Automatic classification of cancer tumors using image annotations and ontologies

International Symposium on Computer-Based Medical Systems, 28th, 2015, São Carlos e Ribeirão Preto.
http://www.producao.usp.br/handle/BDPI/49181

2015 IEEE 28th International Symposium on Computer-Based Medical Systems

Automatic Classification of Cancer Tumors using Image Annotations and Ontologies

Edson F. Luque*
Department of Computer Science
University of São Paulo
São Carlos, Brazil
Email: *edluquem@usp.br

Daniel L. Rubin†
Department of Radiology
Stanford University
Stanford, USA
Email: †dlrubin@stanford.edu

Dilvan A. Moreira‡
Department of Computer Science
University of São Paulo
São Carlos, Brazil
Email: ‡dilvan@usp.br

***Abstract*—Information about cancer stage in a patient is crucial when clinicians assess treatment progress. Determining cancer stage is a process that takes into account the description, location, characteristics and possible metastasis of cancerous tumors in a patient. It should follow classification standards, such as TNM Classification of Malignant Tumors. However, in clinical practice, the implementation of this process can be tedious and error-prone and create uncertainty. In order to alleviate these problems, we intend to assist radiologists by providing a second opinion in the evaluation of cancer stage in patients. For doing this, Semantic Web technologies, such as ontologies and reasoning, will be used to automatically classify cancer stages. This classification will use semantic annotations, made by radiologists (using the ePAD tool) and stored in the AIM format, and rules of an ontology representing the TNM standard. The whole process will be validated through a proof of concept with users from the Radiology Dept. of the Stanford University.**

***Index Terms*—OWL, SWRL, cancer staging, ePAD, cancer**

I. Introduction

In radiology and oncology, evaluating the response to cancer treatments depends critically on the results of image analysis by experts. However, the information obtained from this analysis is not easily interpreted by machines.

Medical images in clinical tasks are important as they allow specialists to diagnose, plan and track patients [1]. Thus, a considerable number of computer applications, aimed at this medical field, have been developed. Most of them are focused on extracting visual features with the help of image processing algorithms.

Although these algorithms can help physicians process image contents for cancer treatment, they have problems when a query is made in the context of cancer patient classification. For instance, if an oncologist wants to know if a tumor is close to spread to any region near the origin of the cancer, but not for other parts of the body (a specific cancer stage) [2]. The algorithms have difficulties during image interpretation, because the semantic information, implicit in the image reports, is not accessible to them.

II. Objective

The objective of this work is to automatically determine the cancer stage of lesions, present in medical images, using semantic web reasoning tools to process semantic annotations made by radiologists to provide clinicians a second opinion on the classification of their patients. This semantic annotations are made using tools, such as ePAD, and saved in the AIM format. Automatic cancer staging can increase the efficiency of radiologists and oncologists and improve the quality and uniformity of image interpretation by experts.

In the case of this research, knowledge for determining cancer stage will be encoded in ontologies (OWL) and rules (SWRL).

III. Related Work

Currently, there are other cancer staging systems, however, they are not open source and their classification methods cannot be analyzed or reused openly. In addition, they also have accuracy problems and contradictions in the results of staging[3].

Cancer staging is a classification process to determine how much cancer there is on the whole body and where it is located. Prior research on cancer staging have used semantic annotations, from a controlled vocabulary, for discovering implicit knowledge:

- Levy et al. [1] developed a methodology to transform the AIM information model[1] in OWL and SWRL in order to automatically classify and calculate tumor burden.
- The work proposed by Zillner et al. [4], [3] uses medical image annotation and reasoning (spatio-anatomical reasoning [5]) technologies in order to automatically classify patients with lymphoma.
- In Meriem et al. [6], the authors propose a methodology to improve the clinical model that performs the score of breast cancer, based on the Nottingham Grading System (NGS). They designed an OWL-DL ontology and SWRL rules based on histopathological images annotations in WFML[2].
- Dameron et al. [7] presented an OWL ontology for automated TMN classification. However, they did not use it to do classification based on image annotations. Their ontology may be reused by us.
- In Racoceano et al. [8], the authors describe a prototype that controls an entire histological image analysis

[1] https://wiki.nci.nih.gov/display/AIM/Annotation+and+Image+Markup+-+AIM

[2] A XML language produced by TRIBVN for its ICS framework

2372-9198/15 $3.00 © 2015 IEEE
DOI 10.1109/CBMS.2015.83

protocol developed in MICO[3] in order to improve the Whole Slide Image (WSI) analysis protocol and become a reliable assessment for breast cancer classification. The prototype consists of two main components, the semantic core and image processing algorithms.

These works use semantic annotations to make their observations consistent, explicit, and machine-accessible. Radiological tools such as the ePAD [9] have been implemented with the purpose of making these annotations facile to record. Recent works using ePAD [10], [11] propose an image retrieval framework based on semantic annotations. This framework incorporates the semantic correlations between terms used to describe these images. This automated approach provides real time support for radiologists, showing them images associated with similar diagnoses.

