A Cognitive Virtual Microscopic Framework for Knowledge-based Exploration of Large Microscopic Images in Breast Cancer Histopathology

Ludovic Roux, Adina Tutac, Nicolas Loméni, Didier Balensi, Daniel Racoceanu, Antoine Veillard, Wee-Kheng Leow, Jacques Klossa and Thomas C. Putti

Abstract—Histopathological examination is a powerful method for prognosis of major diseases such as breast cancer. Analysis of medical images largely remains the work of human experts. Current virtual microscope systems are mainly an emulation of real microscopes with annotation and some image analysis capabilities. However, the lack of effective knowledge management prevents such systems from being computer-aided prognosis platforms. The cognitive virtual microscopic framework, through an extended modeling and use of medical knowledge, has the capacity to analyse histopathological images and to perform grading of breast cancer, providing pathologists with a robust and traceable second opinion.

I. INTRODUCTION

In the last decade, histopathological examination has been widely accepted as a powerful method for prognosis in major diseases such as breast cancer. Currently, analysis of medical images in histopathology largely remains the work of human experts. For pathologists, this consists of hundreds of slides examined daily. Such a tedious manual work is time-consuming, sometimes inconsistent and subjective.

The cognitive microscopic proposes a radical change in medical practice by introducing a knowledge-driven medical imaging environment, enabling safer decision-making for histopathology.

In section II, we introduce breast cancer prognosis using consistent breast cancer grading through robust second opinion. Section III presents the state of the art in virtual microscopy and in ontologies, and our cognitive microscopic framework is described in section IV. Section V concludes our study, by synthesizing the future researches.

II. PURPOSE

The cognitive microscope system integrates dynamic capture, representation and use of specific visual medical knowledge, in order to setup the bases of a knowledge-driven medical imaging. It helps at providing an effective, efficient, reliable and traceable assistance for prognosis. All the time-consuming and repetitive tasks are performed automatically and a second opinion is provided to the doctor who, at last, makes the final prognosis. Pervasive adaptation and a continuous learning of implicit medical knowledge are designed to maintain the quality level of the analysis performed by the system.

Apart from the use for safe and robust decision support, the huge amount of medical information captured and represented by the cognitive microscope constitutes a source of new knowledge for pathologists and medical researchers. Thus, with the cognitive microscope, the pathology prognosis accuracy is enhanced by integrating earlier stage diagnostic modalities including mammography, breast ultrasound imaging, and clinical information (such as history and physical examination) as well as a knowledge-driven prognosis support (such as similar cases, similar virtual slides, expert procedure and method, modeling and simulation of expert diagnosis).

Defining the outline of a generic cognitive approach for visual exploration and analysis of histopathological images, this general and ambitious objective focuses on the domain of breast cancer grading.

III. STATE OF THE ART

A. Existing Virtual Microscope Platforms

Virtual slide (or Whole Slide Images, WSI) systems use a client-server architecture. Most systems are able to work either as standalone or as connected systems. To facilitate access to any image sub-region at the required resolution, the different vendors use a tiled organization, stored in a pyramid where each level contains pre-computed images at lower resolution (e.g. 1/4th, 1/16th . . . ) [1]. Also few systems use JPEG2000 single file system that provides direct access to any sub-region at the desired resolution and at the desired bit rate/quality. Such sophisticated systems need implementation of an additional server for delivering the needed data stream. In compensation, these systems require less space for storing
compressed images and allow an additional benefit while storing three-dimensional data. Finally, such format has less compression artefacts at the tile boundaries, which is more adequate for image analysis processes.

Some platforms are dedicated to a specific WSI format (generally the one provided by the slide scanner vendor) while other platforms are able to handle different formats (tiled format or JPEG2000 format). In a survey published in 2006 [2], a comparison of 31 digital slide commercial systems was reported. These systems allow pathologists to browse the images as if they were using a real microscope (multi-scale virtual browsing) and to perform “slide conferencing”, that is synchronized cooperative sessions. It is possible to annotate images. Several systems also provide image analysis tools (segmentation, classification) for detecting regions of interest, cells, nucleus or membrane.

It exists also a few free digital slide systems designed mainly for education purpose. They allow building microscopic image libraries, providing a tool to navigate through a virtual slide image and to annotate images manually. The most impressive of these tools is WebMicroscope1 [3]. Images of the scanned slides are stored in a database on a server connected to the Internet. On the client side, a plug-in installed on standard web browser allows the user to query the server for a slide, to browse it and to annotate it.

None of these systems are designed to be computer-aided prognosis platforms because they lack the capacity of effective knowledge management.

