Case-Based Reasoning in the Health Sciences: 
What's Next?

Isabelle Bichindaritz¹, Cindy Marling²

¹ University of Washington
1900 Commerce Street
Tacoma, WA 98402, USA
ibichind@u.washington.edu

² School of Electrical Engineering and Computer Science
Ohio University
Athens, Ohio 45701, USA
marling@ohio.edu

Corresponding author:

Dr. Isabelle Bichindaritz
University of Washington
1900 Commerce Street
Tacoma, WA 98402, USA
Telephone: +1 253 692 4605
Fax: +1 253 692 5862

Keywords. Case based reasoning, bioinformatics, medical informatics, artificial intelligence in medicine.
Abstract. This paper presents current work in case-based reasoning in the health sciences, describes current trends and issues, and projects future directions for work in this field. It represents the contributions of researchers at two Workshops on Case-Based Reasoning in the Health Sciences. These workshops were held at the Fifth International Conference on Case-Based Reasoning (ICCBR-03) and the Seventh European Conference on Case-Based Reasoning (ECCBR-04). Current research in CBR in the health sciences is marked by its richness. Highlighted trends include work in bioinformatics, support to the elderly and people with disabilities, formalization of CBR in biomedicine, and feature and case mining. CBR systems are being better designed to account for the complexity of biomedicine, to integrate into clinical settings, and to communicate and interact with diverse systems and methods.

1 Introduction

Case-based reasoning (CBR) is an Artificial Intelligence approach that capitalizes on past experience to solve current problems. It may be viewed, simultaneously, as a research paradigm, a perspective on human cognition, and as a methodology for building practical intelligent systems [21]. CBR has proven to be especially applicable to problem solving and decision support in the health sciences. Among the reasons for this are:

- Case histories have long been essential in the training of health care professionals
- The medical literature is filled with anecdotal accounts of the treatments of individual patients
- Many diseases are not well enough understood for formal models or universally applicable guidelines to be available
- When guidelines are available, they provide a general framework to guide clinicians, and require consequent background knowledge to become operational, which is precisely the kind of information recorded in practice cases; cases complement guidelines very well and help to interpret them
A biological system like the human body is difficult to describe by general models.

Even in domains where a deep model can represent the disease process – such as hypertension or heart disease – often several diagnoses interact to manifest the symptoms.

Reasoning from examples is natural for health care professionals.

Medicine is a highly data intensive field where it is advantageous to develop a system capable of reasoning from preexisting cases from an electronic medical record, for instance, or from cases mined from the data.

Both research and the practical application of CBR in the health sciences are currently experiencing rapid growth and development. Two Workshops on Case-Based Reasoning in the Health Sciences were recently held to provide a forum for identifying important contributions and research opportunities. The first of these was held at the International Conference on Case-Based Reasoning (ICCBR-03), in Trondheim, Norway, in 2003 [29]. The second was held at the European Conference on Case-Based Reasoning (ECCBR-04), in Madrid, Spain, in 2004 [17]. In total, seventeen papers were presented, representing the research and experience of 37 authors, working in ten different countries, on a wide range of problems and projects. The insights and contributions of those who participated in these workshops are manifested in this paper. Here, we reflect on some pioneering efforts in this field, summarize the current state of the art, describe current trends and issues, and look forward to future research directions.

2 Early Work

The first noted application of CBR in the health sciences was a proof of concept that CBR could be applied to the medical field [22]. Anecdotally, these authors decided to collaborate on some research when they met at a conference and were pleased to discover that they had the same last name. From this chance meeting grew the idea for a paper by Janet Kolodner – a CBR researcher – and Robert Kolodner – a psychiatrist and researcher in mental health computing. This paper [22] presents SHRINK, a CBR system designed to be a psychiatric diag-
nostician, capable of reasoning from a combination of category descriptors and cases to guide diagnosis and treatment. One of SHRINK’s characteristics was that it would get better over time, as its memory would learn new cases as it processed them – following the CBR paradigm of learning from success and failure situations. Although SHRINK was not a fully-implemented system, many of its ideas have since been implemented in MNAOMIA [6], a case based diagnosis and treatment system for eating disorders in psychiatry.

