Exploiting the Knowledge Computing and Engineering in Medical Informatics and Healthcare

ABDEL-BADEEH M. SALEM
Medical Informatics and Knowledge Engineering Research Labs
Ain Shams University
Abbasia, Cairo
EGYPT
abmsalem@yahoo.com absalem@cis.asu.edu.eg http://net.shams.edu.eg/27.htm

VIRA SHENDRYK
Department of Computer Science
Sumy State University
Sumy
UKRAINE
v.shendryk@cs.sumdu.edu.ua ve-shen@opm.sumdu.edu.ua http://sumdu.edu.ua

SERGII SHENDRYK
Department of Computer Science
Sumy State University
Sumy
UKRAINE
s.shendryk@cs.sumdu.edu.ua http://sumdu.edu.ua

Abstract: - The field of knowledge computing (KC) has become the most challenging area in the last several years. KC deals with the development of intelligent computing and knowledge-based systems in which knowledge and reasoning play pivotal role. KC consists of three main areas, namely: Document Engineering (DE), Knowledge Engineering (KE), and Reasoning Techniques (RT). KE includes; knowledge acquisition, expert systems, ontologies, knowledge-based systems, knowledge compilation, shells and tools, methodologies, modeling, knowledge management, knowledge discovery, and knowledge representation techniques. The aim of this paper is to make an overview of some of KC techniques and approaches and their applications in medical informatics and healthcare. The paper discusses the following techniques and applications: case-based reasoning approach for cancer and heart diagnosis, ontological engineering for breast cancer knowledge management, and mining patient data using rough sets theory to determine thrombosis disease.

Key-Words: - Knowledge Computing, Knowledge Engineering, Knowledge-Based Systems, Ontological Engineering, Knowledge Discovery, Computational Intelligence, Health Informatics

1 Introduction
During recent decades, knowledge is the aspiring elementary resource mandatorily required by all intelligent information processing systems. Knowledge engineering is an intelligent process by which the gathered raw data is transformed into knowledge. Knowledge engineers use artificial intelligence (AI) concepts and techniques in developing knowledge-based systems [1, 2]. AI is based on many disciplines such as: computer science, philosophy, psychology, mathematics, biology, linguistics, and cognitive sciences. Various intelligent methodologies, computational techniques and knowledge-based systems have been developed for automated reasoning and learning. AI techniques are robust, applied successfully to complex problems, efficiently adaptive, and usually have a parallel computational architecture. For those reasons they have been proved to be effective and efficient in developing intelligent and knowledge-based systems for many tasks in biomedical informatics and health sciences [3, 4]. Applications of AI technology include; general problem solving, expert systems, natural language processing, computer vision, robotics, and education. All of these applications employ knowledge base and inference techniques to solve problems or help make decisions in specific domains [2, 4]. These systems acted primarily as browsing systems and information providers. In this paper we focus our discussion around some of knowledge computing and
engineering techniques and approaches and their applications in medical informatics and healthcare. Herein presented the following techniques and applications: (a) case-based reasoning for cancer and heart diagnosis, (b) ontological engineering for breast cancer knowledge management, and (c) mining patient data using rough sets theory to determine thrombosis disease.

2 The Key Aspects of Knowledge Computing (KC)

To define the scope of the KC discipline, a comprehensive analysis has been done of the publications during the last 10 years [3, 5,6]. Figure 1 shows the KC main disciplines and its research areas. From this figure it can be seen that the scope of the field of KC consists of three main areas, namely: Document Engineering (DE), Knowledge Engineering (KE), and Reasoning Techniques (RT). KE includes; knowledge acquisition, knowledge-based systems, knowledge compilation, Shells and Tools, Methodologies, Modeling, Knowledge Discovery, and Knowledge Representation and Management.

From our analysis of the published literatures about the knowledge-based systems, we can summarize the general features of KC approaches in the following:

- Intelligence can be achieved through formal operations on symbol structures;
- Represent knowledge explicitly and manipulating it through an inference engine;
- Programmed is rather used than trained;
- The computing is constrained to the sequential operations;
- The overlay is sensitive to noise;
- Models emphasize the use of symbols to donate objects and relations in a domain of interpretation;
- Relying upon a single powerful processor;
- Symbolic –based artificial intelligence approach;
- Programs suffer from the “brittleness” problem - perform perfectly until it fails to perform at all;
- Expert systems are the typical examples realization of KC approaches.

