ORIGINAL PAPER

Automatic Staging of Cancer Tumors Using AIM Image Annotations and Ontologies

E. F. Luque[1,2] · N. Miranda[1] · D. L. Rubin[3] · D. A. Moreira[1] (ORCID)

Abstract

A second opinion about cancer stage is crucial when clinicians assess patient treatment progress. Staging is a process that takes into account description, location, characteristics, and possible metastasis of tumors in a patient. It should follow standards, such as the TNM Classification of Malignant Tumors. However, in clinical practice, the implementation of this process can be tedious and error prone. In order to alleviate these problems, we intend to assist radiologists by providing a second opinion in the evaluation of cancer stage. For doing this, we developed a TNM classifier based on semantic annotations, made by radiologists, using the ePAD tool. It transforms the annotations (stored using the AIM format), using axioms and rules, into AIM4-O ontology instances. From then, it automatically calculates the liver TNM cancer stage. The AIM4-O ontology was developed, as part of this work, to represent annotations in the Web Ontology Language (OWL). A dataset of 51 liver radiology reports with staging data, from NCI's Genomic Data Commons (GDC), were used to evaluate our classifier. When compared with the stages attributed by physicians, the classifier stages had a precision of 85.7% and recall of 81.0%. In addition, 3 radiologists from 2 different institutions manually reviewed a random sample of 4 of the 51 records and agreed with the tool staging. AIM4-O was also evaluated with good results. Our classifier can be integrated into AIM aware imaging tools, such as ePAD, to offer a second opinion about staging as part of the cancer treatment workflow.

Keywords Reasoning · Image annotations · Cancer staging · ePAD · SWRL · TNM

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 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 to process image findings for cancer treatment, they have problems when an abstract query is made in the context of cancer patient classification, for example, when an oncologist wants to know if a tumor is at an advanced stage and it expanded to some region near the origin of cancer but not for other parts of the body [2]. There are difficulties during image interpretation, because the semantic information, implicit in the image reports, is not accessible to these algorithms.

Although medical images and reports provide a significant amount of information to physicians, this information is not easily integrated into advanced medical applications, such as clinical decision support systems to treat patients with cancer, specifically when physicians assess individual

✉ D. A. Moreira
  dilvan@usp.br

  E. F. Luque
  edluquem@usp.br

  N. Miranda
  nelson.miranda@hc.famema.br

  D. L. Rubin
  dlrubin@stanford.edu

1 Department of Computer Science, University of São Paulo, São Carlos, SP, Brazil

2 Universidad Nacional de San Agustín de Arequipa, Arequipa, Peru

3 Department of Biomedical Data Science, Radiology, and Medicine (Biomedical Informatics), Stanford University, Stanford, CA, USA

progress of a cancer patient to decide new treatment measures [3]. Appropriate treatment options are supported by information about cancer staging. Cancer staging is a classification process based on characteristics such as tumor size and location in the body. This classification process can be automated in order to optimize the work of physicians, which may become cumbersome and error prone as the number of patients increases [4].

There are few tools that allow radiologists to easily capture semantic structured information as part of their routine research workflow [5]. The Annotation and Image Markup (AIM) [6, 7] project, from the cancer Biomedical Informatics Grid (caBIG) [8], provides a XML schema to describe the anatomical structure and visual observations about images using the RadLex ontology [9]. It allows the representation, storage, and consistent transfer of semantic meanings in images. Tools using the AIM format, such as the ePAD [5] tool, can help to reduce the effort to collect structured semantic information about images. It also permits making inferences about this information (cancer lesions) using biological and physiological relationships between image metadata.

Image metadata, in AIM format, does not allow the representation of information about image findings in a format that is directly suitable for reasoning. AIM provides only a format for data transfer and storage. We can see then that there is a lack of semantic methods to make inferences about cancerous lesions from semantic annotations of images, based on standard formats (such as AIM). Thus, in this work, we developed a reasoning approach based on a staging system: the Tumor-Node-Metastasis (TNM) classification (stages, definitions, and examples can be found in [10] and Fig. 6 respectively), published by the Union for International Cancer Control (UICC).

Part of the novelty of this approach is to apply semantic methods for image-based reasoning to automate the reasoning tasks in TNM currently done by humans. This TNM classifier was evaluated using 51 actual patient's radiology reports (from The NCI's Genomic Data Commons). When compared with the stages attributed by physicians, the classifier had a precision of 85.7% and recall of 81.0%. Furthermore, 3 radiologists from 2 different institutions manually reviewed a random sample of 4 of the 51 records and agreed with the tool staging. We also compared semantic search, using the AIM4-O ontology, to keyword search in the task of searching patient reports. Semantic search had better precision and recall in all but one case.

A classifier, like the one we are proposing, makes sense only if formats for semantic image metadata (like AIM) are adopted more widely by imaging tools. In that sense, it is a glimpse of what semantic metadata could do for image processing.

We recognize that even though radiology imaging is a crucial component in determining cancer stage it alone may not be enough. Physicians may also combine it with EHR and laboratory/pathology data to reach a classification. Our goal is to show a working automated method for staging based on imaging that can be extended to include non-image info in the future. Ultimately, a tool based on this method will meet the requirements to be incorporated into semantic annotation tools for medical images, such as the electronic Physician's Annotation Device (ePAD) [8], to automatically stage cancer patients.

Objective

The objective of this work is to automatically determine the cancer stage of lesions present in medical images, using ontologies and reasoning technologies, to process semantic annotations made by experts and provide clinicians with a second opinion on the classification of their patients. These semantic annotations are made using tools that use the AIM format (used by the ePAD tool) to describe and save image findings. Automatic cancer staging can increase the efficiency of radiologists and oncologists and improve the quality and uniformity of image interpretation by experts. It is important to mention that our work focuses on staging liver cancer due to data availability.

This paper is organized as follows. The "Related Work" section presents related approaches. The "Method Description" section describes our methodology composed of three main components: the ontological representation of the AIM 4.0 model, the conditions to implement the TNM classifier (General Ontology), and the formal representation of cancer staging. The "Experimental Study and Results" section analyses experimental data to assess our TNM classifier relevance. Conclusions are found in the "Conclusion" section.

Related Work

Currently, in clinical research, there are similar cancer staging systems. Cancer staging is a classification process to determine how much cancer there is on the whole body and where it is located. Some efforts on tumor staging, based on a formal representation of a classification system (such as TNM), have been using semantic annotations from a controlled vocabulary for discovering implicit knowledge. However, they are not open source and their classification methods cannot be analyzed or reused openly [5].

Among the proposed approaches, Dameron et al. [11] and Marquet et al. [12] perform reasoning based on classification systems, such as TNM and WHO [13], using an ontology class–based reasoning approach. However,

this approach often leads to an underlying drawback: the creation of unnecessary classes, increasing the complexity of the ontology. In addition, they performed closed-world-based reasoning in a context of open-world assumption (OWA) [14] by modeling patient conditions using classes, avoiding the reasoning based on instances. However, it is possible to perform reasoning based on instances supported by data structures (that will be described in detail in following sections).

Some authors create ontologies in OWL-DL for TNM [15–22]. However, the idea of having an ontology for each type of body organ is undesirable, as in the case of Zillner et al. [19] and Tutac et al. [15]. We believe in the approach of having an ontology representing directly image findings (such as the ontology model of AIM) and, besides that, the classification tasks for cancer staging should be guided only by rules and axioms.