CASEY was one of the first CBR systems built in a health sciences domain [23]. CASEY diagnosed heart failure patients by comparing them to earlier patients whose diagnoses were known. It incorporated and extended an earlier model-based system, which diagnosed heart failures based on a physiological model of the human heart. A diagnosis, in CASEY, was represented as a graph, showing the causality among the different symptoms and features of a patient. CASEY was originally cited for its efficiency, as it produced diagnoses comparable to the model-based system, but did so an order of magnitude faster. It has since been recognized for pioneering the integration of CBR with another reasoning methodology, a trend that continues to this day in health sciences applications.

PROTOS was another early CBR system, which diagnosed audiological disorders [1]. PROTOS incorporated the knowledge of an expert audiologist and approximately two hundred actual patient cases. A diagnosis in PROTOS was a categorical label, assigning a new patient to one of a number of pre-specified diagnostic categories. A new patient was effectively classified as having the same diagnosis as other similar past patients. PROTOS was influential in that its approach was adopted for building systems in other domains. For example, ProtoISIS was developed using the same techniques to select the most appropriate diagnostic imaging procedures for a patient [20]. An in-depth review of early work in CBR in the health sciences is available in several excellent summary articles [6, 16, 32]. Here, we briefly note some other early influential systems.

- MEDIC diagnoses pulmonary disease [40]
- ALEXIA determines a patient's hypertension etiology [10]
- FLORENCE assists with nursing diagnosis, prognosis, and prescription [11]
- ROENTGEN helps to design radiation therapy plans [4]
- MacRad helps to interpret radiological images [26]
3 Current Work

Table 1 provides a brief overview of the systems discussed at the two Workshops on CBR in the Health Sciences. Here, we discuss these systems and the trends and issues they represent. ALEXIA, MNAOMIA and CARE-PARTNER are influential systems that have been presented earlier in other venues [5, 6, 9, 10]. Here, they are of continuing interest in that (a) safety issues are being addressed [7] and (b) they are being integrated to provide a basis for semantic interoperability in CBR systems in biology and medicine [8]. ALEXIA combines CBR with a causal, physiopathological model to diagnose hypertension disorders. MNAOMIA assists clinical staff members with diagnosis, treatment, and research hypotheses for psychiatric eating disorders. It relies on cases, concepts, prototypes and models in its reasoning processes. CARE-PARTNER combines CBR with rule-based reasoning to provide decision support for the long-term follow-up care of stem-cell transplant patients. As stem-cell transplants are still new, CARE-PARTNER implements evidence-based medical practice.
by formalizing guidelines established for the care of stem-cell transplant patients. Some current trends evidenced by these systems include the use of multi-modal reasoning and the need to represent time series data in order to reason about patient progress over time [38].

Table 1. Systems Discussed at the Workshops on CBR in the Health Sciences.

<table>
<thead>
<tr>
<th>System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARE-PARTNER</td>
<td>Supports long-term follow-up care of stem-cell transplant patients [7, 9]</td>
</tr>
<tr>
<td>CASEREC</td>
<td>Interprets digital microscopic images to identify hazardous airborne fungi [34, 35]</td>
</tr>
<tr>
<td>Endocrinology Aid</td>
<td>Supports prescription and monitoring of levothyroxine dosage for endocrine therapy [41, 42]</td>
</tr>
<tr>
<td>FM-Ultranet</td>
<td>Interprets ultrasound scans and diagnoses fetal malformations [14]</td>
</tr>
<tr>
<td>Gene Finder</td>
<td>Identifies coding regions in strands of mammalian DNA [12]</td>
</tr>
<tr>
<td>HR3Modul</td>
<td>Classifies sensor measurements to diagnose stress-related disorders [31, 39]</td>
</tr>
<tr>
<td>Mémoire</td>
<td>Provides a framework for the semantic interoperability of CBR systems in biology and medicine [8]</td>
</tr>
<tr>
<td>NutriGenomics</td>
<td>Provides nutrition counseling based on individual genetics, health and lifestyle [43]</td>
</tr>
<tr>
<td>RHENE</td>
<td>Assesses the efficiency of hemodialysis treatment for patients with end stage renal disease [30]</td>
</tr>
<tr>
<td>SISAIH</td>
<td>Supports decision making in hospital admissions authorization [25]</td>
</tr>
<tr>
<td>SMARTHOME</td>
<td>Matches assistive living technologies to the needs of the elderly and disabled [13]</td>
</tr>
<tr>
<td>Somnus</td>
<td>Supports diagnosis and treatment of obstructive sleep apnea [24]</td>
</tr>
<tr>
<td>WHAT</td>
<td>Prescribes exercise regimens for cardiac and pulmonary patients [15]</td>
</tr>
</tbody>
</table>

