4.1 Breast Cancer Domain

Breast cancer is the form of cancer that either originates in the breast or is primarily present in the breast cells. The disease occurs mostly in women but a small population of men is also affected by it. Early detection of breast cancer saves many thousands of lives each year. Many more could be saved if the patients are offered accurate, timely analysis of their particular type of cancer and the available treatment options. Since the breast tumors whether malignant or benign share structural similarities, it becomes an extremely tedious and time consuming task to manually differentiate them. As seen in Figure 5 there is no visually significant difference between the fine needle biopsy image of the malignant and benign tumor for an untrained eye.

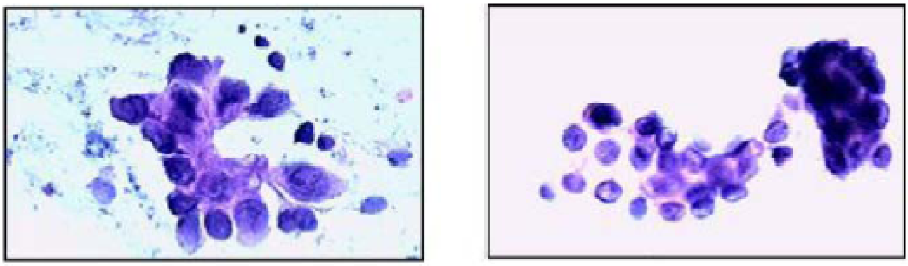


Fig. 5. Fine needle biopsies of breast. Malignant (left) and Benign (right) [15].

Accurate classification is very important as the potency of the cytotoxic drugs administered during the treatment can be life threatening or may develop into another cancer. Laboratory analysis or biopsies of the tumor is a manual, time consuming yet accurate system of prediction. It is however prone to human errors, creating a need for an automated system to provide a faster and more reliable method of diagnosis and prediction for the patients.

4.2 The Open Source jCOLIBRI Framework

jCOLIBRI framework is an evolution of the COLIBRI architecture [16], that consisted of a library of problem solving methods (PSMs) for solving the tasks of a knowledge-intensive CBR system along with ontology, CBROnto [8], with common CBR terminology. COLIBRI was prototyped in LISP using LOOM as knowledge representation technology. The design of the *jCOLIBRI* framework comprises a hierarchy of Java classes plus a number of XML files. The framework is organized around the following elements:

***Tasks and Methods*:** The tasks supported by the framework and the methods that solve them are all stored in a set of XML files.

***Case Base*:** Different connectors are defined to support several types of case determination, from the file system to a database.

***Cases*:** A number of interfaces and classes are included in the framework to provide an abstract representation of cases that support any type of actual case structure.

***Problem Solving Methods*:** The actual code that supports the methods included in the framework.

4.3 The Open Source myCBR Framework

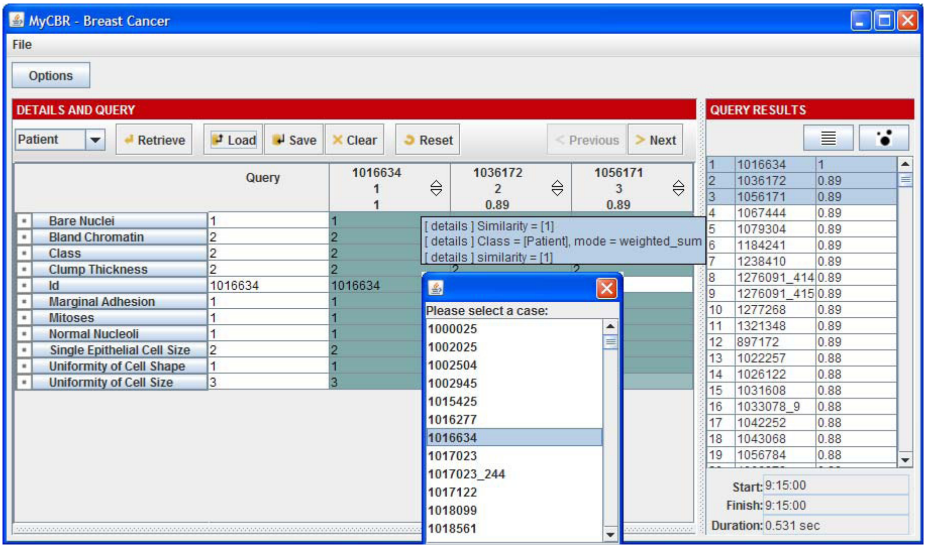
myCBR is an open-source plug-in for the open-source ontology editor *Protégé*. *Protégé* is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development [17]. *Protégé* allows defining classes and attributes in an object-oriented way. Furthermore, it manages instances of these classes, which *myCBR* interprets as cases. So the handling of vocabulary and case base is already provided by *Protégé*. The *myCBR* plug-in provides several editors to define similarity measures for an ontology and a retrieval interface for testing. As the main goal of *myCBR* is to minimize the effort for building CBR applications that require knowledge-intensive similarity measures, *myCBR* provides comfortable GUIs for modeling various kinds of attribute specific similarity measures and for evaluating the resulting retrieval quality. In order to reduce also the effort of the preceding step of defining an appropriate case representation, it includes tools for generating the case representation automatically from existing raw data. The novice as well as the expert knowledge engineer are supported during the development of a *myCBR* project through intelligent support approaches and advanced GUI functionality. Knowledge engineer can go through the following four steps to develop a CBR System; (a) Generation of case representations, (b) Modeling similarity measures, (c) Testing of retrieval functionality, and (d) Implementation of a stand-alone application.

