Toward Translational Incremental Similarity-Based Reasoning in Breast Cancer Grading

Adina Eunice Tutac[a,d], Daniel Racoceanu[a,e], Wee-Kheng Leow[a,c] , Henning Müller[f,g] , Thomas Putti[b], Vladimir Cretu[d]

[a]Image Perception, Access & Language IPAL (UMI CNRS 2955, UJF, NUS, I2R) Singapore;
[b]National University Hospital, Singapore; [c]National University of Singapore;
[d]Politehnica University of Timisoara, Romania; [e]University of Besançon, France;
[f]Medical Informatics Service, University Hospital of Geneva, Switzerland;
[g]University of Applied Sciences, Western Switzerland, Sierre, Switzerland

ABSTRACT

One of the fundamental issues in bridging the gap between the proliferation of Content-Based Image Retrieval (CBIR) systems in the scientific literature and the deficiency of their usage in medical community is based on the characteristic of CBIR to access information by images or/and text only. Yet, the way physicians are reasoning about patients leads intuitively to a case representation. Hence, a proper solution to overcome this gap is to consider a CBIR approach inspired by Case-Based Reasoning (CBR), which naturally introduces medical knowledge structured by cases. Moreover, in a CBR system, the knowledge is incrementally added and learned. The purpose of this study is to initiate a translational solution from CBIR algorithms to clinical practice, using a CBIR/CBR hybrid approach. Therefore, we advance the idea of a translational incremental similarity-based reasoning (TISBR), using combined CBIR and CBR characteristics: incremental learning of medical knowledge, medical case-based structure of the knowledge (CBR), image usage to retrieve similar cases (CBIR), similarity concept (central for both paradigms). For this purpose, three major axes are explored: the indexing, the cases retrieval and the search refinement, applied to Breast Cancer Grading (BCG), a powerful breast cancer prognosis exam. The effectiveness of this strategy is currently evaluated over cases provided by the Pathology Department of Singapore National University Hospital, for the indexing. With its current accuracy, TISBR launches interesting perspectives for complex reasoning in future medical research, opening the way to a better knowledge traceability and a better acceptance rate of computer-aided diagnosis assistance among practitioners.

Keywords: content-based image retrieval, case-based reasoning, breast cancer grading, translational approach, similarity-based reasoning, incremental learning, histopathology

1. INTRODUCTION

This section introduces our motivation for such kind of study starting from a general presentation of CBIR and CBR approaches. It summarizes their main characteristics, problems and trends, based on the existing work.

Content-Based Image Retrieval is generally seen as a technology using content-similarity-based methods to solve problems, particularly, to access images from image database by visual content, according to the users' interest[1, 2]. In the first approaches, CBIR consisted of two main phases: images indexing and similar images retrieval with a given query - typically based on visual similarity (Query-By-Visual-Example). More recently, the need of relevance feedback has been perceived as an integrated part of the CBIR demarche, yet considered as a key-issue in CBIR[10]. Problems solving with respect to types of query, similarity computation, relevance of results retrieved and so forth, are still open questions CBIR faces in its development, especially in the medical field[7]. Another evolution direction is based on the main functionality of CBIR – the visual content processing and analysis. New trends of using high level semantics concepts combined with the low level visual features for an efficient indexing and retrieval have been proposed in the literature[28, 29,30].We consider that such an approach has deep implications if used in medical applications, being able to provide more effective diagnosis and prognosis assistance. To illustrate the CBIR approach, a functional diagram is presented in Fig.1, adapted from Long[1].

From its early stages of theoretical foundations and the first systems development hitherto, CBIR has opened promising perspectives in the research area. The reasons are manifold but foremost, the idea of achieving valid retrieval results when given a query, challenged research community to define advanced indexing and retrieval techniques. Consequently, it enriched the core functionality of organizing increasing digitized data. Related to this, one of the main characteristics of CBIR is the *single-image-way of structuring information*. This characteristic becomes an issue when developing medical CBIR due to the different way patient information is usually structured: by cases. Thus, we introduce CBR as a solution to this problem, based on its resemblance with the medical reasoning and the information representation by cases.

Also designed for problem solving but from another perspective, Case-Based Reasoning has been proposed as a novel approach in Artificial Intelligence, particularly in Knowledge-Based systems[3]. Essentially, it is defined as the four Res cycle: Retrieve, Reuse, Revise and Retain as main principles, each of them having particular phases. The hallmark of CBR is working with structured information by cases and trying to retrieve the similar cases based on past experience (reasoning by remembering[4]) when given a new problem. Unlike CBIR, the latter goes beyond retrieval process, by using a suggested matched cases solution for the new case that will be then reused and tested for success. Furthermore, the solution may be revised if the retrieved case is not similar with the given problem, producing a new case to be retained in the case base. Various issues CBR faces in each phase of its cycle, along with their possible solutions are identified. For instance, weak domain knowledge and slow retrieval have been overcome by generating CBR prototypes[5]. Hybrid reasoning paradigm consisting of multi-modal approaches[6] (CBR, Rule-Based Reasoning, Model-Based Reasoning, etc) combined together proved to be more efficient in some particular situation of higher complexity. Fig.2. presents a clear description of how CBR functions.