In three articles [16, 23, 24], the authors used semantic image annotations and perform a classification based on the Nottingham Grading System (NGS) supported by OWL and SWRL. The fact of creating a new ontology, depending on the conditions you want to analyze, is a limiting factor. It does not occur in our approach, which provides an ontology based on the AIM standard.

Möller et al. [25] and Zillner et al. [3] use the closest approach to our proposal. However, ignoring the fact that the data used are not available for all the necessary analysis, the lymphoma staging system that has been implemented in this study is relatively simpler than TNM staging system. For example, we can see that Zillner et al. [26] do not consider the size of a lesion as an important factor. However, this fact is very important in staging systems such as TNM for liver, lung, etc. Moreover, we can see that the process of aligning all ontologies generated in this study is not described explicitly.

Gimenez et al. [27] and Kurtz et al. [28] are recent works that use the ePAD tool. The authors propose an image retrieval framework based on semantic annotations. They used semantic correlations based on semantic terms that are used to describe medical image findings. Their automated approach helps radiologists by showing them images associated with a similar diagnosis.

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 the creation of image annotations in AIM format, but these are not suitable for reasoning. AIM 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 is made using semantic reasoning on annotations encoded in AIM and these annotations, made by radiologists, describe lesions in images.

Method Description

In this section, we describe our methodology. It is comprised by three main tasks:

1. Ontological representation of the AIM 4.0 model.
2. Creating conditions to implement reasoning based on TNM rules, using OWL instances.
3. Formal representation of cancer staging.

Ontological Representation of AIM Model

In order to perform inference and classify image annotations based on the AIM standard, we need a language equipped with formal semantic. Using this semantic, inferences about an ontology together with a set of individual instances of classes can be made. In this context, the Web Ontology Language (OWL), a language for building ontological representations of information models, was used. In our work, we transformed the AIM data model into an equivalent ontological representation, using OWL2. This transformation was performed by creating classes and properties in OWL that are user understandable and suitable for inference.

We developed an OWL model, based on the ontology provides by Bulu, Hakan and Rubin, Daniel L. [29][1], which represented an older version of AIM. Therefore, in order to represent the AIM 4.0 model, which is the version used to store image annotations generated using tools (such as ePAD [30]), we modified Bulu's ontology to represent AIM 4.0 model concepts. Our ontology is called **AIM4-O**.

AIM4-O Ontology

AIM4-O Classes

In general, the AIM 4.0 model is an extension of the AIM Foundation model. There are nine classes that capture lesion results and measurements derived during image-based clinical analysis. In this work, we considered the following six classes sufficient to achieve our goal:

1. **AIM:Entity:** It is an abstract class that represents the existence of a thing, concept, observation, calculation, measurement, and graphic design in AIM.
2. **AIM:AnnotationCollection:** It is an abstract concept of a container that collects elements such as

---

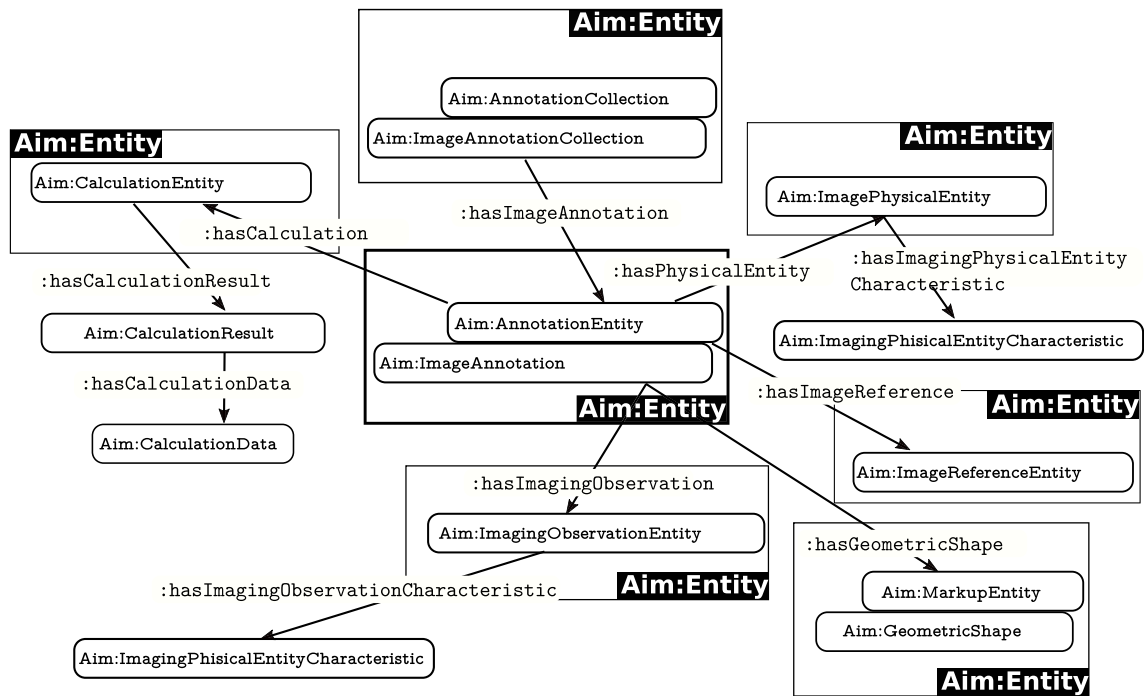[1] https://wiki.nci.nih.gov/display/AIM/Extending+the+AIM+Model

Fig. 1 This diagram shows the **AIM4-O** ontology including the six classes modified from Bulu's ontology. Ovals represent extended abstract classes and instances of **AIM:Entity**, the relationships are represented as arrows

AnnotationOfAnnotation or ImageAnnotation entities.

3. **AIM:imageAnnotationCollection:** Stores instances of the ***AnnotationImages*** class. It associates with the ***Person*** class that contains patient demographic information.

4. **AIM:ImagingPhysicalEntity:** This class stores an anatomical location as a coded term (i.e. RID2662, femur, RadLex) based on controlled vocabularies such as RadLex®, SNOMED CT®, or Unified Medical Language System (UMLS).

5. **AIM:ImagingPhysicalCharacteristic:** This class describes the ***ImagingPhysicalEntity*** as a coded term.

6. **AIM:MarkupEntity:** This class captures textual information and graphical representation as DICOM-SR and SCOORD3D, for tridimensional and bidimensional spatial coordinates.

AIM4-O Relationships

Relationships describe semantically how different concepts relate to each other. For example, an ***ImageAnnotation*** having an ***Observation*** entity that describes a lesion. These relationships in our ontology enable the semantic

reasoning, which is a prerequisite for semantic classification and searching.

One of the basic concepts in the AIM4-O ontology is the ***ImageAnnotation*** entity. It allows us to describe data properties of an image annotation in OWL, such as comments, name, date, and time of its creation. For instance, the statement ***imageAnnotation1 dateTime {2014-09-26T17:07:58^^datetime}*** says that an annotation, referred to as ***imageAnnotation1***, was created at ***2014-09-26T17:07:58***. There are more ***ImageAnnotation*** entity relationships to other concepts, such as physical location and observations on lesions found. These relations can be seen in Fig. 1.