The issue of patient safety is raised and addressed by CARE-PARTNER [7]. CARE-PARTNER provides advice to hometown health care providers after stem-cell transplant patients return to their home communities. Because these patients are immuno-compromised, seemingly small problems can escalate quickly, and timely, accurate advice is essential. The possibility that a patient could be harmed by system error was a major issue to overcome in fielding this system. CARE-PARTNER employs an introspective reasoning process in which it determines which signs, symptoms and recommendations are safety-critical. It provides non-safety-critical advice immediately, but contacts expert clinicians to assist in critical situations. Safety is an important issue for all CBR systems that influence patient treatment, espe-
cially those that automatically generate patient recommendations, rather than simply providing retrieved cases to health care professionals to support their own reasoning.

Most recently, ALEXIA, MNAOMIA and CARE-PARTNER have provided the test bed for Mémoire, a framework for unifying and sharing case bases and biomedical CBR systems [8]. Most CBR systems to date have been standalone systems, which makes it difficult for system users to find all the support they need and for CBR systems developers to build upon each others' efforts. Issues raised include patient privacy, the security of patient data, and adopting the right technological approach to facilitate interoperability. Mémoire employs a semantic approach, similar to that developed for the semantic Web, for sharing and distributing cases, prototypical cases, case based ontologies, and CBR systems in biology and medicine.

Gene Finder [12] and the system for NutriGenomics [43] are two research prototypes that portend CBR's promise for applications in bioinformatics. Gene Finder analyzes nucleotide sequences to identify DNA segments that encode biological functionality. It employs a case library of nucleotide segments that have already been classified as coding or non-coding to find the coding regions in previously unseen strands of DNA. The long range goal of this system is to advance gene therapy, the identification and eventual alteration of genes that cause disease. The system for NutriGenomics provides personalized nutrition counseling by leveraging emerging knowledge of how molecular nutrients affect gene expression to promote health and prevent disease. It makes use of ontologies to represent domain knowledge, data mining techniques to extract knowledge from external databases, and a distributed case base structure. It configures a dietary strategy for an individual, considering their own genetic makeup as well as their current health status and dietary lifestyle. While neither system is fully implemented, both provide frameworks for continuing work in case-based bioinformatics. Additional work in this area was recently reported in [19]. Here, it was noted that CBR is a good fit for bioinformatics applications as there are massive amounts of data, many unknowns, and an incomplete, evolving, domain theory.

CASEREC, which identifies hazardous airborne fungi [34, 35], and HR3Modul, which classifies respiratory sinus arrhythmia (RSA) sequences [31, 39], illustrate a growing capacity to automate analysis of medical images and biophysical signals. While the use of images and continuous measurements is not new to medical CBR systems, many prior sys-
tems have depended upon manual extraction of relevant features and/or manual analysis of retrieved images. To automate these tasks, there is a focus on developing novel metrics to detect similarity in nontextual data. CASEREC analyzes digital microscopic images to detect contaminants in working environments. It combines CBR with model-based image analysis to recognize fungi spores by learning shape models from past images and applying them to new images. In this way, it detects objects in images that are highly likely to be fungi spores and then identifies the types of fungi spores detected. HR3Modul classifies RSA sequences by analyzing sensor readings of the variation in heart rate that occurs with breathing. Each measurement is taken for ten to fifteen minutes and includes sixty to eighty respiration cycles. The sequence analysis is used to diagnose stress-related disorders.