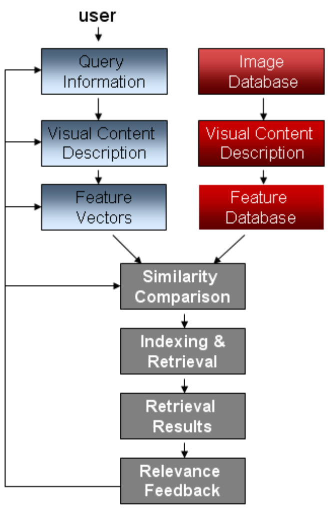


Fig. 1. CBIR functional diagram

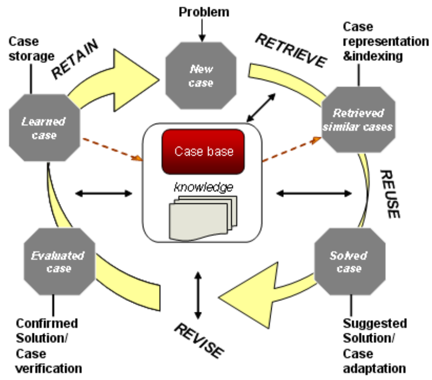


Fig. 2. CBR Res cycle

Apart from Content-Based Image Retrieval's and Case- Based Reasoning's individual promising applicability in medical communities, both face significant issues with respect to clinical practice[7, 8, 9], and inherent questions are raising above. Is it efficient enough to analyze only images, for identifying relevant features in a retrieval process? Is the CBIR process the end of a medical procedure in order to give a good diagnosis or prognosis? On the other hand, is textual description of a case representation accurate enough to provide a good diagnosis as a valid solution? Would a combined approach solve their limitations?

To answer these questions, our paper presents a comparative analysis of CBIR and CBR with the scope of identifying common characteristics and advantages that can be used to propose an integrated framework: *a translational incremental similarity-based reasoning (TISBR) for Breast Cancer Grading (BCG)*. Textual information describing symptoms and prognosis - as it is a commonly representation of a case in CBR - combined with image content analysis- typical for CBIR, as well as the concept of similarity, common for both , give us the insight for the hybrid approach. The paper is organized as follows: Section 2 introduces generic considerations on CBIR and CBR in terms of methodology/technology context and interconnected fields. Identifying similarities and differences between the two approaches at the indexing, retrieval and refinement levels, with the contrast further applied to medical applications is the objective of section 3. An illustration of the hybrid approach is proposed in section 4 with respect to BCG application. Conclusions and future trends related to research and clinical practice are discussed in section 5.

2. GENERAL CONSIDERATIONS

In order to define our analysis strategy, some prerequisites need to be considered. Firstly, we make a distinction between methodology and technology applied to CBIR and CBR. Secondly, we position CBIR and CBR in their context, showing the interconnection with other fields.

2.1 Methodology versus technology

From the concept point of view, CBIR is often referred as a technology[1, 2, 10], which uses various techniques to solve specific problems. Similarly, Kolodner and Richer[11, 12] consider CBR as a technology, whilst Watson[13] emphasizes that CBR is an organized *set of principles* which guide action in problem solving matters rather than an isolated technique, limited to handle only very specific tasks. Hence, it verifies the definition of a methodology given by Checkland[14]. The reason for viewing CBR as a methodology has several implications. On one hand, since it doesn't have its own technology, it can use *any* technology that applies CBR principles. On the other hand, we can build *hybrid systems*, in terms of *hybrid methodologies* and not *hybrid technologies*. Furthermore, seeing CBR as a methodology supports the idea of future research, which is important, since many technologies for each CBR phase are commonly used and some - already mature. We adopt the same approach as Watson's and moreover, we propel CBIR at the same level, of methodology. We envisage that CBIR made significant development in the recent years in terms of concepts, techniques and application domain. In our opinion, CBIR of today doesn't only organize digital data, as it was the main objective in its early years due to the semantic web development. In this setting, we consider that the principles of CBR can also define CBIR as a methodology, except the last principle (Retain). It is not necessary to have all four principles of CBR to see CBIR a methodology; the key idea is to *define CBIR in terms of basic principles and use any techniques in line with its principle*. Our general paradigm is depicted by Fig. 3. The reasons and the implications for CBIR and CBR technology versus methodology are detailed in Table. 1.

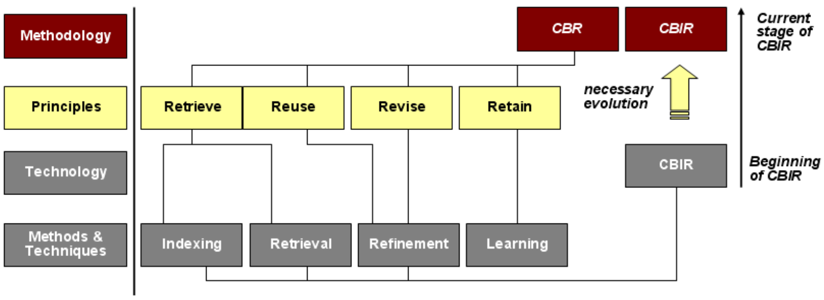


Fig. 3. CBIR and CBR. Methodology versus Technology

Table 1. Reasons and implications for methodology/technology in CBIR and CBR

Methodology versus Technology	Definition	Reason	Implication
Content- Based Image Retrieval	Technology [1,2]	set of *methods* to solve problems[10]	manifold applications
	Methodology	continuous development	knowledge-based systems flexibility
Case-Based Reasoning	Technology [11]	AI Technology description [11]	task limitation research limitation
	Methodology[13]	set of *principles* to solve problems[15]	can use any technology hybrid systems future research

To illustrate how CBIR follows the principles of CBR methodology, we construct the Res cycle based on the CBIR functional diagram (see Fig. 4).