These relations can also be specified using OWL relationships. A small example is given in Fig. 2. It includes an OWL representation of an AIM 4.0 Ontology instance and its usual relationships. This example shows semantically equivalent concepts between AIM-XML format and OWL model (Manchester serialization format) that were used in this work. For example, an AIM ***Image Annotation*** instance with the identifier value **''unique Identifier:9gs43xqj1kyl13l...''**, can be represented in OWL and used for reasoning purposes. Similarly, Fig. 3 shows that the concept ***PhysicalEntity***, with the identifier value **''uniqueIdentifier:g08jn**

Fig. 2 Sample of image annotations using AIM-XML and OWL Manchester serialization

m9ow79ti...'', is related to the **ImageAnnotation** instance by the **hasPhysicalEntity** Object Property. Figure 3 shows more information about the syntax of a **PhysicalEntity** instance in AIM 4.0 XML and the equivalent OWL Manchester syntax for the same instance.

Creating Conditions to Implement Reasoning Based on TNM Rules Using OWL Instances

The second step is to transform existing AIM-XML documents to their equivalent in OWL (using the AIM4-O ontology). To achieve this, we developed scripts, in the Groovy language.[2] First, we automatically map AIM-XML entities to AIM Java classes, based on the AIM UML[3] model (Fig. 4). We then create instances of the AIM4-O ontology from these AIM Java classes, using the OWL API, a Java API for creating, manipulating, and serializing OWL

ontologies. Finally, these instances populate a semantic web knowledge base. This base is suitable for classification-based and rule-based inference.

In order to automatically stage cancer, our approach must have the support of an ontology to specify the semantics of image observations from a particular domain. In this case, this ontology should be able to represent the topology of a human body organ (the organ in which cancer starts growing and has its own TNM system). In this work, this organ was the liver. Furthermore, it is considered necessary to include an OWL representation of RadLex®[31] vocabulary in order to facilitate handling AIM4-O individuals, because these individuals have Radlex terminology in their structure. Finally, a set of rules to do the actual staging, based on the TNM liver system, was added to the ontology. The final result was the **General Ontology**.

General Ontology

The General ontology is divided into 4 files, as seen in Fig. 5:

Fig. 3 Sample of AIM PhysicalEntity in AIM-XML and OWL Manchester serialization format

Fig. 4 Mechanism to transform AIM-XML documents to encoded AIM4-O individuals using OWL
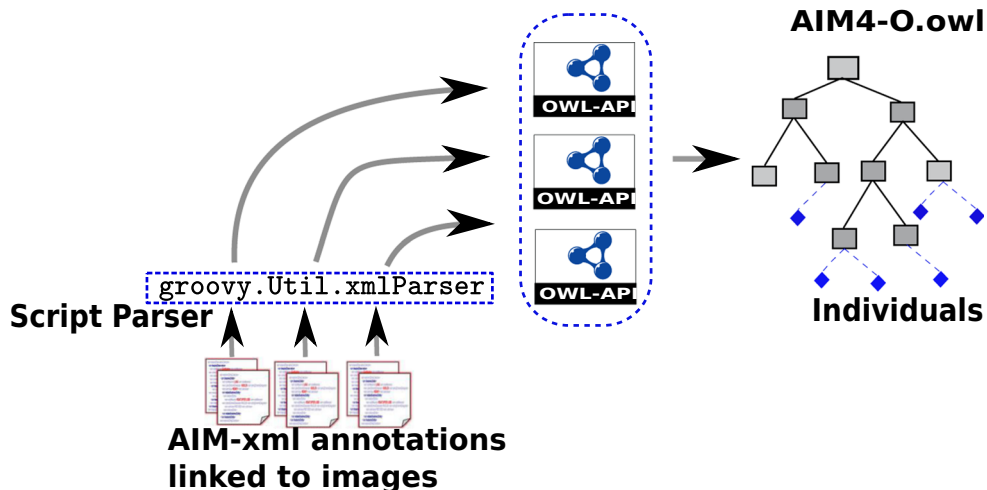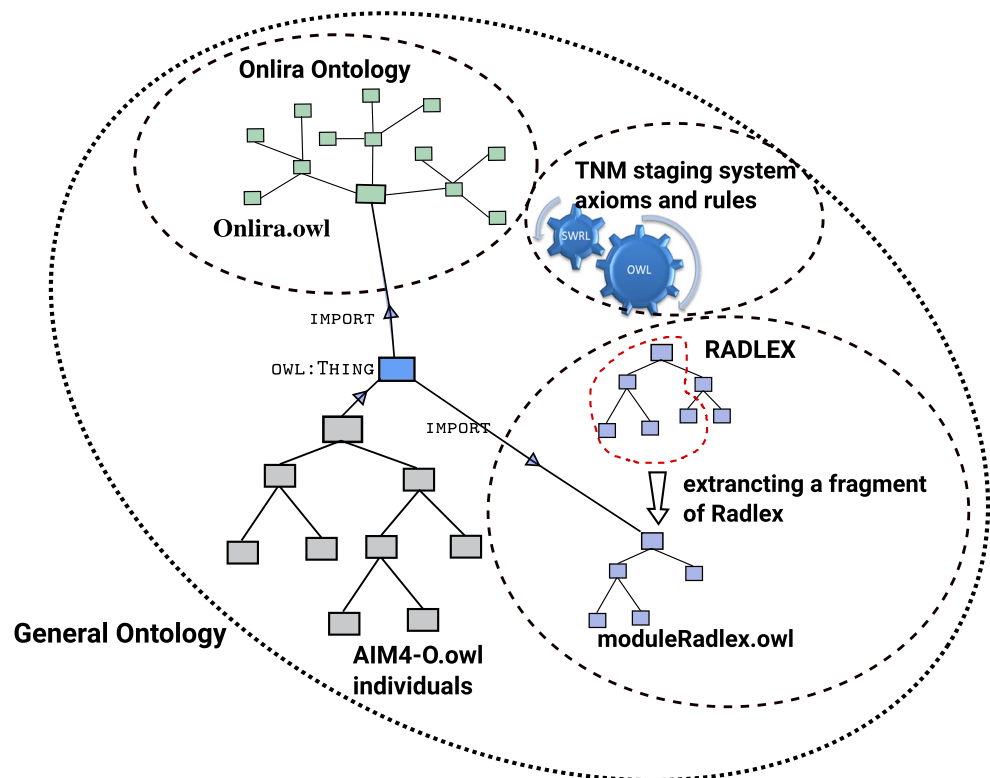
Fig. 5 The General Ontology imports (TNM rules, axioms, moduleRadlex.owl, and Onlira.owl) needed in order to classify liver cancer

- The AIM4-O ontology with individuals ("AIM4-O Ontology").
- Onlira.owl: The Liver Ontology (based on the Onlira ontology [32]).
- Radlex lexicon module (ModuleRadlex.owl[4]).
- General concepts of TNM (axioms and SWRL rules).

The following sections give a description of the ontologies used as part of the **General Ontology**.

Ontology of the Liver for Radiology

The Ontology of the Liver for Radiology (ONLIRA) ontology[5] was developed as part of the CaReRa project. It aims to model imaging observations of the liver domain with an emphasis on properties and relations between the liver, hepatic veins, and liver lesions. This ontology is used as an ontological representation of the liver and its topological features.