If we consider Knowledge Engineering for Health Informatics, then it can be represented in the form of the sequence shown in the figure 2.

2.1 Knowledge Representation Techniques

Knowledge is the main key of knowledge computing and engineering. In order to act intelligently, a computer must have knowledge about the domain of interest. The knowledge of the domain must be collected and codified. It must be organized, outlined, or otherwise arranged in a systematic order. This process of collecting and organizing the knowledge is called knowledge engineering. It is the most difficult and time-consuming stage of any AI software development process. The knowledge consists of facts, concepts, theories, and procedures. A variety of knowledge representation schemes are used including; lists, trees, semantic networks, frames, scripts, production rules and ontology [7]. Lists are used to represent hierarchical knowledge. Hierarchical knowledge can also be represented visually with graphs called trees. Semantic networks use circles called nodes that represent objects or events. The nodes are interconnected with lines called arcs that show relationships. Frames and scripts are two types of schemes dealing with stereotyped knowledge. Frames are used represent facts about objects and events. And details are given in sub-elements called slots. Scripts describe knowledge that is a sequence of events or procedures. Frames and scripts permit a system to infer details of specific common objects and events. Production rules are the most commonly used knowledge representation methods. The rules are two parts statements with a premise and a conclusion and are written in the form of an if-then statement. They also may state a situation and corresponding action. For more details we refer to the books [1, 2, 3].

In fact, these representation schemes share two common characteristics. First, they can be programmed with computer languages and stored in memory. Second, they are designed so that the facts and other knowledge contained within them can be manipulated by an inference system, the other major part of the knowledge-based system. The inference system uses search and pattern matching techniques on the knowledge base to answer questions, draw conclusions, or otherwise perform an intelligent function.

2.2 Reasoning Methodologies and Techniques

The field of reasoning is very important for the development of knowledge-based systems. The research area in this field covers a variety of topics, e.g.; automated reasoning, case-based reasoning, commonsense reasoning, fuzzy reasoning, geometric reasoning, non-monotonic reasoning, model-based reasoning, probabilistic reasoning, causal reasoning, qualitative reasoning, spatial reasoning and temporal reasoning [1, 8]. In fact, these methodologies receive increasing attention within the knowledge computing
and engineering in biomedical and health informatics communities [4].

2.2.1 Reasoning with Production Rules
Rules are easily manipulated by reasoning systems. Forward chaining can be used to produce new facts (hence the term “production” rules), and backward chaining can deduce whether statements are true or not. Rule-based systems were one of the first large-scale commercial successes of AI research. On the other side, backward chaining is often called goal-directed inference, because a particular consequence or goal clause is evaluated first, and then we go backward through the rules. Unlike forward chaining, which uses-rules to produce new information, backward chaining uses rules to answer questions about whether a goal clause is true or not. Backward chaining is more focused than forward chaining, because it only processes rules that are relevant to the question. It is like to how resolution is used in predicate logic. However, it does not use contradiction. It simply traverses the rule base trying to prove that clauses are true in a systematic manner. Backward chaining is used for advisory systems, where users ask questions and get asked leading questions to find an answer. One advantage of backward chaining is that, because the inference is directed, information can be requested from the user when it is needed. Some reasoning systems also provide a trace capability which allows the user to ask the inference engine why it is asking for some piece of information, or why it came to some conclusion.