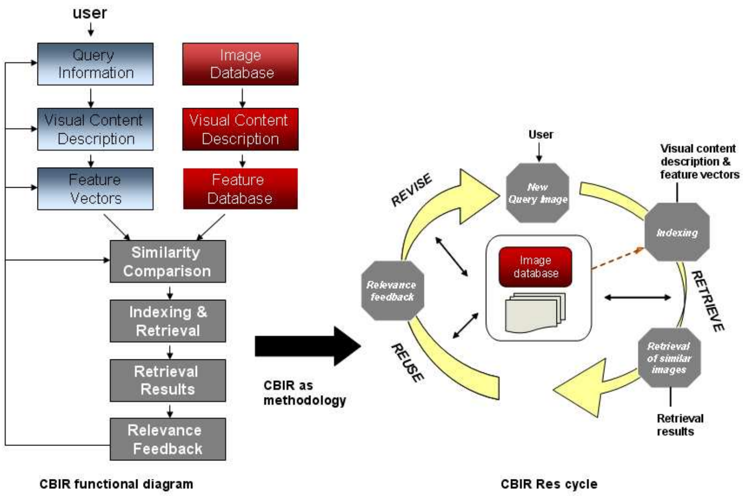


Fig. 4. CBIR methodology

Both CBIR and CBR are positioned at the confluence of some related areas, revealing their cross-discipline orientation. These crossroads also show the domains from where they emerged and extended afterwards. The great success of CBIR and CBR witnessed in the scientific literature emphasized a potentially significant impact for diagnosis and prognosis assistance in medical communities (see Fig. 5).

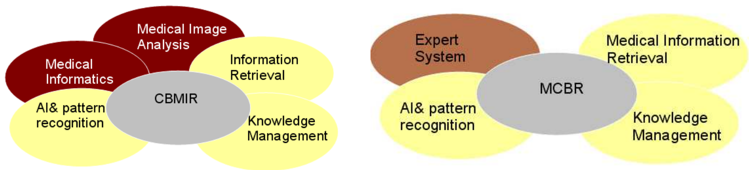


Fig. 5. Content-Based Medical Image Retrieval's (CBMIR) and Medical Case-Based Reasoning's (MCBR) related fields

3. COMPARATIVE ANALYSIS OF CBIR & CBR

Since we consider both CBIR and CBR as methodologies, the aim of this section is to present the similarities and the differences of three major technologies common for both CBIR and CBR: the indexing, the retrieval and the search refinement. Thus, to benefit of their merits and to overcome their week points for our hybrid framework proposal.

3.1 Indexing in CBIR & CBR

As shown in Table. 2, the core distinction is the principle-based orientation of CBR, unlike CBIR that was generally considered technique-based oriented until now. At the techniques level, there are some similar works proposed (learning-based in CBIR and inductive-learning in CBR) as well as some different approaches (no correspondence in CBIR for similarity & explanation-based technique found in CBR). *Semantic indexing* represents another category related to our analysis. An important issue identified here is the semantic gap[10], defined as the discrepancy between the low level visual features and the high-level semantic concepts. The essence is that this gap is biased by the versatility of visual image content. No difficulties are now encountered in extracting the objects from a raw image to label them [1,7,16]. Most common research into bridging the semantic gap is actually tackling the descriptors and object labels level. By which means, to associate meaning (to index images by semantic means) according to a specialized or generalized knowledge to some features retrieved from the image, is the cue of semantics[17]. The ultimate trend is to design and model ontologies[18] that can work at the semantic level with domain knowledge support, thus emphasizing the solution of apriori knowledge injection. Hence, the similarity with CBR design principle can be recognized here. Furthermore, we advance the idea that *ontologies* can also be *applied to CBR*, in the sense of structuring and representing knowledge contained in the cases. Ontologies are defined as a formal and explicit specification of an abstraction[19]. Similarly, CBR knowledge is explicitly stored in concrete cases, thus implying the fact that the case is not a general rule, but an instantiation of a formal specification. Without any doubt, semantic deixis conveys to finding the most accurate reality representation, in our approach particularly oriented to problem solving in medical applications.

Table 2. Indexing in CBIR & CBR

Indexing	**CBIR**[1,2,10]	**CBR** [3,12,13]
Indexing Principles	not until now	➢ predictive ➢ purpose oriented ➢ abstract/concrete enough
Similar Indexing Techniques	➢ feature-based ➢ structural features	➢ features &dimensions
	➢ salient-features	➢ difference-based
	➢ learning-based	➢ inductive learning
Different Indexing Techniques		➢ similarity & explanation-based
Characteristics	➢ feature indexing	➢ case indexing

3.2 Retrieval in CBIR & CBR

A similar analysis is applied to retrieval phase in CBIR and CBR. Once again, the need to have guiding principles in CBIR is pointed out, in Table 3. Modeling similarity is a central element to navigate through the space of possible solutions, in CBIR as well as in CBR, different approaches depending on image indexing/case representation. A general account of CBIR techniques carried out in the literature is given in the overview of Deb[31], extended by our table. Smeulders[10] presents a fine distinction of CBIR from the user's and system's point of view. Trends of combining elementary methods of retrieval with techniques of AI domain are discussed by Datta[2] (for instance a new Bayesian learning framework for automatic image annotation proposed by Shi[20]). A retrieval technique based on semantic example in CBIR is correspondently found in CBR as a knowledge-guided retrieval. One of the most common retrieval techniques successfully applied in both fields is the Nearest-Neighbor Retrieval. From the dissimilarity point of view, there are some techniques that are used either in one or in the other (validated retrieval in CBR, or Query-by-Keyword in CBIR). To this end, a classification of CBR into two categories is to be considered. Most CBR systems fall in the problem- solving category, which uses previous cases to only suggest the most likely solution to be applied to the new case. In contrast, interpretive CBR[21] are based on reference cases – previous cases, per se, to solve the new problem. In the same paper, a summary of soft CBR, implying combination with AI techniques is given to emphasize the idea of methods integration when it comes to evaluate the results from the reliability standpoint.