RadLex Terminology

The AIM model provides an XML schema that describes anatomical structures, visual observations and other information relevant to images using the RadLex terminology. We extracted a module from the Radlex lexicon to represent this information. The RadLex module is used by the General ontology to permit a formal representation (in OWL) of TNM criteria. The TNM criteria are based on knowledge about the way cancer develops and disseminates. For this reason, it is important that the General ontology represents not only the anatomical entities mentioned in TNM but also other direct and indirect related anatomical entities to consider the relative proximity between them. For example, to the N and M criteria (in TNM) we added 2 super classes, the ***adjacentOrganGroup***, which describes the set of organs adjacent to a main organ (e.g., the liver), and the ***noadjacentOrganGroup***, which describes organs based on the most common sites of tumor dissemination [10]. For the liver, we included lungs and bones as no adjacent organs, as seen in Fig. 8.

Classes to Represent TNM System Concepts

In order to create an OWL representation for each TNM stage, we had to interpret each stage definition. Although it is not mentioned explicitly, the TNM criteria are exclusive, so the corresponding OWL classes were made disjoint. For example, the ***T2*** stage is represented by two constraints: a single tumor (of any size) that has grown into blood vessels concept (***T2_a*** class) and a single tumor no larger than "x" cm concept (***T2_b*** class).

---

[4]http://bioportal.bioontology.org/ontologies/RADLEX/classes

[5]https://bioportal.bioontology.org/ontologies/ONLIRA

Table 1 American Joint Committee on Cancer/International Union against Cancer TNM classification system

Primary tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Solitary tumor (any size) without vascular invasion
T2 = T2_a or T2_b	Solitary tumor (any size) with vascular invasion or multiple tumors none < 5 cm
T3a	Multiple tumors, with at least one tumor > 5 cm
T3b	Single tumor or multiple tumors of any size involving a major branch of the portal vein or hepatic vein
T4	Tumors with direct invasion of adjacent organs other than the gallbladder or with perforation of visceral peritoneum
Regional lymph nodes (N)	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
Distant metastasis (M)	
M0	No distant metastasis
N0	Distant metastasis

Formal Representation of TNM Cancer Staging

In the previous sections, several steps were necessary to create the General ontology for a TNM classifier. First, we created classes and properties in order to fill the semantic gap between the tumor features and the AIM4-0 classes definitions. Then, we provided formal definitions for the TNM stages, liver's topological features, and RadLex terminology, in order to represent them in OWL. Finally, we defined formal mechanisms for reasoning (using only OWL and SWRL expressivity) such as OWL classes, intersections, equivalences, disjunctions between classes, and a set of rules in order to determine cancer stage from image annotations. This last step will be described below.

In order to discover the limits of the OWL concepts and SWRL rules, we attempted to formally define and implement the conditions that TNM staging demands. TNM cancer staging is divided in two main steps. The first step consists in giving a score starting from the description of the tumor (T), its spreading into lymphatic nodes (N), and possible metastasis (M) (see Table 1).

The second step consists in determining the stage according to the previous scores (see Fig. 6). To make the aforementioned tasks possible, we decided that the following conditions reflect a desirable staging process:

– Condition 1 : Staging should consider the existence of solitary or multiple tumors on the same site.
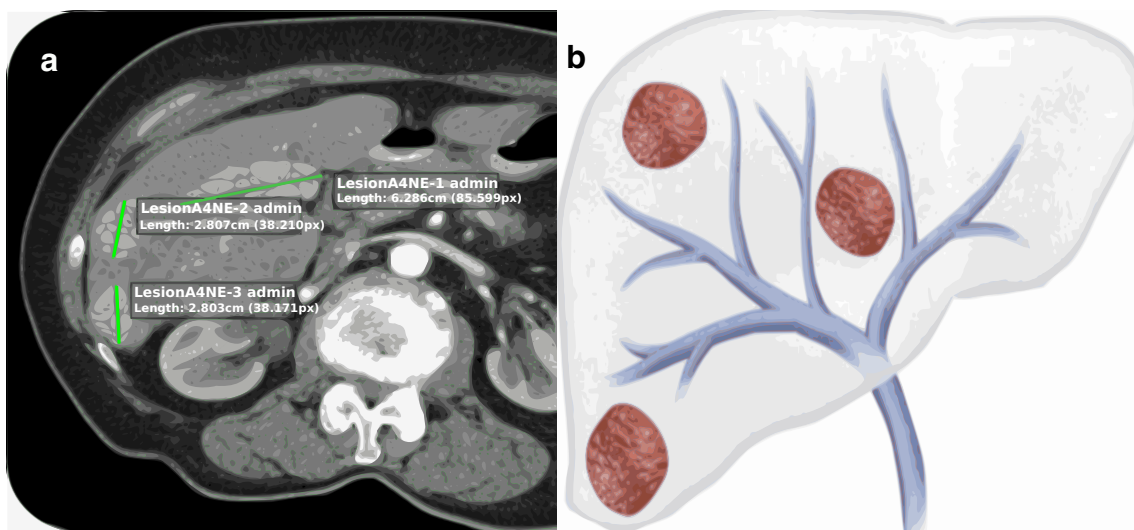

Fig. 6 **a** Axial, contrasted CT image shows multiple HCC tumors (green lines), identified and annotated using the ePAD tool. There was no regional lymph node involvement or metastasis. **b** The diagram shows multiple HCCs with at least one > 5 cm. This patient was classified as having TNM stageIIIA (T3a, N0, M0). Adapted from [10]

– Condition 2 : Staging should consider if tumors are either bigger or minor than a certain size in cm.
– Condition 3 : Staging should consider lesions in adjacent organs.

Asserting Conditions Using OWL

Condition 1: Staging should consider the existence of solitary or multiple tumors on the same site.

$$ImageStudy(?X) \land isImageStudyOf(?x, ?y) \land isImageStudyOf(?x, ?z) \land$$
$$DifferentFrom(?y, ?z) \rightarrow MoreThanOneTumor(?x)$$

This rule classifies an **ImageStudy** as a member of the **MorethanOneTumor** class if an image study "X" is referenced by more than one image annotation. In order to classify something as **MorethanOneTumor**, we created a new concept called **isImageStudyOf**. This concept is the inverse of the **hasImageStudy** object property. The **hasImageStudy** property relates an **ImageAnnotation** entity to an **ImageStudy** entity.

In the scenario of classifying patients with one solitary tumor, we did not find axioms or rules that satisfied this requirement, due to the fact that OWL works under the open-world assumption. Open world means that just because something is not said it does not mean that it is not true. For example, I can say that the patient annotation describes a cancer lesion, using the **ImagingObservation** entity of the AIM4-O ontology model, but unless I explicitly say that there are no other lesions, it is assumed that there may be other lesions that I just have not mentioned or described.

We have tried to solve this problem (the open-world assumption) by considering some alternatives such as modeling again our AIM4-O ontology (e.g., setting the **hasImagingObservation** object property as a primitive class). But, this did not seem intuitive to us. Instead, we decided to state the number of lesions explicitly by creating one new concept named **singleLesion**, as a data property of an **ImageStudy** entity. This concept

The AIM4-O ontology does not give us the explicit mechanics such as classes, subclasses, or properties, that allows us to infer whether a patient has a single or multiple tumors. In the case of multiple tumors, we constructed the following rule **MoreThanOneTumor** (in SWRL notation):

denotes if an **ImageStudy** describes exactly one solitary tumor. We assumed that an **ImageStudy** entity describes only one tumor (**"singleLesion {true}"**) if and only if it is referenced by only one **ImageAnnotation** entity. However, it was not possible to formulate this using only OWL. Instead, this information was provided by a data structure that was generated as part of the process of parsing the AIM-XML image annotations to create AIM4-O individuals. Finally, to classify annotations that describe a single lesion, we constructed the rule **SingleTumor** (in SWRL notation):

$$ImageStudy(?X) \land singleLesion(?X, ?val) \land$$
$$equal(?val, true) \rightarrow singleTumor(?X)$$

Condition 2: Staging should consider if tumors are either bigger or minor than a certain size in cm.