2.2.2 Reasoning with Fuzzy Rules
Unlike Boolean logic, which has only two states, true or false, fuzzy logic deals with truth values which range continuously from 0 to 1. The use of fuzzy logic in reasoning systems impacts not only the inference engine but the knowledge representation itself. Reasoning with fuzzy rule systems is a forward-chaining procedure. The initial numeric data values are fuzzified, that is, turned into fuzzy values using the membership functions. Instead of a match and conflict resolution phase where we select a triggered rule to fire, in fuzzy systems, all rules are evaluated, because all fuzzy rules can be true to some degree (ranging from 0.0 to 1.0). The antecedent clause truth values are combined using fuzzy logic operators (a fuzzy conjunction or and operation takes the minimum value of the two fuzzy clauses). Next, the fuzzy sets specified in the consequent clauses of all rules are combined, using the rule truth values as scaling factors. The result is a single fuzzy set, which is then defuzzified to return a crisp output value.

2.2.3 Reasoning with Cases
Reasoning from experience is a natural way of human thinking, one remembers an apparently similar situation, what one has done and what the outcome has been; accordingly one acts in the present situation. The case-based reasoning (CBR) draws from this paradigm and tries to formalize it for use on the computer [9]. CBR is the scientific method (or collection of methods) to imitate and enhance, if possible, this human behavior to find useful and applicable old cases and to reuse them either directly or after adaptation. In addition, the success of adaptation has to be verified and cases have to be collected for future use. CBR, as a computational intelligence method, assumes a memory model for representing, indexing and organizing past cases and a process model for retrieving and modifying old cases and assimilating new ones [9, 10]. CBR has already been applied in a number of application areas, such as customer support and environmental monitoring applications, medical and health informatics, and intelligent robotics [9, 10, 11, 12, 15].

The case is a list of features that lead to a particular outcome (e.g. The information on a patient history and the associated diagnosis). The complex case is a connected set of subcases that form the problem solving task’s structure (e.g. The design of an airplane). Determining the appropriate case features is the main knowledge engineering task in case-based knowledge-based software. This task involves defining the terminology of the domain and gathering representative cases of problem solving by the expert. Representation of cases can be in any of several forms (predicate, frames). The idea of case-based reasoning is becoming popular in developing knowledge-based systems because it automates applications that are based on precedent or that contain incomplete causal models. In a rule-based system an incomplete mode or an environment which does not take into account all variables could result in either an answer built on incomplete data or simply no answer at all. Case-based methodology attempt to get around this shortcoming by inputting and analyzing problem data. For more technical information, see [9, 10, 13].

2.2.4 Benefits and advantages of Rough Set Theory
Rough set theory was proposed as an efficient approach to vague concept description from incomplete data. The rough set theory is one of the most useful techniques in many real-life applications such as medicine, pharmacology, engineering, banking and market analysis. This theory provides a
powerful foundation to reveal and discover important structures in data and to classify complex objects. In accordance with [14], we can distinguish the following benefits and advantages of rough set theory:
- deals with vagueness data and uncertainty;
- deals with reasoning from imprecise data;
- used to develop a method for discovering relationships in data;
- provides a powerful foundation to reveal and discover important structures in data and to classify complex objects;
- do not need any preliminary or additional information about data;
- concerned with three basics: granularity of knowledge, approximation of sets and data mining.

3 Intelligent Medical Informatics and Healthcare

Figure 2, illustrates the main disciplines of the health informatics. While figure 3, shows the multidisciplinary field of research of intelligent medical informatics. The main research areas are: Robotic Surgery, Medical Education/Training, Medical Imaging, Medical Knowledge Engineering, and Medical Expert Systems. Currently, most of the researches are focused in the following topics:
- Machine learning for personal identification and authentication using EEG and ECG bio-signals;
- Visualization algorithms for orthopedic surgery;
- Machine Intelligence and deep learning for detecting brain tumors from magnetic resonance imaging (MRI images);
- Bio-inspired computing for predicting thrombosis disease;
- Visualization algorithms for medical data mining;
- Mobile computing for cancer management using computational intelligence techniques;
- CBR techniques for emergency medical responders;
- Web-based lesson planning system based on CBR;
- Using ubiquitous computing with intelligent wireless sensors, ad-hoc, and mesh networks in teaching and learning processes;
- Virtual reality technologies in the healthcare, and medical education;
- Cloud computing and Internet of Things (IoT), for healthcare;
- Ontological engineering for caner diseases.