In essence, the process of retrieval highly depends on the indexing phase and the similarity computation step. The higher is the efficiency of indexing, the better the retrieval. This conclusion is also encountered with respect to medical applications.

Table 3. Retrieval in CBIR & CBR

Retrieval	CBIR[7,17]	CBR[3,12,13]
Retrieval Principles	not until now	➢ criteria selection &memory model
Similar Retrieval Techniques	➢ Query-By-Semantic-Example ➢ semantic retrieval	➢ knowledge-guided
	➢ Nearest-neighbor retrieval	➢ Nearest-neighbor retrieval
Different Retrieval Techniques	➢ Query-by-Keyword ➢ Query-by-Visual-Example(QBVE)	➢ inductive ➢ validated retrieval

3.3 Refinement in CBIR & CBR

Relevance feedback is defined as supervised active learning query modification/adaptation technique to improve the effectiveness of the information systems[2]. Likewise, case adaptation of CBR reuse principle, focuses on the proposed solution refinement of the similar cases extracted at retrieval time. Our rationale to consider relevance feedback and case adaptation as correspondent is due to their basic idea: refinement. The difference between the two approaches appears with respect to the *target of refinement*: in CBIR, *the query* is to be refined, while in CBR, *the solution* is refined. A relevance feedback step integrated in the CBIR process has several implications. Firstly, it's possible to create the link between the low level features and the high level concepts, capturing user and query specific semantics. Secondly, it refines the ranks accordingly to the query adaptation, thus improving system recall. There are however, some drawbacks: increasing of the user involvement (multiple rounds of feedback affect user's patience) or the fact that the changes won't be done at the low level features (they will remain the same). Also, human perception of image similarity is highly subjective, task-dependent and it's sometimes hard to establish why the obtained images are similar and how to exactly improve the performance of the system. Hence, it is necessary to have a relevance feedback in a CBIR system. Yet, many approaches provide no relevance feedback or a naïve feedback. Similar situation can be encountered in a CBR system, the so-called null adaptation technique. The analogy between the techniques used in CBIR and in CBR is described in Table.4.

To conclude this section, a strong point for the CBR regards its closed loop characteristic. CBR process doesn't stop at the retrieval phase as it is most likely to happen in a CBIR system. Thus the CBR is incrementally learning, the knowledge is continuous expanding, unlike CBIR where there is no such step beyond. Therefore, we consider CBIR as an open loop; even if there is a week relevance feedback, the process starts over again and so, there is no recording of how the problem was solved in the past. However, there are also some tradeoffs at the case storage level of CBR too. Storing too many cases may affect the speed of the execution and may introduce overfitting problems. To face this issue, Tadrat[32] proposed rough set theory (RST) combined with formal concept analysis (FCA). Nevertheless, the relevance feedback is also of current interest in medical applications.

Table 4. Refinement in CBIR & CBR

Refinement	CBIR [1,2,21]	CBR [3,12,13]
RF Principles	not until now	➢ structural & derivational
Similar Techniques	➢ no RF/naïve RF	➢ null adaptation
	➢ feature re-weighting	➢ parameter adjustment
	➢ specialized user-driven	➢ critic-based
	➢ memory-retrieval	➢ model-based
	➢ active-learning	➢ abstraction& respecialization/reinstantiation
Different Techniques	➢ probabilistic	➢ derivational replay ➢ case base substitution
Characteristics	➢ query refinement	➢ solution refinement

3.4 CBIR and CBR in medical applications

Within the last years, the number of digital images produced in the medical field is continuing to increase and hence a crucial need to design CBIR system to assist in the diagnosis, prognosis, has been emphasized. Deserno[8] presents a classification of CBIR medical applications, describing the most representative ones: PACS (Picture Achieving and Communication System) with the extended version cbPACS[33], IRMA[34] and MedGift[35], in an a posteriori approach. In the same time, CBR systems are highly promising to be used in clinical practice due mainly, to their cognitive adequateness characteristic and the explicit experience involved. Their baseline principle is more closed to the medical reasoning. Yet, issues such unreliability, adaptation or concentration on reference (a stored case-based is essential to consider the system as source of previous experiences) are still facts to be accounted for, in medical CBR. Similarly, the promising development on CBIR in the scientific community, however, did not accrue in the same manner in the medical field. One of the reasons for such a lack is attributed to various gaps of CBIR[2, 7,8,10]. Despite of some minor disagreements of the authors, a compilation of all gaps is given by Table. 5, with our own emphasis on *perception gap* instead of *aesthetic gap* proposed by Datta[2] (we think perception is more appropriate to be considered in the medical field, rather than an aesthetic gap).

Table. 5. CBMIR gaps

CBMIR gaps	Characteristics
Content	modeling & understanding image/information- real image/information
Features	computational numerical features- real image/information
Performance	application, integration, indexing, evaluation
Usability	query, feedback, refinement
Perception	visual information perception- real image/information perception
Sensory	information description – real image/information

Another aspect mentioned bellow, concerns the semantic gap. Müller[7] goes beyond Smeulders[10] by defining a generic content gap, which includes both semantic gap and the context gap. Table 6 shows CBIR and CBR main advantages and drawbacks with focus on medical area.