4.4 Experimental Results of Breast Cancer Classifications

In our study, the two mentioned CBR frameworks are tested by developing a CBR application that classifies the condition of the breast cancer tumor whether it is benign or malignant. Wisconsin breast cancer data set was used for building the case-bases. It is obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg [14]. Samples inside the data set arrive periodically as Dr. Wolberg reports his clinical cases. The number of instances inside the dataset is 699 (as of 15 July 1992). Each record contains ten attributes plus the class attribute.

During the implantation of the breast cancer diagnostic application using jCOLIBRI we found that jCOLIBRI is user-friendly and efficient to develop a quick application. The classifier was successful in classification of the selected data set. Results given by the testing of the developed application are quite satisfactory. The system can precisely retrieve most similar cases although the input query has a missing attribute value. Deployment of a standalone application is completely failed.

During the implantation of the breast cancer classifier using myCBR we noticed that myCBR is a really a tool for rapid prototyping of a new CBR application. In seconds, users may have a running standalone CBR application by using the CSV importing feature. myCBR is intelligent enough to build the case structure and the case base by parsing the provided CSV file. myCBR avoids reinventing the wheel by making the development of a new CBR application done inside Protégé. The classifier was successful in classification of the selected data set.



	Query	1016634 1 1
Bare Nuclei	1	1
Bland Chromatin	2	2
Class	2	2
Clump Thickness	2	2
Id	1016634	1016634
Marginal Adhesion	1	1
Mitoses	1	1
Normal Nucleoli	1	1
Single Epithelial Cell Size	2	2
Uniformity of Cell Shape	1	1
Uniformity of Cell Size	3	3

Other columns: 1036172, 2, 0.89; 1056171, 3, 0.89

[details] Similarity = [1]
[details] Class = [Patient], mode = weighted_sum
[details] similarity = [1]

Please select a case:
1000025
1002025
1002504
1002945
1015425
1016277
1016634
1017023
1017023_244
1017122
1018099
1018561

1	1016634	1
2	1036172	0.89
3	1056171	0.89
4	1067444	0.89
5	1079304	0.89
6	1184241	0.89
7	1238410	0.89
8	1276091_414	0.89
9	1276091_415	0.89
10	1277268	0.89
11	1321348	0.89
12	897172	0.89
13	1022257	0.88
14	1026122	0.88
15	1031608	0.88
16	1033078_9	0.88
17	1042252	0.88
18	1043068	0.88
19	1056784	0.88

Start: 9:15:00
Finish: 9:15:00
Duration: 0.531 sec

Fig. 6. Breast Cancer as a Stand-Alone Application

5 Genetic Algorithms Approach for Data Mining Breast Cancer Classification

5.1 Genetic Algorithms Approach

Genetic Algorithms (GA) provide an approach to learning that based loosely on simulated evolution. The GA methodology hinges on a population of potential solutions, and as such exploits the mechanisms of natural selection well known in evolution. Rather than searching from general to specific hypothesis or from simple to complex GA generates successive hypotheses by repeatedly mutating and recombining parts of the best currently known hypotheses. The GA algorithm operates by iteratively updating a poll of hypotheses (population). One each iteration, old members of the population are evaluated according a fitness function. A new generation is then generated by probabilistically selecting the fittest individuals form the current population. Some of these selected individuals are carried forward into the next generation population others are used as the bases for creating new off springs individuals by applying genetic operations such as crossover and mutation.