B. Cognitive Virtual Microscopy

Focusing on histopathology and dealing with microscopic images of breast cancer tissue, the cognitive microscope continues the efforts developed by the medical device manufacturers in the field of virtual microscopy for histopathology and cyto-hematoLOGY. Cognitive virtual microscopy is an emerging field. Adding knowledge to virtual slides is for instance done by the Center for Neuroscience, University of California, USA, for the purpose of research and teaching students the brain structure and function [4]. Thus, their system BrainMaps.org is an interactive high-resolution digital brain atlas and virtual microscope that is based on scanned images of serial sections of both primate and non-primate brains. However, different from our cognitive microscope, their system is not intended to support medical decision-making. Our system does not rely only on anatomical descriptions. It also has reasoning capability.

One of the most advanced research team in virtual microscope is with Dr. Ronald S. Weinstein (Health Sciences Center, The University of Arizona, Tucson, USA) that established the Rapid Breast Care Service to provide a woman who has a positive digital mammography study with the results of her laboratory analysis the same day [5]. DMetrix company, in connection with this research team, has developed a digital image archiving system called the “Arizona” system [6]. Users are able to access image data and metadata through a secure Web site. However, this highly efficient system does not use expertise modeling for having a knowledge enhanced virtual microscope.

C. Ontologies

Domain knowledge-based representation evolved significantly within last years, with the ontologies’ coming in the semantic web world. Ontologies emerged from the Artificial Intelligence field [7] into medical realm where making explicit assumptions and sharing common understanding of the domain knowledge became a self-evident issue due to consensus inconsistencies amongst experts [8], [9]. Hence, various knowledge-modeling approaches have been proposed, ranging from comprehensive, standardized vocabularies such as UMLS2, SNOMED-CT3, NCI4, and reference ontologies (e.g. Open Biomedical Ontologies), to specific application-oriented ontologies (FMA-RadLex) [10].

An evidence-based solution to assist in establishing a consensus in breast pathology domain is given by [11], where SNOMED-CT terminology resource and available diagnostic classification are used as a basis for building an ontology of morphological abnormalities (e.g. ductal carcinoma in situ). Semantic similarities between abnormality descriptions are computed to help in finding agreement. For lung pathology, [12] proposes formalised medical reports using UMLS concepts represented in Web Ontology Language (OWL), in order to have semantic-based retrieval for text and image data in digital virtual microscope environment. However, due to the limitations of OWL in expressivity, rules support for ontology has been pursued in a more rigorous way, dating back to Description Logics. It is a current research topic in semantic web [13].

IV. OUR METHOD

The cognitive microscope is a synergy between knowledge, context, cognition and experience based on a user-centred approach to provide visual prognosis assistance to pathologists. The key features of the cognitive microscope platform are pervasive adaptation, context-awareness, human-computer confluence, modularity and openness.

A. Medical Knowledge Representation / Ontologies

In this section, we present a combined scientific-clinical perspective of domain ontology-based knowledge modeling with focus on breast cancer grading application.

Although there are some similarities in terms of using textual description and visual data from slide analysis, or reference ontology support, what we propose is different. One asset is the cognitive virtual microscopic framework where knowledge-modeling plays a key role not only for semantic annotation and retrieval, but also for semantic exploration and compression of large microscopic images. Another novel
direction represents the modeling of Nottingham Grading System (NGS), since histopathological grading of breast carcinoma has become a highly relevant assessment tool for prognosis in modern pathology [14].

There are various ways of designing an ontology. However, no single right methodology has been acknowledged. Therefore, we adjust the principles of Uschold and Grüninger [15] to fit our specific application and we identify a three-phase procedure: Knowledge acquisition, knowledge translating and knowledge refining. A top-down development process starting with the definition of the most general concepts in the domain is subsequently followed by specialization of the concepts and properties adding. Fig. 1 illustrates the knowledge modeling procedure.

1) Knowledge acquisition: As an important prerequisite, we define the domain as representation of Breast Cancer Grading (BCG) and we plan to use this ontology to improve the intra- and inter-observer agreement when it comes to sharing and finding harmony of prognosis opinion among pathologists. Based on the hypothesis that ontologies are mirrors of the real world and that formal ontologies can bring significant benefits for the development and maintenance of application ontologies, we use National Cancer Institute (NCI) thesaurus to feed our own application ontology with the concepts related to the breast cancer grading. Nevertheless, NCI does not handle all specific concepts for grading. Only a small subset of concepts was selected from NCI thesaurus and transmitted to our ontology (e.g. generic concepts such as Disease, Patient and some specialized concepts DuctalCarcinomaInSitu, CancerPatient). For the concepts related to the histopathological grading, like Assessment, TubuleFormationScoring, we add knowledge provided by pathologist experts to adjust the medical representation we target at. We use Protégé ontology editor and knowledge-based framework for our breast cancer grading modeling. In this way, the grading of breast carcinoma could be integrated into upper-ontologies of breast pathology.