Table 6. CBIR and CBR in medical field

	Advantages	**Drawbacks**
Medical CBIR[7]	increasing rate of image production applications in diagnosis, teaching & research	relevance feedback, user interfaces performance
Medical CBR[9]	cognitive adequateness explicit experience duality of objective & subjective knowledge system integration application in diagnosis, teaching & research	adaptation unreliability concentration on reference

Nilsson[9] classifies various influential medical CBR systems of last years, from the purpose-oriented and construction-oriented perspectives. As a conclusion of this section, it is important to specify that, although challenging in theory, few system have been actually used in practice. Our desire to propose a translational approach came to meet this need.

4. TISBR FOR BREAST CANCER GRADING

Our desideratum is to make progress without losing what we already posses, with respect to CBIR and CBR. Table 7 contains the outline ideas of the main differences of CBIR and CBR. Our hybrid approach combines some of their characteristics within a single framework. The knowledge is structured by cases, as in CBR, thus creating the knowledge case-base in an incremental manner. From CBIR, image plays a prominent part in our approach, since medical assessment procedure hardly can work without it. Both CBIR and CBR are context dependent. It is yet difficult to overcome the context chasm, by propelling a generic solution, but some inner specific problem can be solved in a fusion paradigm. Hence the context modeling is applied in TISBR, related to BCG. This section will present in summary the BCG and the indexing axis, which represents the first level developed on TISBR. The retrieval and the refinement processes will be implemented in the near future. However, the results obtained for this first step are encouraging and show that the complete work will have strong impact on medical practice.

Table 7. CBIR versus CBR

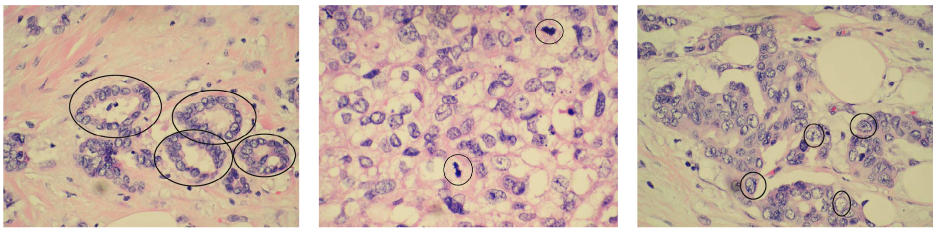
Content-Based Image Retrieval (CBIR)	Case-Based Reasoning (CBR)
➢ image based	➢ textual information based
➢ limited to retrieval phase	➢ integrated new case after adaptation & revise
➢ query expansion	➢ no query
➢ lack of knowledge injection	➢ a priori knowledge
➢ semantic web-based	➢ knowledge-based
➢ structured by image only	➢ structured by cases
➢ week learning/static database	➢ incrementally learning/dynamic database
➢ context-dependent	➢ context-modeling
➢ open loop : naïve relevance feedback	➢ close loop: case adaptation- case storage

4.1 Breast Cancer Grading

The reason of choosing breast cancer grading as our application is due to various facts: firstly, the statistics show a high global rate of breast cancer[1]. Therefore, we think there is a vital need to translate theoretical ideas into medical practice and CBIR and CBR together, are a possible solution. Secondly, breast cancer grading is a powerful prognosis exam worldwide and to our knowledge, not much work has been done with respect to it. The majority of proposals focus on diagnosis and not on prognosis. For instance, an example of a CBR diagnosis system is provided by Jaulent[23] and later extended to IDEM framework[24], while a CBIR system for BCG has not been implemented yet.

Histological grading is nowadays considered an exam of high relevance in breast cancer prognosis of modern pathology. Among the standard grading systems, Nottingham Grading System (NGS) represents the gold standard (ground-truth) due to its objectiveness for the three components of grading, described bellow. The scores for the three separate criteria (tubules, nuclei and mitoses) are added to give the overall grade[25].

- Tubule Formation score (TF) - are referred as the density of the Tubule Formations - white blobs (lumina) surrounded by a continuous string of cell nuclei.
- Mitosis Count (MC) score represent the number of Mitoses - diving cells nuclei. MC is assessed in the peripheral areas of the neoplasm and it's based on the number of mitoses per 10 High Power Field's (HPF's) – high resolution (usually 400x) frames obtained using microscopic acquisition.
- Nuclear Pleomorphism Score (NPS) - categorizes cells nuclei based on two main features: size and shape.



a *b* *c*

Fig. 6. NGS components: a) Tubule formation: lumina surrounded by string of cell nuclei, b) Mitosis: dividing cells nuclei, c) Big size / irregular shape nuclei-- NPS grade 3

We develop a *feature-based method* for a semi-*automated knowledge-guided semantic indexing process*, thus combining characteristics of both CBIR and CBR. To better understand our paradigm, Fig. 7 illustrates the main steps accomplished.

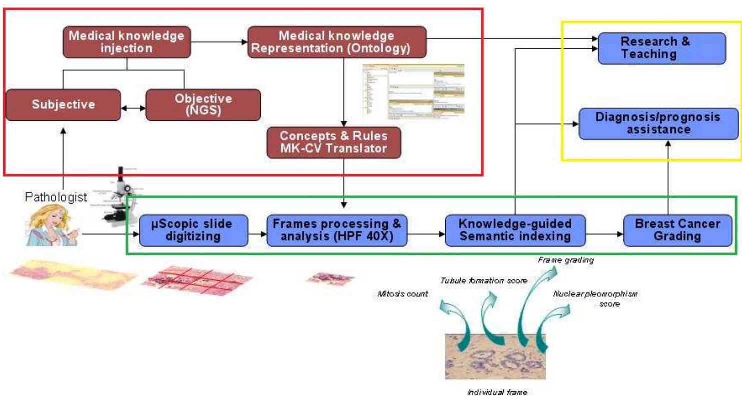


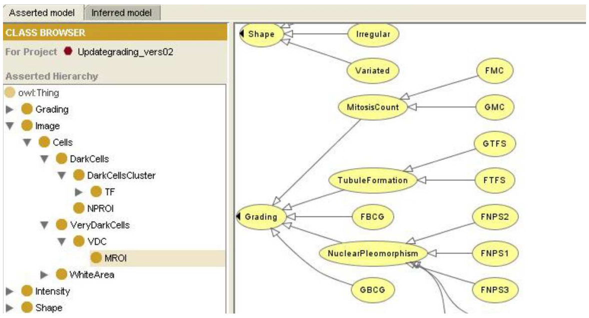
Fig. 7. Knowledge-guided semantic indexing workflow (one time processing/training – red/upper left frame, each case testing- green/bottom frame, services provided upon request – yellow/upper right frame).