The Wellworks/Heartworks Advisor/Trainer (WHAT) [15], SMARTHOME [13], and the support system for endocrine therapy [41, 42] are prescriptive systems. WHAT helps sports medicine students learn to prescribe exercise regimens for patients already diagnosed with cardiac or pulmonary problems. This system serves an educational role by automatically generating independent case-based and rule-based prescriptions for the same patient and allowing students to compare them. The SMARTHOME prototype recommends the most effective technological aids for individuals impaired in the capabilities needed for successful independent living. Using the commercial ReCall shell with an iterative decision tree retrieval, this prototype was able to produce recommendations comparable to an expert's. The endocrinology system prescribes and monitors levothyroxine dosages for patients diagnosed with hypothyroidism. This system extends the prescriptive capabilities of WHAT and SMARTHOME, in that it follows patients over time and adjusts dosages in response to patient conditions. It also provides a framework for other medical conditions in which initial therapies may not be effective and may need adjustment. Other systems with prescriptive components include MNAOMIA, CARE-PARTNER, NutriGenomics and Somnus. Other systems with tutorial functionality include CARE-PARTNER, Somnus and FM-Ultranet. Other systems that follow patients and adjust treatment over time are CARE-PARTNER, MNAOMIA and RHENE.

Assessing treatment efficacy using case-based retrieval is the focus of RHENE, which monitors hemodialysis sessions for patients with end stage renal disease [30]. A typical hemodialysis session lasts for four hours, during which twenty-five variables are auto-
matically sampled at one second intervals. Each hemodialysis session is represented as a case, containing the time series data collected and additional static features that characterize the patient. This system allows physicians to analyze current problems by considering past patterns of failure over time and the solutions provided in terms of prescribed dialysis flow rates. It also supports quality assessment of the hemodialysis service. FM-Ultranet [14] and Somnus [24] are primarily diagnostic systems. Diagnosis was the first task supported by medical CBR systems, and diagnostic systems continue to make important contributions.

FM-Ultranet [14] supports ultrasonagraphists in detecting fetal malformations. When a practitioner detects an abnormality during a routine ultrasound screening, FM-Ultranet helps to determine if the abnormality is dangerous. Future plans for this system include incorporating automated image analysis and pattern recognition techniques. FM-Ultranet was implemented using the commercial CBR-Works shell. Somnus [24] is a prototypical system for respiratory therapy students in a sleep disorders clinic. It helps students diagnose and treat obstructive sleep apnea, a breathing disorder that repeatedly wakes patients from sleep. Somnus provides support by retrieving cases similar to a student’s current case for comparison. It also allows students to review patient profiles and to analyze treatments and patient compliance. Somnus combines semiotics, fuzzy logic, and database query techniques with CBR.

SISAIH [25] provides decision support for authorizing admissions to Brazilian hospitals. It supports financial accountability by ensuring that admissions are justified. It detects billing errors by comparing current admissions records to past ones, focusing on multiple admissions of the same patient. As the requisite financial data was already maintained online, SISAIH was able to streamline a cumbersome manual review process using a standard CBR approach.

4 What’s Next?

As pioneering work and the work presented at these workshops illustrate, the health sciences have proven to be a fruitful application domain for CBR. CBR systems in the health sciences have started from systems aimed at modeling medical expertise [22], and case based learning as the process by which physicians learn their expertise through medical practice. They have
evolved into the design of case based assistants for clinical practice, quality control, and clinical forecast. It is expected that the interest in CBR systems in the health sciences is going to grow, as longer life expectancy forces a focus on health care and human biology.

A physician – a psychiatrist – once portrayed what he envisioned as the perfect computer program to assist him in his clinical work. He described representing in a database all the patients he had been treating, with all the details of their history, environment, symptoms, diseases, treatments, and evaluations. He also imagined that he could get access to patient cases of his colleagues. When encountering a new patient or a new disease episode, the system would find the most similar patients and patient episodes, would show them to him, explain the similarities and differences, how each episode was solved successfully, and provide recommendations for diagnosis and treatment for the particular episode he was dealing with. Ultimately, realizing this vision is one of the main goals of CBR in the health sciences. With this in mind, we here propose a roadmap for the future and point out some pitfalls that may arise along the way.