In the literature, we found similar systems where semantic annotations are stored in different formats that do not allow their integration for reasoning processes. Often these formats are also proprietary. Some of these studies also allow creating image annotations in AIM format (in XML), but these are not suitable for reasoning. AIM format provides only a transfer and storage format.

Our work is focused on helping cancer specialists in automatic patient classification (staging) using semantic annotations in images. The classification shall be made using semantic reasoning on annotations encoded in AIM, these annotations, made by radiologists, describe lesions in images.

IV. Methods

In this section we describe the methodology to be followed. It is divided in the following steps:

- Develop a mechanism to transform the semantic annotations from AIM documents, generated by ePAD, to instances in the OWL model.
- Develop (or extend) an ontology focused on the TNM classification criteria (Classification of Malignant Tumors).
- Integration of knowledge through reasoning. The reasoning process allows the classification of a patient. It integrates the knowledge captured by the annotations and medical ontologies. The ontological model will capture the individual cancer staging of each patient. Information about staging of patients in the database will be queried using SPARQL .
- Implementation of ePAD interfaces, in partnership with the Dept. of Radiology, Stanford University.
- Evaluation of results with radiologists / oncologists of the Dept. of Radiology, Stanford University.

V. Conclusion and Justification

The process of cancer staging in patients by image analysis is a task performed by specialists, such as oncologists and radiologists, that often entails intensive work that requires precision in the interpretation of cancer lesions. Expert accuracy is achieved through training and experience[12], but variations in image interpretation is a limitation of human observers. In this context, the development of an automatic classification system is a strong medical need. It can help experts obtain a higher accuracy rate in interpretation.

Moreover, although systems, such as ePAD, enable the creation of image annotations (in the AIM format), they do not represent them in a format that is directly suitable for reasoning. AIM is a format for data transfer and storage. Other systems, such as Mint Lesion (MintMedical GmbH, Dossenheim, Germany) and syngo.via (Siemens Healthcare, Malvern) are commercial software where all the image data are stored internally in a proprietary format that cannot be accessed by third parties [9].

There is currently a lack of semantic reasoning methods to make inferences about cancerous lesions from semantic annotations. Thus, the main motivation for this work is the possibility of developing reasoning methods, based on cancer staging standards such as TNM, for the automatic staging of cancer patients (using AIM image annotations) and incorporate them into open source image annotation systems, such as ePAD.

This work is partially funded by the National Council for Scientific and Technological Development - CNPq and CAPES.

[3]http://www.ipal.cnrs.fr/?q=project/mico-cognitive-virtual-microscopy

References

[1] M. Levy, M. J. O'Connor, and D. L. Rubin, "Semantic reasoning with image annotations for tumor assessment." *AMIA Annual Symposium proceedings*, vol. 2009, pp. 359–63, Jan. 2009.

[2] P. Wennerberg, K. Schulz, and P. Buitelaar, "Ontology modularization to improve semantic medical image annotation." *Journal of biomedical informatics*, vol. 44, no. 1, pp. 155–62, Feb. 2011.

[3] S. Zillner and D. Sonntag, "Image metadata reasoning for improved clinical decision support," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 1, no. 1-2, pp. 37–46, Mar. 2012.

[4] S. Zillner, "Towards the ontology-based classification of lymphoma patients using semantic image annotations," in *SWAT4LS'09*, 2009, pp. –1–1.

[5] M. Manuel, D. Sonntag, and P. Ernst, "A Spatio-Anatomical Medical Ontology and Automatic Plausibility Checks."

[6] B. Meriem, T. Yamina, and A. D. Pathology, "Interpretation breast cancer imaging by using ontology," no. I, pp. 1–6, 2012.

[7] O. Dameron, É. Roques, D. Rubin, G. Marquet, and A. Burgun, "Grading lung tumors using owl-dl based reasoning," in *9th International Protege Conference - Presentation Abstracts*, 2006, pp. 69–72.

[8] D. Racoceanu and F. Capron, "Towards semantic-driven high-content image analysis: An operational instantiation for mitosis detection in digital histopathology," *Computerized Medical Imaging and Graphics*, Oct. 2014.

[9] D. L. Rubin, D. Willrett, M. J. O'Connor, C. Hage, C. Kurtz, and D. A. Moreira, "Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials," *Translational Oncology*, vol. 7, no. 1, pp. 23–35, Feb. 2014.

[10] C. Kurtz, A. Depeursinge, S. Napel, C. F. Beaulieu, and D. L. Rubin, "On combining image-based and ontological semantic dissimilarities for medical image retrieval applications." *Medical image analysis*, vol. 18, no. 7, pp. 1082–1100, Jul. 2014.

[11] F. Gimenez, J. Xu, Y. Liu, T. T. Liu, C. F. Beaulieu, D. L. Rubin, and S. Napel, "On the Feasibility of Predicting Radiological Observations from Computational Imaging Features of Liver Lesions in CT Scans," *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, pp. 346–350, Jul. 2011.

[12] A. Depeursinge, C. Kurtz, C. Beaulieu, S. Napel, and D. Rubin, "From Radiological Image Data : Preliminary Results With Liver Lesions in CT," vol. 33, no. 8, pp. 1669–1676, 2014.