4 Case-Based Expert System for Cancer Diagnosis

Case-Based expert system (CES), uses CBR methodology, can reason from analogy from the past cases. This system contains what is called “case-memory” which contains the knowledge in the form of old cases (experiences). CES solves new problems by adapting solutions that were used for previous and similar problems [2]. CESs directly addresses the problems found in rule-based expert systems namely: knowledge acquisition, performance, adaptive solution, and maintenance. For more technical information, see [11, 12]. 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.

From the knowledge engineering point of view, the system consists of three main modules; user interface, case base reasoning model and computational model all are interacted with the main environment of cancer diseases. 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 model 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 (nearest neighbor matching). The system is implemented using Visual Prolog for Windows. 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. The system uses the case-based reasoning strategy to record and retrieve its knowledge. The initial diagnostic process is done through firing of rules in the rule-based inference. These rules encode information about patient’s symptoms and pathological examinations. For more technical information, see [16].

5 Expert system for Diagnosis of Heart Diseases

Heart disease is a vital health care problem affecting millions of people. Heart disease are of 25 different ones; e.g. left-sided heart failure, right-sided heart failure, angina pectoris, myocardial infraction and essential hypertension. This research deals with the
development of a rule based expert system for
diagnosis of heart diseases. The system can give an
appropriate diagnosis for the presented symptoms,
signs and investigations done to a cardiac patient with
the corresponding certainty factor. It aims to serve as
doctor diagnostic assistant and support the education
for the undergraduate and postgraduate young
physicians [17].

In our case-based expert system, we have represented
the knowledge in the form of frames and built the
case memory for 4 heart diseases namely; mistral
stenosis, left-sided heart failure, left-sided heart
failure, stable angina pectoris and essential
tension. For the case retrieval, we have
developed two algorithms namely; nearest-neighbor
algorithm and induction algorithm so that we can
measure the system performance in both cases. The
system searches for the most adequate cases for the
current case, a similarity value between each retrieved
case and the current case is calculated and
the retrieved cases are ranked according to these
values. The system has been implemented in Visual
Prolog for Windows and has trained set of 42 cases
for Egyptian cardiac patients and has been tested by
another 13 different cases. Each case contains 33
significant attributes resettled from the statistical
analysis performed to 110 cases. The system has been
tested for 13 real cases. The experimental results
have shown 100% accuracy in estimating the correct
results for using nearest neighbor algorithm and this
percentage is dropped to 53.8% in case of using the
induction algorithm.

6 Mining patient’s data using rough
sets theory to determine thrombosis
disease.

Intelligent data mining methodology aims to extract
useful knowledge and discover some hidden patterns
from huge amount of databases which statistical
approaches cannot discover. Knowledge discovery in
databases (KDD) process involves the following
phases; (a) using the database along with any
required selection, preprocessing, sub-sampling, and
transformations of it, (b) applying data mining
methods [18,19] to enumerate patterns from it, and
(c) evaluating the products of data mining to identify
the subset of the enumerated patterns deemed
knowledge. The data mining components of the KDD
process is concerned with the algorithmic means by
which patterns are extracted and enumerated from
data. The overall KDD process includes the
evaluation and possible interpretation of the mined
patterns to determine which patterns can be
considered new Knowledge. Data mining is
supported by a host that captures the character of data
in several different ways (e.g. clustering,
classification, link analysis, summarization,
regression models, and sequence analysis). Figure 4
shows the main functional phases of the knowledge
discovery process [20].

Recently, knowledge computing paradigms plays an
essential role in the area of medical data mining, e.g.:
- hospitals analyze historical data to help decide
which new patients are likely to respond best to
which treatments;
- extract knowledge from medical and biological
data.

In modern period Data Mining and Knowledge
Discovery model very often are used for solving
medical problems in such way:
- translating the medical problem into Machine
Learning (ML) domain, in which the goals, current
solutions and domain terminology are defined;
- understanding the data, where the corresponding
medical data is described and analyzed with respect
to the underlying ML problem;
- preparation of the data, in which the data
preprocessing methods are applied;
- data mining, in which the prepared data is processed
with ML techniques;
- evaluation of the discovered knowledge, where the
results provided in the previous step are evaluated;
- using the discovered knowledge, in which the
generated knowledge is deployed.