[1]American Cancer Society, Breast Cancer Facts & Figures 2005-2006, http://www.cancer.org/downloads/STT/CAFF2005BrF.pdf

We will present a synthesis of the main steps as they are followed to reach the objective of TISBR first axis, namely the knowledge-guided semantic indexing of BCG.

- 1. Structuring BCG medical knowledge

For this aim, we structure the knowledge (specific to CBR) using the OWL-DL sublanguage, in a *BCG ontology* model validated under Pellet reasoner[27] (see Fig. 8).



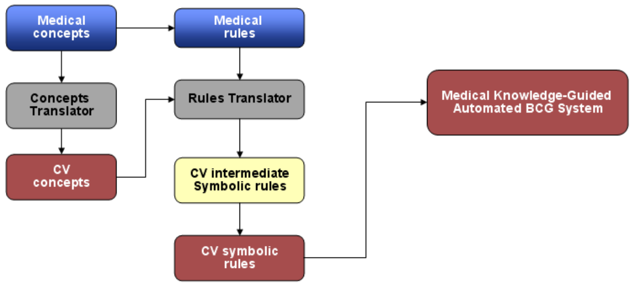


Fig. 8. OWLviz Breast Cancer Grading hierarchy

Fig. 9. Generic Translation Framework

- 2. Translating Medical Knowledge into Computer understandable terms

The key idea is to use the medical knowledge (MK-tubule formation, nuclear pleomorphism and mitosis count criteria for grading) in terms of concepts and rules, for creating a Computer Vision (CV) concepts and rules correspondence (see Fig. 9). A detailed presentation of the translation is given by our previous work[26].

- 3. Processing and analysis of BCG histopathology images

The structured information is then used for the image processing and analysis step (specific to CBIR). The *features vector*, similar with a case indexing in terms of CBR, contains the *features-symptoms* (TubuleFormationROI, MitosisROI, NucleiROI) and *their values*, *together with the prognosis*, *defined as the local and global grading* (per frame/per entire histopathology slide). Fig. 10 shows an example of two explicit cases, part of the BCG case-base in the CBR manner. The processing technique along with the semantic indexing details is also discussed in Tutac[27].

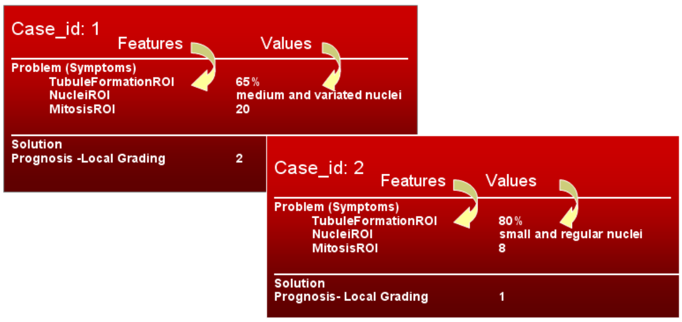


Fig. 10. BCG cases representation

- 4. Grading results

We evaluated six breast core-biopsy cases stained with H&E marker, consisting of 7000 frames scanned from the tumor tissue slides and obtained from the Pathology Department of National University Hospital of Singapore (NUH). The database is composed by two sets: 1400 frames used for the training algorithm phase and 5600 frames used for the testing and validation phase. The slides were scanned on a sequence of frames at 10X40 (400X) magnification with a 1080 x 1024 resolution. Based on previous steps, the grading is given for each frame and for the entire slide. Individually, the most accurate results were obtained for the mitosis count. Although, a 7, 33 % error was registered for the training dataset and 11% for the testing dataset in a local grading, for the global grading we obtained no computation errors. Compared with the manual grading given by the pathologists, we achieved an accuracy of 80% for the breast cancer global grading.

Table 8. BCG grading results

Manual Grading				Semi-Automated Grading					
Tubule score	Nuclear score	Mitosis count	BCG	Tubule score	Nuclear score	Mitosis count	BCG	Case ID	Data type
1	1	3	*1*	1	1	3	*1*	1000	Training Database (1400 images)
1	2	1	*1*	2	2	1	*1*	2000	
3	3	3	*3*	3	2	3	*3*	4895	
2	3	3	*3*	3	2	3	*3*	5020	Testing Database (5600 images)
3	3	3	*3*	3	2	3	*3*	5042	
3	2	1	*2*	3	2	1	*2*	5075	

Table 9. Local and global grading errors

Data base	Tubule score	Nuclear score	Mitosis count	Component scores error	Global BCG error
Training errors	11%	11%	0	7,33%	0%
Testing errors	11%	22%	0	11%	0%