2) Knowledge translation: This phase unfolds into two steps, essentially defining a hybrid ontology made up of structural and rule components: Structural modeling (structural component) using Web Ontology Language – Description Logics (OWL-DL)\(^5\) and rule modeling (rule component) with Semantic Web Rules Language (SWRL)\(^6\) which provides a high level of expressiveness and reasoning as described in [16]. OWL-DL has its foundations in SHOIN(D) formal language, enabling us to define concepts, relationships between concepts and to perform reasoning. We use SWRL for more accurate descriptions, such that reflection of reality could be achieved.

We construct the knowledge-base (KB) by defining the pair \(KB = (T, A)\) which contains the terminology (TBox) and the assertions (ABox).

The TBox comprises the set of axioms that formally describes the domain structure. Here, concepts (classes) and relations between concepts (properties) are defined. The main roots of the class hierarchy consist of three main primitive classes ConceptualEntity, DescriptorEntity and MicroscopicEntity, each of them specializing in related concepts. ConceptualEntity class contains structured information with respect to persons (Person) that are patients (Patient) having a disease (Disease) which need to be assessed (Assessment), using specific criteria (Criterion) of NGS (NottinghamGrading) by a specimen (Specimen) analysis. MicroscopicEntity handles NGS relevant objects identified on a specimen analyzed under the microscope, like cell, nucleus, mitosis, tubule, lumina. DescriptorEntity represents the formalisation of visual low-level features in terms of spatial (TopologicDescriptor such as Inclusion) and geometry attributes (Size, Shape, Intensity, etc).

\(^5\)http://www.obitko.com/tutorials/ontologies-semantic-web/owl-di-semantics.html

\(^6\)http://www.w3.org/Submission/SWRL/
With this image analysis information representation linked to the semantic level description, complex modeling is carried out. From the structural modeling point of view, all sibling classes are made disjoint, and is-a relationship stands as the basis for the class hierarchy construction.

In the ABox, individuals of classes are instantiated for example, for the Slide/Frame classes - SlideID or FrameID can be introduced and connected through a relationship with the PatientID.

Adding the medical definition/rules to the aforementioned concepts is done in terms of condition specifications, which contain property assertions in order to encapsulate a definition of the reality. Properties could as well have sub-properties specialization. For instance, hasTopologicDescriptor has $\text{surroundedBy}$ as sub-property, and its functional inverse property $\text{isSurrounding}$. We focus here only on one important characteristic, yet difficult to understand, of modeling in OWL-DL: Open World Assumption with negation restriction. The Mitosis definition provides an appropriate example (see Fig. 1), stating that a mitosis is any nuclei division that amongst other things is not included in some Tubule and hasLuminance some VeryLow and hasEccentricity hasValue MitosisEccentricity ranging between $[0, 1]$. At this point, to define the MitosisEccentricity we interact with the SWRL and the definition is represented according to the SWRL syntax. Similarly, the rules for scoring and grading are defined (see Fig. 2 for an illustration). Although hybrid modeling offers an increase of expressivity, the trade-off is the decidability. An excellent discussion regarding expressivity versus complexity and decidability is given by [16]. However, an OWL-DL + SWRL combination is suitable for modeling, under certain circumstances of computational expectations.

3) Knowledge Refining: Interactivity is essential for proving an accurate formalization. Therefore, we revise and refine the ontology, considering both computer vision and clinical perspective. Based on the principles in [17], we use the tableau-based algorithm of Pellet reasoner for our application ontology, due to its powerful advantage to support high expressivity and rules in SWRL. Although there are multiple reasoning tasks with regard to concepts, TBox, ABox or all KB, the tableau-based paradigm essentially reduces all inference problems to concept subsumption which practically verifies concept satisfiability using De Morgan rules with Negation Normal Form (NNF). In formal logics, a concept $C$ is satisfiable wrt $T$ iff there is an interpretation $I$ of $T$ such that $C^I \neq 0$. Hence, $I$ is called a model of $C$.

Given the definition of $\text{Nucleus}$ and the instantiation of $\text{Nucleus}$ in $\text{Nucleus}_I$, the question is whether we can subsume $\text{Nucleus}_I$ is a $\text{MicroscopicEntity}$. Fig. 3 provides the result of the subsumption test.

By testing concept subsumption for all concepts from the ontology, the KB consistency is thus verified (satisfies the coherence postulate). This reasoning task represents the central element of reasoning and is involved in computing the implicit knowledge from the explicit knowledge (the inferred model of ontology from the asserted model) to obtain the operational knowledge used further in the exploration from which results the pervasive implicit knowledge. The latter is then integrated in the operational knowledge thus providing a complex refined ontology. OWLViz enables the visualization of the conceptual graph, which represents the hierarchy built in the ontology (see Fig. 4).