This condition was easily implemented by getting the value from the data property **values** on the **CalculationResult** entity. This entity is related to the **ImageAnnotation** entity through the **hasCalculationEntity** object property. In order to satisfy this condition, we assert the following rules taking 5 centimeters as the longest dimension of the target liver lesion (in SWRL notation):

LessThan5cmTumor:

$$ImageAnnotation(?x) \land hasCalculationEntity(?x, ?y) \land$$
$$hasCalculationResult(?y, ?z) \land$$
$$values(?z, ?val) \land lessThan(?val, 5) \rightarrow lessThan5cmTumor(?x)$$

MoreThan5cmTumor:

$$ImageAnnotation(?x) \land hasCalculationEntity(?x, ?y) \land$$
$$hasCalculationResult(?y, ?z) \land$$
$$values(?z, ?val) \land greaterThan(?val, 5) \rightarrow MoreThan5cmTumor(?x)$$

Condition 3:   Staging should consider lesions in adjacent organs.

To satisfy this condition, the most complicated criterion of classification, we had to consider the fact that a cancerous tumor can spread throughout the body. For that, we needed to create one new concept, based on the **Lesion** class from the Onlira ontology [32]. The **Lesion** class handles important characteristics of a lesion, such as composition, density, size, and shape. But, unfortunately, they are not enough for TNM classification and reasoning. For this reason, we added 3 properties to it and created the subclass **OutsideLesion**; these properties are:

–   **hasLocation** (object property): This property indicates the lesions location based on RadLex taxonomy. This property relates Onlira **Lesion** class instances to RadLex **AnatomicalEntity** class instances (see Fig. 7).
–   **isRegionalLymphNodeAffected** (data property): This property denotes whether a lesion is found in some lymph node. It was useful to enable classification criteria such as N0 and N1 (see Fig. 7).
–   **isAdjacentOrgan** (data property): This property denotes whether a lesion with a **hasLocation** value ``X`` is close to any adjacent organ. In accordance with the TNM liver classification criterion, which is the case of study in this work, we considered as adjacent organs to the liver [10]; the pancreas, duodenum, and colon (see Fig. 8). Furthermore, we grouped these concepts as organs in RadLex representation, creating two new classes, **AdjacentOrganGroup** and **NoAdjacentOrganGroup**:

$$NoAdjacentOrganGroup \sqsubseteq RadlexEntity.Organ$$
$$AdjacentOrganGroup \sqsubseteq RadlexEntity.Organ$$

AdjacentOrganGroup and **NoAdjacentOrganGroup** classes indicate whether a body organ is considered adjacent or not to the organ where the primary tumor was located. The primary organ defines the type of staging system to use; in our case, this organ was the liver. Finally, we constructed the following rule (in SWRL notation) to indicates whether an **OutsideLesion** is located in an adjacent organ:

$$Lesion(?x) \wedge adjacentOrganGroup(?y) \wedge$$
$$hasLocation(?x, ?y) \rightarrow isAdjacentOrgan(?x, true)$$

Once the above requirements were adequately covered using OWL and SWRL rules, we constructed the axioms and rules in order to be able to automatically classify cancer lesions, based on the TNM system. We noticed that the way we modeled things mattered. For example, it was easier to define N1a and N0 criteria and reuse their definitions for M0, rather than to start with the definition of M0 and end up handling complex closures. With the use of the AIM4-O ontology, anatomical concepts can easily be related to each other as demonstrated previously.

Experimental Study and Results

In this section, we first describe our experimental datasets, based on actual medical images and reports. Then, we evaluate the expressivity of the AIM4-O ontology. Finally, we present a quantitative evaluation of our TNM classifier for semantic image findings, which is the objective of this work, using precision and recall.

Datasets

Our first dataset is a set of real clinical reports of Hepatocellular Carcinoma (HCC) patients from The NCI's Genomic Data Commons (GDC). In this work, all
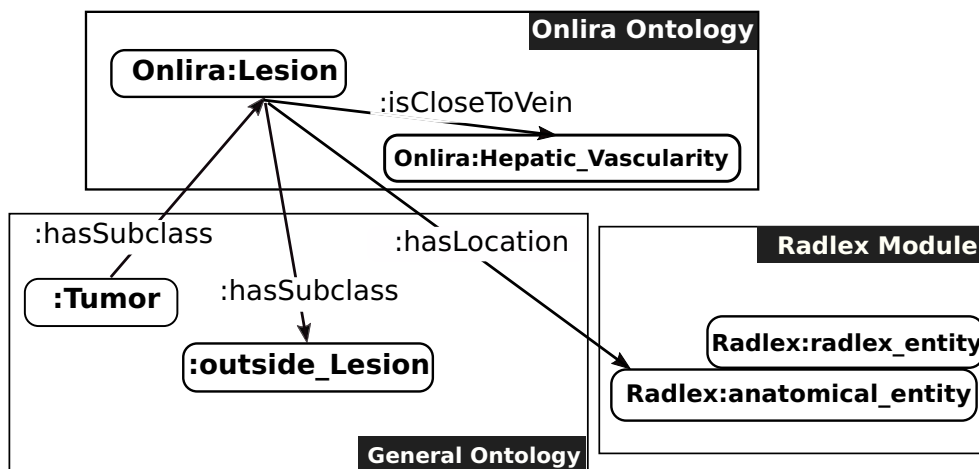

Fig. 7 General Ontology adds to the **Onlira:Lesion** class 2 properties defined as necessary for the TNM representation
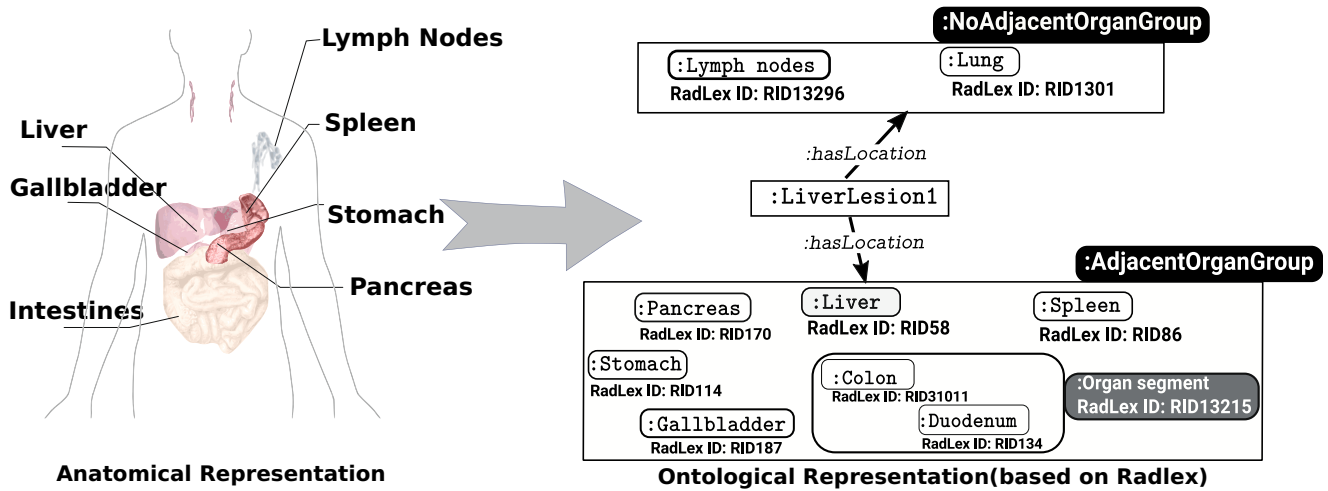
Fig. 8 Getting the subclass-hierarchy from Radlex. *AdjacentOrganGroup (Pancreas, Spleen, Stomach, Gallblader, and Colon)* and *NoAdjacentOrganGroup* *(Lymph nodes, Lung)* classes are created regarding the organ where the primary tumor was located (in this case, the liver)