This section presents an application of genetic algorithms approach for data mining classification task [18,19]. Actually two C4.5 based classifiers were developed, the first classifier; RFC4.5 uses the RainForest framework approach while the second; GARFC4.5 is a hybrid classifier uses Genetic Algorithm. The role of C4.5 classifier is to construct a simple decision tree. The role of RainForest is to keep the scalability aspects in constructing the classifier. The role of Genetic algorithms is working as online or dynamic training. In what follow a brief description of both classifiers is given [17].

5.2 Breast Cancer Database

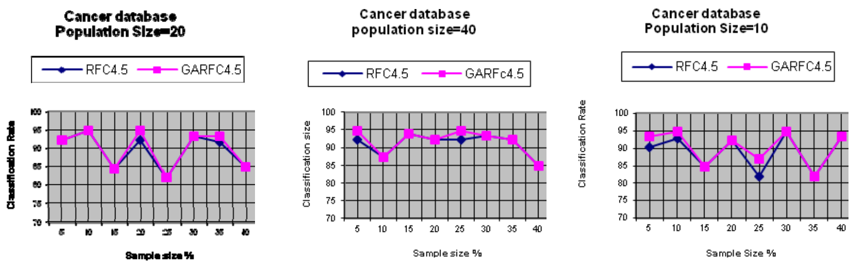
The Breast cancer database contains special laboratory examinations for breast cancer diseases. The problem is to predict whether a tissue sample taken from a patient's breast is malignant or benign. The original data tables have been imported and converted into access format. The database includes a number of instances (about 699 patients). The 16 instances with missing attribute values are removed from the database, leaving 683 instances. The main goal was to discover some sensitive and specific patterns as well as achieving a high prediction rate for breast cancer disease. The data record has 11 numerical attributes, attributes namely Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses and Class. These attributes are used to represent patient instances. The target attribute is the class attribute that has two possible values: benign or malignant. Class distribution: Benign: 458 (65.5%) and Malignant: 241 (34.5%).

5.3 Experimental Results of GAFRC4.5 Classifier

The GARFC4.5 hybrid classifier has two main components. The first one, RFC4.5 classifier, combines the previous described advantage of RainForest with the traditional C4.5 algorithm. RFC4.5 takes a set of training records from the database and produces a decision tree according to the randomly selected set of records. The second component, GA, performs the following operations: (a) taking, as input, the generated trees by RFC4.5, and (b) performs the GA operations. Crossover takes two generated trees from two different training data sets by RFC4.5, exchange two or more sub-tree, and produce another two different trees with different fitness function value (classification accuracy). Mutation operation takes one generated tree by first component RFC4.5, exchange two or more sub-tree of the same given tree or convert one or more subtree to a leaf then produce a new tree with different fitness function value. The GA operation proceed until the best tree is obtained from all the decision trees generated from randomly selected set of training data by RFC4.5 classifier component [20].

In our GARFC4.5 experiments, the probability of crossover is set to 0.9 and mutation rate to 0.02 after trying some other values and discovering that the previous values are the best (as known this value are problem specific). Twenty generation as the stopping criterion as noticed that the fitness of the individuals doesn't improve anymore. The sampling percentages vary from 5% to 40% of the original dataset and the population size from 10 to 50. The database was divided randomly into 70% for training and 30% for testing.

Figure 7 shows the computational results for our genetic algorithms based classifier from different combination of population size and the sampling percentages comparative to RFC4.5 classifier (classifier without the genetic algorithms component). The result curves at figure 7 show that the genetic algorithms based classifier vibrates between high and low classification accuracy and this is due to the numerical nature of the breast cancer database attributes. The results show that GARFC4.5 was not good with numerical nature databases. The results showed that genetic algorithms approach improves the classification accuracy over the traditional decision tree approach. The classification rates are 94% and 81% for our genetic classifier and decision tree algorithm respectively. Moreover the genetic classifier has the privilege of online learning over traditional decision tree C4.5 algorithm.



(a) (b) (c)

Fig. 7. Computational results of the classifiers GARFC4.5 and RFC4.5 for different population size

6 Conclusions

1. Machine Learning algorithms offer intelligent computational methods for accumulating, changing and updating knowledge in intelligent systems, and in particular learning mechanisms that will help us to induce knowledge from information or data.
2. CBR processes and techniques are very useful to develop CBR-based applications in medical domains but extensive effort is required to enhance their learning curve, usability and understandability.
3. Ontology-based CBR frameworks are very useful to develop robust CBR-based breast cancer classifiers that can play a very important role to help for early detecting the disease and hence right medications can be used to save lives
4. Ontological engineering approach is an effective methodology to manage and represent cancer knowledge.