4.1 Roadmap for the future

Highlights of the trends and opportunities that may develop for CBR in the health sciences are the following:

- **Formalization of CBR in the health sciences.** Many case based reasoning systems have been successfully developed in medical and biological domains. A next step in this area of research is to study what is common among these systems, starting from case representation formalism. In particular, formalized case representation language [8], case structure, and reasoning processes [41, 42] will become more available and more important. We can see this as a necessary evolution paralleling the evolution of the World Wide Web from HTML, to XML, and on to OWL as a representation language for Web resources. The benefits of this evolution are numerous, among them the ability to make case bases and CBR systems interoperable, to provide a universal query language for case bases, and to allow the integration of CBR with information search and retrieval, as well as with other systems and methods [15].
Bioinformatics applications. Decoding the Human Genome and the growing availability of genetic data are going to foster the development of CBR applications in bioinformatics, and the integration of genetic data into the medical profile of a patient case. Since biomedical informatics is rapidly developing at the frontier between medical informatics and bioinformatics, CBR systems will also have to reason from both types of data, and to integrate methods from both fields. CBR potentially can be a method of interest in bioinformatics itself [12, 19, 43]. We also note here the work of some researchers who see CBR as a method to assist biomedical research [9].

Integration with electronic patient records and in healthcare settings. Methods from the cognitive sciences will assist in this process, as will close cooperation with health informatics specialists. CBR systems in healthcare environments will have to be safe [7], usable, and incorporate standards of care, in particular, evidence-based medical practice [9]. The formalization efforts mentioned above will also play an important role in this integration. When CBR systems are able to take advantage of patients’ representations in electronic health records, they will become applicable to a wide range of diseases. It is exciting to think of the vast number of potential case bases, the knowledge of which is dormant simply because CBR systems have been developed from a different tradition – Artificial Intelligence versus Health Informatics. CBR researchers need to invest time learning from the Informatics community decades of work representing patient data, exchanging patient records, and formalizing clinical knowledge. Efforts and achievements such as HL7 are highly pertinent for CBR research. CBR researchers will then be able to integrate this work and bridge the gap between their medical case representations and those developed for electronic patient records.

Integration with information retrieval. CBR shares a common framework with information retrieval, and with the growing availability of textual information, including that in medical records, cooperations between textual CBR and information retrieval will develop. Methods for learning cases from text will facilitate the develop-
ment of CBR systems from medical literature [43] and CBR systems without initial cases. Many medical institutions either encourage or force care providers - physicians, physician assistants, or nurses – to take electronic notes about patient visits, consultations, or hospitalizations. An example is the VA system, which has reached a completely electronic state for patient records. These textual traces of the clinical problem solving process are a currently untapped source of case bases that textual CBR could take advantage of.

- **Feature mining.** In several applications, the case representation will be more complex than in past CBR systems. For example, case data will come partly from sensors [31, 32], images [14, 34, 35], or time series [30]. Feature mining consists of synthesizing adequate features for CBR. For instance, feature mining can involve reducing the number of features in highly dimensional data, such as the ones found in micro arrays for bioinformatics cases [19]. Another example is identifying features from physiological signals [31]. From textual records, finding the features among documents is another variant of feature mining.

- **Case mining.** Mining for cases from existing data sources will become more and more important, particularly as the integration of CBR with electronic health records will progress. It consists of learning cases by mining databases or datasets, sometimes involving the consolidation of data from diverse tables, records, or other sources into a unified case structure. Currently, few CBR systems take advantage of this method, and this is why more research is needed in this area. It is often noted by researchers that information about the diagnostic and/or treatment process is not well recorded in patient databases; the information is incomplete. With electronic health records, a more comprehensive view of patient problem situations becomes available to facilitate the case mining process. Here again, mining for cases will often take the form of text mining in progress notes.

- **Support for the elderly and disabled.** As the population ages, these applications will grow in number and significance. CBR researchers have already tackled the prob-
lens of enabling independent living for the elderly [13] and of providing support for patients with Alzheimer’s Disease [27]. Many additional opportunities await.

- **Accounting for the complexity of biomedical domains.** Systems will tackle the growing mass of information, in particular for quality control and monitoring [25, 30], mixing their recommendations with probabilities and statistics, and taking into account the co-occurrence of several ailments [24]. One important focus will be how case based reasoning can associate probabilities and statistics with its results, and how the two different methodologies can complement one another.