Table 2 Description logic and keyword representation for four queries

Query ID	DL query	Keyword query
Q1	hasAnnotations some (hasImageAnnotations some (hasImagingObservation some (ImagingObservationEntity and label value "Lesion type")))	Tumor
Q2	hasAnnotations min 2	Multiple tumor
Q3	hasAnnotations some (hasImageAnnotations some (hasCalculationEntity some (hasCalculationResult some (some values float [¿ 8.0f]))))	Tumor size greater than 8 cm
Q4	hasImagingObservationCharacteristic 1 min	Vascular tumor invasion mass

Table 3 Table showing precision and recall using the two gold standards (Stanford and Marilia) for the four queries

id	Semantic search—DL				Keyword search			
	Precision		Recall		Precision		Recall	
	Stanford	Marilia	Stanford	Marilia	Stanford	Marilia	Stanford	Marilia
Q1	15/15	13/15	15/15	13/13	12/12	12/12	12/15	12/13
	1.0	0.87	1.0	1.0	1.0	1.0	0.8	0.92
Q2	5/5	5/5	5/5	5/5	2/5	2/5	2/5	2/5
	1.0	1.0	1.0	1.0	0.4	0.4	0.4	0.4
Q3	9/10	9/10	9/10	9/10	0/10	0/10	0/10	0/10
	0.9	0.9	0.9	0.9	0.0	0.0	0.0	0.0
Q4	10/15	9/15	10/10	10/10	7/7	7/7	7/10	7/10
	0.67	0.6	1.0	1.0	1.0	1.0	0.7	0.7

experiments were supported by the GDC data. An important requirement to enable a feasible clinical evaluation was to have an image dataset to validate the results of the GDC clinical reports. To cover this requirement, we used the TCIA database [33]. It hosts a large archive of medical images about cancer, it is accessible for public download and it is related to the GDC records by a patient **subject ID**. The imaging modality selected was computed tomography (CT). The downloaded images were loaded into the ePAD annotation tool and annotated.

While TNM staging could be applied to other types of cancer, this work focuses on staging liver cancer. One reason was the availability of clinical data and images for this kind of cancer. For a given patient, the input to our TNM classifier consists of AIM files (image annotations) and the output consists of the Cancer Staging for this patient.

Quantitative Assessment of AIM4-O Ontology

According to Blomqvist et al. [34], "the ontological evaluation is the process of assessing an ontology with respect to certain criteria, using certain measures." In this work, we undertook the evaluation of the AIM4-O ontology from the functional point of view. To achieve this, we carried out a task-focused assessment and inference requirements [32]. In order to evaluate the AIM4-O ontology, we studied and evaluated how it could help in searching clinical reports that describe image findings (reports about cancer). For this purpose, we compared two different approaches:

– Ontology-based (semantic) search: If the clinical reports are described as AIM4-O individuals, these reports can be searched using description logic query languages (DL query).
– Natural Language process-based (keyword) search: Clinical reports and image findings are usually written in natural language. There are many ways to implement keyword search. We decided to use a very popular full-text search engine that can be used from various programming languages: the Apache Lucene.[6]

In the literature, ontology-based search performs better than keyword-based search [32]

One of the reasons is that ontology search can search for information not explicitly mentioned in the text. For instance, it is possible to search for reports not having some features: Find all records of tumors not in the liver. Using keyword search, the best one can do is to find reports without the word liver. Many reports with the word liver may talk about tumors in other organs.

If an ontology-based search system, using the AIM4-O ontology, outperforms a keyword search system

implemented using Apache Lucene it is a good positive quantitative assessment for the ontology.

In order to highlight the differences between the two approaches, we used four queries expressed both in DL (DL query) and keywords (see Table 2):

1. Q1—Find all reports related to an image observation (tumor observation).
2. Q2—Find all reports that describe multiple tumors.
3. Q3—Find all reports that contain a tumor observation that has a size greater than 8 cm.
4. Q4—Find all reports that contain a tumor observation with descriptors (e.g., invasion, mass, vascular).

The DL queries were processed using the ontology editor Protégé[7] (using its default reasoner HermiT). For the keywords, a small Java program was created to read the report texts, read the keywords and use the class StandardAnalyser[8], from the Apache Lucene library, to make the search.

We have considered the following points in order to evaluate both approaches:

– The evaluation was based on GDC reports: We randomly took 15 radiology reports of different patients written in natural language and converted them into AIM4-O instances.
– A report was retrieved if it satisfied the DL query or it collects all keywords in the search query.
– Finally, we compared the precision and recall against a gold standard. Precision is the proportion of truly retrieved reports to the total number of reports retrieved. Recall is the proportion of truly retrieved reports to the total number of reports that should have been retrieved [32].

The gold standard was determined manually by 3 radiology professors from two different institutions: one from Stanford University School of Medicine and two from the Faculty of Medicine of Marilia (Brazil). They manually evaluated each query to decide which of the 15 reports should be retrieved.

The four queries with corresponding precision and recall results are shown in Table 3.

By analyzing the four queries, we can see that the semantic search has the greatest number of relevant documents retrieved:

Q1: With the semantic approach, 15 reports were retrieved with an average precision of 0.95 and recall of
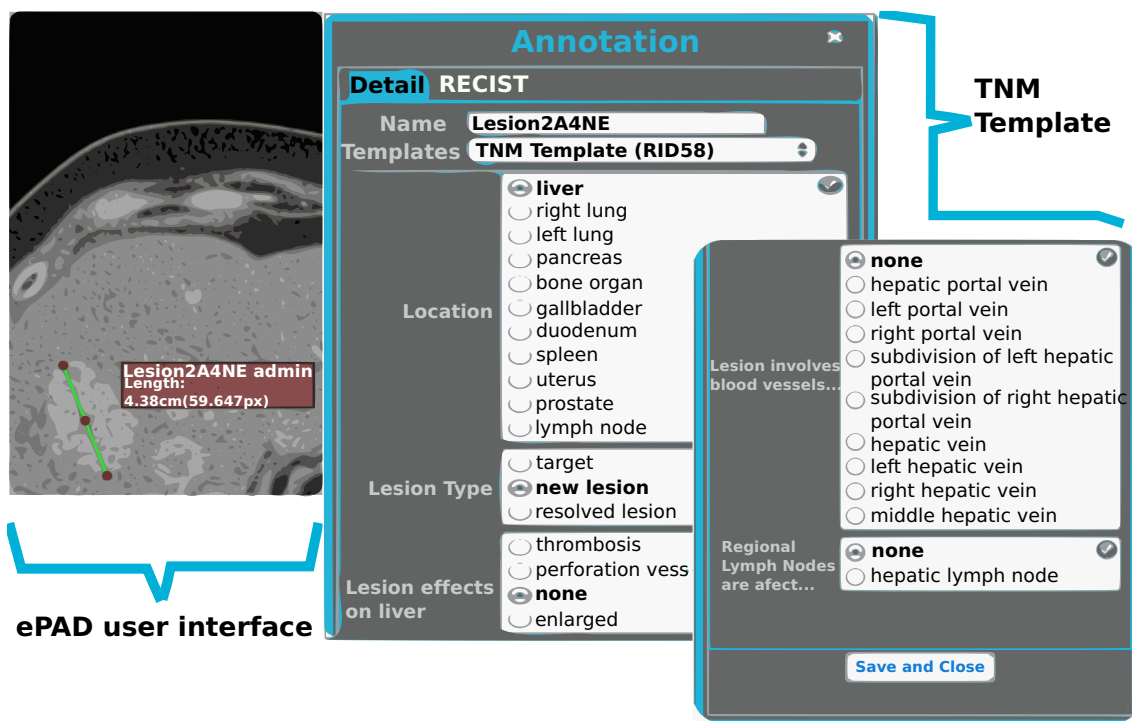

Fig. 9 A CT image of the liver annotated using the TNM template (on the right of the image)

1.0. The keyword approach returned 12 reports with 1.0 precision and 0.96 average recall.

Q2: With the semantic approach, 5 reports were retrieved with a 1.0 precision and recall. Much better than the 2 reports retrieved by the keyword with just 0.4 precision and recall.

Q3: In this question, the keyword approach was very poor with no reports retrieved. There were no reports containing the queried words (i.e., "lesion size greater than 8 cm"). Keyword search is not well suited for queries with numerical relations, but such relations are very important when searching for tumors. The semantic approach returned 10 reports with 0.9 precision and recall.

Q4: The semantic approach retrieved more reports (15 vs 7) with 0.67 precision and 1.0. When compared with the keyword approach (precision 1.0 and recall 0.7), it had a similar performance with 0.80 F1 versus 0.82 of the keyword search.

The semantic search approach performance was better, with recall values close to 1 and always better than the keyword search in both golden ratio values. Also, in all but one case, precision values were better for the semantic search. That shows that the AIM4-O ontology is able to semantically represent the information in the reports well enough to outperform keyword search. Its representation can also be used by reasoners to successfully compute information note

explicitly stated in the report (such as the fact of a tumor be bigger than a given size, question Q3).

Automatic TNM Clinical Stage

In this section, we calculated the classification rate of the TNM classifier. At first, we created an ePAD template named "TNM template" in order to provide radiologists with a prespecified set of semantic terms for image annotations. These image annotations, which are compatible with the ePAD tool, were stored in the AIM-XML format. An example of an annotated image is presented in Fig. 9.

After, the generated image annotations (in AIM-XML format) were classified automatically using the TNM criteria. This process was duly evaluated and correctly

Table 4 Confusion matrix of cancer stages predicted by the TNM classifier versus the values the physicians placed in the reports

n = 51		Actual stages				
		I	II	IIIA	IIIB	IVA
Predicted stages	I	24	2	0	0	0
	II	0	10	2	0	0
	IIIA	0	1	8	0	0
	IIIB	0	0	0	2	1
	IVA	0	0	0	0	1
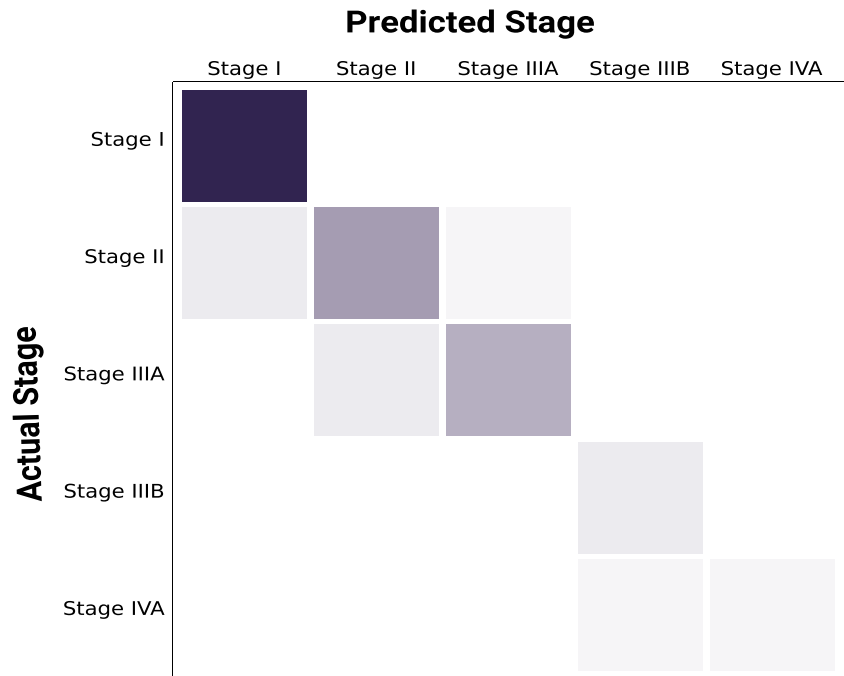
accepted by two radiology professors from two different institutions (Stanford University School of Medicine and Faculty of Medicine of Marilia). They also analyzed the accuracy of the generated annotations in terms of semantics. The process we followed was:

– The data set used came from the following open databases:

- The NCI's Genomic Data Commons (GDC)[9].
- The Cancer Imaging Archive (TCIA)[10] (collects only images, the number of series and studies): As we were working with TNM classification from the liver, we searched for "LIHC - Liver hepatocellular carcinoma" obtaining 52 patients with information available in both databases (images and reports). However, the information about tumor size was obtained by manual review of the medical reports. These reports are also available in The NCI's Genomic Data Commons.

– After reading the medical reports, the radiologist was provided with an excel spreadsheet that provided information about medical findings, such as lesion size, vascular invasion, and others.

– Based on this excel file and the GDC data, we created AIM-XML annotations and integrated them into our knowledge base (as AIM4-O ontology individuals).

---

[9] https://gdc.cancer.gov/

[10] https://public.cancerimagingarchive.net/ncia/login.jsf

– The AIM files were used as inputs for our TNM classifier. The produced output was compared with the TNM values that physicians reported.

The AIM image annotations were generated based on 52 different clinical reports. Our automatic staging approach was evaluated by using precision and recall values. The cancer stages generated by our TNM classifier were compared to those described by the physicians who created the original clinical reports (our gold standard). We used the 7th edition of TNM [10]. One patient, with the subject ID "TCGA-DD-A1EJ," was removed from this analysis. Our radiology professors considered that the TNM classification, reported by the physician in his respective clinical report, was incorrect (more information below).

For the calculation of precision and recall, the result is considered positive when the automatic staging coincides with the stage given by physicians who created the original clinical reports (see Table 4).