At this point, an intervention from medical side is also required when axioms need to be refined / completed or rules added. The evaluation of the ontology is however, a complex process and depends on a lot of factors [18], thus it is still ongoing. An ontology fragment is publicly available at http://ipal.i2r.a-star.edu.sg/project_MICO.htm.

B. Semantic Compression

All the histopathological cases submitted to the system for analysis are stored, together with their analysis and grading, and be available for future information retrieval. To limit the quantity of information to be stored to a manageable amount, a semantic compression of histopathology images is
performed. For instance, only the regions of interest denoted by the analysis to be relevant for the pathology being studied are stored at the highest magnification. All the other parts of the slide are stored at a low magnification level.

C. Image Reasoning

The cognitive microscope combines visual perception, context, cognition and experience to reinforce a visual prognosis assistance following an approach centred on pathologist behaviour. It is intended to be capable of cognitive functions ranging from active perception to symbolic description, from symbolic manipulation to communication, and from an ad hoc approach to a capitalisation of previous experiences. The cognitive microscope behaves in consistent way with standard medical practice, following a uniform representation of image analysis, reasoning and context elements. This transparency throughout the whole process is exploited in order to permit confluence between the user and the platform, and medical validation of the results through technologies of content-based semantic image retrieval. The platform will be fully configurable and operable through a knowledge explorer allowing, in an incremental way, new medical knowledge capture, representation, use and discovery.

Multi-scale analysis and description of the slide is performed automatically by the system. Visual content of the slide is captured, analyzed and described dynamically in an active cognitive process. Descriptions produced by the analysis are stored and can be used at a later stage for validation purpose.

The challenge is to provide pathologists with a pertinent robust second opinion breast cancer grading, by detecting scale-dependent meaningful regions of interest in the microscopic biopsy images, annotating them, analyzing them and providing a breast cancer grade and validation capabilities. The cognitive microscope improves robustness and stability of grading by providing the pathologist with hints of a first automatic analysis as a second opinion from which to establish a final prognosis. In addition, it will be a suitable innovative working environment system with, until now, no competitor in the e-health market.

Fig. 5 shows an example of our prototype. Ontologies for breast cancer grading drive the image processing algorithm. All the visual elements detected in the image are given a name and are linked to a structural element defined in the ontology as shown in Fig. 5(a). On the virtual slide browser (see Fig. 5(b)), the visual elements are annotated by their name. There is a direct link between the slide browser and the ontologies.

The cognitive microscope is intended to be useful for pathology research. With the cognitive microscope, virtual slide images will be effectively using dynamic knowledge bases, rules and medical ontologies. This combination of images and knowledge would also enhance the information provided by the image annotation/analysis supporting new medical knowledge discovery from huge medical image and related information databases.

V. CONCLUSION

A first prototype has been developed. It can perform multi-scale image analysis using medical knowledge, breast cancer grading, semantic indexing, and validation through CBIR (content-based image retrieval).

We modeled a breast cancer grading application ontology with Protégé OWL-DL + SWRL, which is a different approach from our previous one [19]. The knowledge base contains axioms and assertions to capture the domain knowledge and we enriched the representation with rules. The formalization of the knowledge is highly needed to be able to perform reasoning. Thus, we evaluated the ontology using Pellet reasoner and refined it according to the clinical feedback.

Reasoning in ontologies and knowledge-bases is the cue rationale for a domain specification to be formalised. A reasoner has the capability to infer facts not explicitly represented in the ontology. In this light, description logics initiative was designed for tractable reasoning and automated processing. Therefore, our approach enables a future integration of new knowledge discovered when mapping the high-level representation of concepts with low-level image processing.

We envision that knowledge modeling for visual exploration of images in breast cancer grading domain using this approach opens new avenues for scientific and clinical applications. The idea to integrate it into a virtual microscope platform to assist in research as well as to assess more accurate prognosis is a significant improvement. The next step will be the extension of such knowledge modeling to both morphological and functional data at the cellular and at the molecular levels which is the key for transmodality correlation (clinical/imagery/biology/microscopy) and new intrinsic markers discovery.
(a) Ontologies for breast cancer grading have been instantiated by visual entities detected in the slide.

(b) Virtual slide: The detected visual elements are annotated by the system.

Fig. 5. Virtual slide and ontologies of our cognitive microscope prototype.

ACKNOWLEDGMENTS

This study has been partially supported by the MMedWeb grant A*STAR SERC 052 101 0103 (NUS R-252-000-319-305).

REFERENCES