5. Cancer ontologies are very useful in medical knowledge-based systems to facilitate knowledge sharing, refine, search and reuse.
6. The application of genetic algorithms approach improves the cancer classification accuracy over the traditional decision tree approach.

References

1. Cortes, C., Vapnik, V.: Support Vector Networks. Machine Learning 20, 273–297 (1995)
2. Quinlan, J.R.: C4.5: Programming for Machine Learning. Morgan Kaufman Publishers (1993)
3. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Welsey Publishing Company, Reading (1989)
4. Kolonder, J.: Case-Based Reasoning. Morgan Kaufmann (1993)
5. Abdrabou, E.A., Salem, M.A.B.: Case-Based Reasoning Tools from Shells to Object-Oriented Frameworks. Advanced Studies in Software and Knowledge Engineering-Supplement to the Intern. J. “Information Technologies & Knowledge”, 37–44 (2008); Markov, K., Ivanova, K., Mitov, I. (eds.). Institute of Information Theories and Applications FOI ITHEA, Sofia
6. Abdrabou, E.A.M., Salem, A.B.: eZ-CBR: A Generic Tool for Case-Based Reasoning. In: Proc. of 2nd International Conf. on e-Health and Bioengineering - EHB 2009, Iaşi-Constanța, Romania (2009)
7. Salem, A.B.M., Roushdy, M., El-Bagoury, B.M.: An Expert System for Diagnosis of Cancer Diseases. In: Proc. of the 7th Intern. Conf. on Soft Computing, MENDEL, pp. 300–305 (2001)
8. Salem, A.-B.M., Nagaty, K.A., El-Bagoury, B.M.: A Hybrid Case-Based Adaptation Model For Thyroid Cancer Diagnosis. In: Proc. of 5th Int. Conf. on Enterprise Information Systems, ICEIS 2003, Angres, France, pp. 58–65 (2003)
9. Fernández-López, M., Gómez-Pérez, A.: Deliverable 1.4: A survey on methodologies for developing, maintaining, evaluating and reengineering ontologies. Part of a Research Project Funded by the IST Programme of the Commission of the European Communities as project number IST-2000-29243 (2002)
10. Salem, M.A.B., Alfonse, M.: Ontology versus Semantic Networks for Medical Knowledge Representation. In: Proceedings of 12th WSEAS CSCC Multiconference (Computers), Heraklion, Crete Island, Greece, pp. 769–774 (2008)
11. Salem, M.A.B., Alfonse, M.: Building Web-Based Lung Cancer Ontology. In: 1st National Symposium on e-Health and Bioengineering, EHB 2007, Iaşi, Romania, pp. 177–182 (2007)
12. Salem, M.A.B., Alfonse, M.: Ontological Engineering Approach for Breast Cancer Knowledge Management. In: Proc. of Med-e-Tel the International eHealth, Telemedicine and Health ICT for Education, Networking and Business, Luxembourg, pp. 320–324 (2009)
13. Abdrabou, E.A., Salem, M.A.B.: Application of Case-Based Reasoning Frameworks in a Medical Classifier. In: Proc. of Fourth International Conference on Intelligent Computing and Information Systems (ICICIS-4), Cairo, EG, pp. 253–259 (2009)
14. Abdrabou, E.A., Salem, M.A.B.: A Breast Cancer Classifier based on a Combination of Case-Based Reasoning and Ontology Approach. In: Proc. of 2nd International Multi-Conference on Computer Science and Information Technology, IMCSIT 2010, Wisła, Poland (2010)

15. Sewak, M., Vaidya, P., Chan, C.C., Duan, Z.H.: SVM Approach to Breast Cancer Classification. IMSCCS 2, 32–37 (2007)
16. Recio-García, J.A., Díaz-Agudo, B., González-Calero, P.A.: jCOLIBRI2 Tutorial, Group of Artificial Intelligence Application (GAIA). University Complutense of Madrid. Document Version 1.2 (2008)
17. Stahl, A., Roth-Berghofer, T.R.: Rapid Prototyping of CBR Applications with the Open Source Tool myCBR. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 615–629. Springer, Heidelberg (2008)
18. Cios, K.J., Pedrycz, W., Swiniarski, R.W.: Data Mining Methods for Knowledge Discovery. Kluwer (1998)
19. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. ACM Computing Surveys 31(3), 264–323 (1999)
20. Salem, M.A.B., Abeer, M.M.: A Hybrid Genetic Algorithm-Decision Tree Classifier. In: Proc. of the 3rd Intern. Conf. on New Trends in Intelligent Information Processing and Web Mining, Zakopane, Poland, pp. 221–232 (2003)