5. CONCLUSIONS AND PERSPECTIVES

The scope of this paper is to firstly set up a theoretical foundation for this CBIR and CBR generic and systematic overview to emphasize the need of TISBR approach. To do that, we consider both CBIR and CBR as methodologies; hence it is possible to build a hybrid framework in terms of hybrid methodologies and not technologies. We identify three common axes of CBIR and CBR- indexing, retrieval, refinement- thus, a comparative analysis of CBIR and CBR from these perspectives comes intuitively. One should consider that in this approach we did not discussed techniques in detail; many papers did that. We only mentioned from time to time, a solution to an issue or introduced a technique, if it was necessary. The novelty of our approach is given by its very nature, a general comprehensive survey. Secondly, we present the CBIR and CBR in medical applications. Reasons such as digital visual data increasing production, partially annotation of images (due to the non-standardized, subjective, error biased procedure) and diagnosis support (reference database for education, standardization, computer-aided diagnosis, etc) are the pros for having CBIR in medical applications. Some shortcomings are however, not missing; for instance, the gaps, the page zero problem, the real-world system use. CBR fits better with the medical field, due to its similarity with physician reasoning.

From the application standpoint, the spotlight is set to Breast Cancer Grading. We discussed the motivations for it in section 4.1. To build the TISBR framework, we combined characteristics of CBR with CBIR and emphasized the usage of the integration, in the indexing axis. Related to the context gap, some inner problems that could be solved in a fusion paradigm are for instance the subjectivity of manual procedure (as usually pathologist adopt), time consuming and tedious tasks. These are alleviated by a semi-automated grading provided based on a semantic indexing method of histopathology images. Thus, a computer system that performs the automatic grading assists the pathologist by giving a second opinion. As already stated, this paper presents the general workflow for the semantic indexing; the specific details for each step are described in our previous works. The new idea here is the usage of ontology to structure CBR in line with the importance of ontologies for semantics in CBIR.

Regarding the perspectives, there are some directions we consider to follow in the near future. For the retrieval part, we envision of adopting an interpretive CBR in our approach. Furthermore, to make the retrieval phase more complex and efficient, TISBR will provide a query expansion procedure related to the new case matched against the case-base. In this way, we will be able to make queries on the new problem. To be consistent with the indexing, a semantic retrieval could be proposed[36, 37]. From our point of view, it is also highly important to have a continuity of the retrieval phase, starting with an efficient case adaptation and ending with the case storage and case-base maintenance. Interesting features to be added on a standard CBR, is the query, characteristic of CBIR. When it comes to selecting the most similar case, we propose a clustering of the obtained results to improve the accuracy of the retrieval phase To have a glimpse of how TISBR will function, see Fig.11.

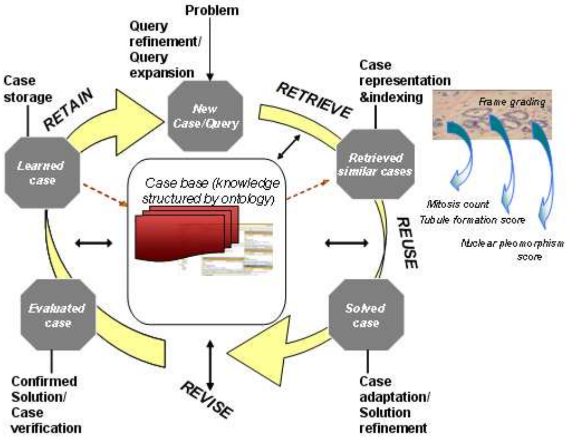


Fig. 11. TISBR strategy

Despite the progress done in both CBIR and CBR, the need of improvement still stands; and this comes in line with the open door for future research methodologies are not bereft of, as mentioned in section 2.1. With an accuracy of currently 80% for the indexing, TISBR strategy launches interesting perspectives for complex reasoning in medical research, mapped to specific applications. Also, it opens the way to better knowledge traceability and a better acceptance rate of computer-aided diagnosis assistance systems among practitioners in the future.

ACKNOWLEDGEMENTS

This project is partially supported by the ONCO-MEDIA[2] project and MMedWeb[3] project. We would like to thank to Pr. Teh Ming, head of the Pathology Department of the National University Hospital from Singapore for his constant support in this joint study.

REFERENCES

[1] Long F., Zhang H.and Feng D., [Multimedia Information Retrieval and Management – Technological Fundamentals and Application], D. Feng, W. C. Siu & Hongjing Zhang , Springer-Verlag, Germany, 1-26 (2003)

[2] Datta R., Joshi D., Li J. and Wang J., "Image Retrieval: Ideas, Influences, and Trends of New Age", ACM Transactions on Computing Surveys, 40(2), 1-66 (2008)

[3] Aamodt A. and Plaza E., "Case Based Reasoning: Foundational Issues, Methodological Variations and System Approaches", AI Communications, 7(1), 39-59 (1994)

[4] Leake D.B., [CBR in context : the present and the future], D.B. Leake, 1-35,AAAI/MIT press, Menlo Park, (1996)

[5] Schmidt R.and Gierl L., "Medical Case-Based Reasoning Systems: Experiences with Architectures for Prototypical Cases", Proc.MEDINFO, 518-522 (2001)

[6] Montani S., Magni P., Roudsari A. V., Carson E. R. and Bellazzi R., "Integrating different methodologies for insulin therapy support in type 1 diabetic patients", AIME, 121–130 (2001)

[7] Müller H., Michoux N., BandonD. And Geissbuhler A. "A Review of Content-Based Image Retrieval System in Medical Applications- Clinical Benefits and Future Directions", IJMI, 73, 1-23 (2004)

[8] Deserno T., Antani S. and Long R., "Gaps in content-based image retrieval", Proc SPIE, 6516,1-11(2007)

---

[2]ONCO-MEDIA (ONtology and COntext related MEdical image Distributed Intelligent Access) ICT ASIA project www.onco-media.com