At the Artificial Intelligence and Knowledge
Engineering Research Labs at Ain Shams University,
Egypt, a rough set-based intelligent system for
mining patient data for predictive rules to determine
thrombosis disease was developed [20] this system
aims to search for patterns specific/sensitive to
thrombosis disease. The data was made through
“Discovery Challenge Competition”, organized in
the frame of the 3rd European Conference on
Principals and Practice of KDD in Prague, 1999. This
system reduced the number of attributes that describe
the thrombosis disease from 60 to 16 significant
attributes in addition to extracting some decision
rules, through decision applying decision algorithms,
which can help young physicians to predict the
thrombosis disease.

7 Ontological Engineering approach
for breast cancer knowledge
management

Ontological engineering approach have become an
efficient methodology for knowledge engineering
and computing in many domains [21, 22, 23]. Ontology is the foundation of describing a domain of interest. It consists of a collection of terms organized in a hierarchical structure that shape the reality. The components of ontology are, the following: concepts, relations between concepts, properties, and attributes. Ontologies are now ubiquitous in many information-systems enterprises. They constitute the backbone for the semantic web and all of the electronic-activities services (e.g. e-Government, e-Learning, e-Health, e-Business, etc.) [24, 25]. Ontologies' usage in information systems may be approached from various points of view: as a common vocabulary for multi-agent systems, as a chain between heterogeneous systems, ontologies for information resources sharing or for sharing data and ontologies used to mediate the search of specific materials on the Internet.

At our research Labs, a breast cancer ontology was developed [26]. The developed ontology was encoded in OWL-DL format using the Protégé-OWL editing environment. 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 four sub classes; Diseases, Medical_Interventions, Pathological_Category, References, and “People” which has two 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 5. 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 in figure 5.

8 Conclusion and Future Research Directions
The fusion of computational intelligence and knowledge computing paradigms with conventional knowledge acquisition from subject matter experts solves many of the problems in recent generation of healthcare knowledge-based systems. These paradigms play a key role in developing smart and robust tools for medical and healthcare tasks. Case-based reasoning methodology seems best suited for the core process of knowledge management because cases can represent knowledge well in terms of knowledge creation, storage and retrieval. Ontological engineering is an effective methodology for accumulating, representing, management, changing and updating knowledge in intelligent systems. Rough sets theory provides a powerful foundation to reveal and discover important structures and hidden patterns in big data and to classify complex objects. Currently we have three research projects in the areas of biomedical and health informatics. The first is dealing with the applications of genetic algorithms for developing an efficient classifier for thrombosis disease prediction. The second is dealing with visualization techniques in biological data mining to perform clustering tasks. The third is concerned with the using ontological engineering in developing smart medical Learning systems that provide students with intelligent browsing and searching support in their requests for relevant material on the web in the medical topics. In the future, we have two directions of further research (a) developing a web-based lesson planning system and (b) using ubiquitous computing environments with the intelligent wireless sensors, ad-hoc and mesh networks in teaching and learning process.

References:
Fig. 1 Areas of Research in Knowledge Computing and Engineering
Health Informatics

Intelligent Medical Informatics

Medical Education / Training
Medical Imaging
Medical Knowledge Engineering

Robotic Surgery

Medical Sciences + Informatics + Artificial Intelligence

Medical Expert Systems
- Machine Learning for Personal Authentication using Palm Vein.
- Visualization algorithms for Orthopedic Surgery.
- Machine Learning for Detecting Brain Tumors from Magnetic resonance imaging.
- Genetic algorithms classifier for predicting thrombosis disease.
- Visualization algorithms for medical data mining.
- Mobile system for cancer management using CBR.
- CBR Approach for Processing for Emergency medical responders.

Medical Education:
- Web-based lesson planning system based on CBR.
- Using ubiquitous computing with intelligent wireless sensors, ad-hoc, and mesh networks in teaching and learning processes.
- Virtual reality technologies in the health care, and medical education.
Fig. 4 Phases of knowledge discovery process

Fig. 5 Breast Cancer Ontology [26]