- **Life long learning.** CBR systems in medicine should not be designed only to reuse past episodes with little modification. Medicine is a rapidly changing field, and medical practice guidelines are regularly updated. Medications change, and new treatments emerge constantly. Therefore, the recommendations for diagnosis, assessment and treatment that were the norm five years ago may today be obsolete in part. The long-term follow-up CBR system [9] should be designed as a life long learning system, integrating case base maintenance to propagate new findings into the case base, or take recent developments into account at reuse time. Another facet of this is to prioritize recent cases over old cases when a choice is available.

- **Guidelines complementarity.** The evolution of medicine toward evidence-based practice is a world-wide trend that will foster the development of decision support in medicine. Clinicians are supposed to keep updated in their knowledge of recent guidelines in several medical domains, eventually with assistance from software for disseminating the state-of-the-art in medical science. Cases are a necessary complement to guidelines because they provide the detailed context that is generally lacking in the interpretation of guidelines by computerized systems. They also provide some common sense knowledge that is crucially missing in expert systems. Another role envisioned for cases is to generate hypotheses for amending the guidelines from clinical practice, either complementing guidelines where there are gaps, or refining
the interpretation context, or even providing alternate courses of action and ideas for improving future guidelines.

- **Data mining for CBR.** CBR systems implement a variety of data mining tasks and methods such as classification, for instance. MNAOMIA [6] proposed to go one step further by using the memory structures learned during CBR to perform a data mining task as a side effect of case based reasoning to assist the diagnosis and treatment tasks. The memory structures learned were called concepts, or trends, because they synthesized information common to sub-hierarchies of concepts and cases in the memory. When this declarative type of learning occurs, data mining can be added as a complement to CBR, as a meta-level technique capable of answering such questions from clinicians as “What has your system learned?” and eventually providing clinicians with research hypotheses. A particular form of data mining of interest for CBR systems would be prototype mining. Many CBR systems use some kind of prototypical cases, variously known as prototypes, classes, or abstract cases. These entities could be mined from cases, and may again be of interest to clinicians in addition to being of value for CBR.

### 4.2 Pitfalls along the way

Biomedicine is certainly a domain with many pitfalls and puzzling aspects. Among the main ones that may hinder the development of CBR in the health sciences are the following:

- **Legal issues.** Legal issues surrounding patient data restrict access to medical data and require advanced methods for data protection. Related to this are the issues of safety [7] and medical responsibility with regard to using computer-based systems in decision support

- **Reliance on biostatistics over other computational methods.** In the medical field, statistics add a scientific dimension to clinical research. Complementary connections
between statistics and CBR should be made to ensure acceptance of CBR among clinicians.

- Building medical knowledge bases is a long process, for which the availability of practice guidelines and their computerization provides a starting point. Future case based reasoning systems will have to cooperate with one another and with other methodologies in order to grasp the vastness of this specialized knowledge.

- Differential diagnosis. Medical diagnosis involves providing a network and ranking of underlying causes accounting for exhibited symptoms. This again links to statistical considerations and the ability to connect different computerized systems with one another.

7 Summary and Conclusion

This paper has described the current state of the art in case-based reasoning in the health sciences, projected future opportunities and challenges, and warned of potential pitfalls. Current research is marked by its richness, with its wide range of application domains and its integration with numerous complementary approaches and technologies. Opportunities abound, including new contributions to bioinformatics, support for the elderly and disabled, formalization of CBR in biomedicine, feature and case mining, and ultimately the development of systems better designed to account for the complexity of biomedicine and to integrate into clinical settings.

This paper grew out of work presented and discussions generated at two international workshops on case-based reasoning in the health sciences. A third workshop will be held at the Sixth International Conference on Case-Based Reasoning (ICCBR-05) to build upon progress made to date. As researchers join together to exploit the opportunities, while avoiding the pitfalls, CBR will find an even more fruitful niche in the health sciences.
Acknowledgments

The authors wish to thank the participants of the first and second Workshops on Case-Based Reasoning in the Health Sciences for their insights and contributions.

References