[3]MMedWeb (Multimedia Medical Conceptual Web for Intelligent Information Access) project : http://www.comp.nus.edu.sg/~leowwk/MMedWeb/ A*STAR SERC 052 101 0103 (NUS R-252-000-319-305)

[9] Nillson M. and Sollenborn M., “Advancements and Trends in Medical Case-Based Reasoning: An overview of Systems and System Development”, Proc.FLAIRS, 178-183 (2004)
[10] Smeulders A., Worring M., Santini S., Gupta A. and Jain R., “Content -Based at the End of the Early Years”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349- 1380 (2000)
[11] Kolodner J., [Case-Based Reasoning], Morgan Kaufmann , San Francisco (1993)
[12] Richter M., Informal Introduction to Similarity-Based and Case-Based Reasoning (2003)
[13] Watson I., “Case based reasoning is a methodology not a technology”, KBS, 12(5-6), 303- 308 (1999)
[14] Checkland P.and Scholes J. [Soft Systems Methodology in Action], Wiley, NY (1990)
[15] KampG., Lange S. and Globig C, [Case-based Reasoning Technology: from Foundations to Application], M. Lenz, Springer, Berlin, 327 (1998)
[16] Veltkamp R. and Tanase M., [Content-Based Image and Video Retrieval], Kluwer Academic, 47-101 (2002)
[17] Vasconcelos N., “From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval”, Computer, 40(7), 20 -26 (2007)
[18] Smith B., “Beyond Concepts: Ontology as Reality Representation”, Proc.FOIS, 1- 12 (2004)
[19] Gruber T., “A Translation Approach to Portable Ontology Specification”, KA, 5(2), 199-220 (1993)
[20] Shi R., Chua T-S., Lee C-H. and and Gao S, [Image and Video Retrieval], H Sundaram et al, Spinger Verlag, Berlin, 102-106 (2006)
[21] Pal S. and Shiu S., [Foundations of Soft-Case-Based Reasoning], Willey & Sons , 4 -32 (2004)
[22] Zhao R. and Groski W., [Bridging the Semantic Gap in Image Retrieval, Distributed Multimedia Databases:Techniques and Applications], T. K. Shih, Idea Group, 14-36 (2001)
[23] Jaulent M.-C., Bozec C. L., Zapletal E. and Degoulet P, “A case-based reasoning method for computer assisted diagnosis in hisopathology”, AIM, 239–242 (1997)
[24] Jaulent M.-C., Bennani A., Le Bozec C. and Degoulet P, “A Customizabile Similarity Measure Between Histological Cases”, Proc. AMIA Symp, 350-354 (2002)
[25] Tutac A.E., “Histological Grading on Breast Cancer”, IPAL internal report, 1-13 (2007)
[26] Tutac A.E., Racoceanu D., Putti T., Xiong W., Leow W.-K. and Cretu V. , “Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images”, BioMedical Engineering and Informatics: New Development and the Future, Yonghong Peng and Yufeng Zhang, 107-112 (2008)
[27] Tutac A. E., Racoceanu D., Leow W.-K., Dalle J.-R., Putti T., Xiong W. and Cretu V, “Translational Approach for Semi-Automatic Breast Cancer Grading Using a Knowledge-Guided Semantic Indexing of Histopathology Images”, Proc.MIAAB, 1-8 (2008)
[28] Little S. and Hunter J., “Rules-By-Example- A Novel Approach to Semantic Indexing and Querying of Images”, Proc.ISWC, 534-548(2004)
[29] Kalfoglou Y., Dasmahapatra S., Dupplow D., Hu B., Lewis P. and Shadbolt N., “Living with the Semantic Gap: Experiences and Remedies in the Context of Medical Imaging”, Proc.SDMT, 46-47 (2006)
[30] Carneiro G., Chan A., Moreno P. and Vasconcelos N., “Supervised Learning of Semantic Classes for Image Annotation and Retrieval”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(3),394-410 (2007)
[31] Deb S. and Zhang Y., “An Overview of Content-Based Image Retrieval Systems”, Proc. AINA, 1, 59-64, (2004)
[32] Tadrat J., Boonjing V. and Pattaraintakorn P., “A Framework for using Rough Sets and Formal Concept Analysis in Case Based Reasoning”, Proc.IRI, 227-232 (2007)
[33] Traina J.C., Traina A.J., Araujo M.R., Bueno J.M., Chino F.J., Razente H. and Azevedo-Marques P., "Using an image-extended relational database to support content-based image retrieval in a PACS" Computer Methods and Programs in Biomedicine, 80(1), 71-83 (2005)
[34] Hidki A., Depeursinge A., Iavindrasana J., Pitkanen M., Zhou X. and Müller H., “The medGIFT project: perspective of a medical doctor”, Journal of Medical Imaging Technology, 25(5), 356-361 (2007)
[35] Lehmann T.M., Deselaers T., Schubert H., Güld M.O., Thies C., Fischer B. and Spitzer K., “IRMA - a content-based approach to Image Retrieval in Medical Applications”, Proc. IRMA, 911-912 (2006)
[36] Liu Y., Lazar N., Rothfus W., Dellaert F., Moore A.,Schneider J. and T.Kanade, “Semantic - based Biomedical Image Indexing and Retrieval”, Trends and Advances in Content- Based Image and Video Retrieval”, Shapiro, Kriegel and Veltkamp ,1-20 (2004)
[37] Tang H., Hanka R. and Ip H., “Histological Image Retrieval Based on Semantic Content Analysis”, IEEE Transaction on Information Technology Medicine, 7(1), 26-36 (2003)