Precision was 85.7% and recall 81.0% (for 51 patients). This means that, for precision, at 85.7% of the time the system agreed with the staging given by physicians. For the recall, this means that, of all the times that a given stage was reported by a physician, in 81.0% of cases the system agreed with him/her. It is important to note that, even when the system diverged from physicians, the maximum difference between them was only one stage.

In Fig. 10 we show the results of the evaluation summarized in the color scale matrix. It represents our confusion matrix for a multi-stage classification. The darker the square in the diagonal of the matrix means that the
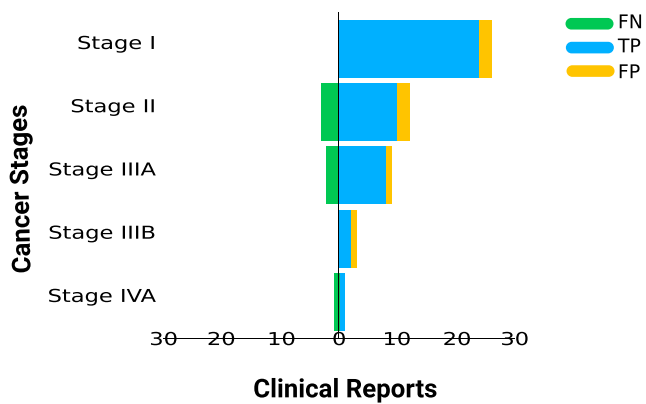
Fig. 11 Summary of histograms for each TNM stage from the confusion matrix (51 reports): FN—false negatives, FP—false positives, and TP—true positives

respective class was better classified. The other squares in grays, outside of the diagonal, indicates that the class in the vertical axis was confused by the classifier with the corresponding class on the vertical axis.

For early stages of cancer, such as I, II, and IIIA, the percentage of misclassifications (e.g., false positives and false negatives) was very small. They are represented by the highlighted diagonal of the matrix (Fig. 10). For more advanced stages of cancer, such as IIIB or IVA, it was larger (Fig. 11). This may have happened simply because we had few patients at these stages or because these stages are described by relatively more complex concepts.

We also performed a sanity check of what was recorded in the AIM and the output produced by the classifier. The 3 radiologists that participated in this validation manually reviewed the whole process (including the stage assigned in the patient report), using the patient images and reports, for 3 randomly chosen patients. In all cases, the whole process for generating the AIM representation and staging classification was correct.

Our classifier also revealed the fact that there are clinical reports with inaccurate staging diagnosis. An example of this situation was the clinical case with subject ID "TCGA-DD-A1EJ." This case was the only in which the difference between the classifier's stage and the physician's evaluation differed by more than one level, our radiologists decided to analyze how thus case was processed. They concluded that the case has been processed correctly and that the result of the classifier was also correct. The stage predicted was Stage I; however, the stage described by the medical report was Stage III.[11] They recommended us to not use this patient's data, so this report was excluded from our analysis. Examples, like this, serve to warrant the importance of

---

[11] https://portal.gdc.cancer.gov/cases/52292ffc-0902-4d97-b461-20723987a177

improving clinical decision support systems (through the use of image metadata in cancer treatment).

Conclusion

Cancer staging entails an intensive work, this often requires an accurate interpretation of the cancer findings in images by medical experts (oncologists and radiologists). Expert accuracy is achieved through training and experience [35], but variations in image interpretation is a human observer limitation. In this context, we developed an automatic staging approach (a TNM classifier). It can help physicians to obtain a higher accuracy rate for image interpretation.

To achieve this, first an ontology to represent AIM4 annotations, called AIM4-O, was developed and validated using a task-focused assessment of actual clinical cases. Using the ontology to semantically search reports, we got much higher precision and recall values when compared to keyword (no semantics) search. Subsequently, the General ontology, integrating the AIM4 ontology, Onlira ontology (a subset of the RadLex vocabulary) and SWRL rules for TNM staging, was developed and used to develop a TNM classifier using the Groovy language.

This TNM classifier was evaluated using actual cancer cases. Our experimental data showed that, when compared to 51 staging values given in actual physician reports, the classifier generated results had 85.1% precision and 81.0% recall. When the classifier stages differed from physician's reported stages, that difference was, at most, of 1 stage.

The TNM classifier also revealed one patient report with inconsistencies in the diagnostics. It is important to note that this automatic staging procedure does not give clinicians new information. It is merely a second opinion for the purposes of quality in clinical diagnosis. We also highlighted some limitations of description logics, such as the open-world assumption.

Our TNM classifier can be used in automated clinical workflows, where AIM based image annotations are produced by imaging systems. Automatic TNM staging can be as easy as pushing a button in such systems.

We believe that our approach could be also applied to other kinds of cancer such as lung or colon, by modifying only the rules and axioms that represent the TNM criteria. That can be done avoiding the creation of an entirely new ontology for each type of cancer.

Besides cancer staging, other tasks, such as RECIST cancer criteria can also be automated using this combination of AIM, OWL ontologies, and SWRL rules.

Future work will include more varied data sets for evaluation, expansion of the classifier to other organs, and incorporation into existing information systems (such as ePAD). The TNM classifier has the potential to be integrated

into larger software systems. A Specific Domain Language (DSL) to describe TNM criteria can be developed as a communication tool between physicians and TNM criteria formal representation (axioms and SWRL rules). It would allow physicians to modify the classifier rules themselves.

A limitation to this work is that a relatively small dataset was used in our evaluation. One reason is the requirement that both medical images (CT) and clinical reports have to be present for the same patient for optimal validation. Another is the time constraint for radiologists and the difficulties to get them to review large datasets.

References

1. Levy M, O'Connor MJ, Rubin DL: Semantic reasoning with image annotations for tumor assessment. AMIA Ann Symp Proc 2009:359–63, 2009

2. Wennerberg P, Schulz K, Buitelaar P: Ontology modularization to improve semantic medical image annotation. J Biomed Inform 44:155–162, 2011. https://doi.org/10.1016/j.jbi.2010.12.005

3. Bretschneider C, Zillner S, Hammon M (2013) Grammar-based lexicon enhancement for aligning German radiology text and images. In: Proceedings of the Recent Advances in Natural Language Processing (RANLP 2013), Hissar, Bulgaria, pp 105–112

4. Zillner S: Reasoning-based Patient Classification for Enhanced Medical Image Annotation. In: Extended Semantic Web Conference. Springer, Berlin, 2010, pp 243–257

5. Rubin DL, Willrett D, O'Connor MJ, Hage C, Kurtz C, Moreira DA: Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials. Transl Oncol 7:23–35, 2014. https://doi.org/10.1593/tlo.13796

6. Channin DS, Mongkolwat P, Kleper V, Rubin DL: The Annotation and Image Mark-up Project. Radiology 253:590–592, 2009. https://doi.org/10.1148/radiol.2533090135. pMID: 19952021

7. Channin DS, Mongkolwat P, Kleper V, Sepukar K, Rubin DL: The caBIG^TM Annotation and Image Markup Project. J Digit Imaging 23:217–225, 2010. https://doi.org/10.1007/s10278-009-9193-9

8. Rubin DL, Rodriguez C, Shah P, Beaulieu C (2008) Ipad: semantic annotation and markup of radiological images. In: AMIA Annual Symposium Proceedings, pp 626–630

9. Kundu S, Itkin M, Gervais D. a., Krishnamurthy VN, Wallace MJ, Cardella JF, Rubin DL, Langlotz CP: The IR Radlex Project: An Interventional Radiology Lexicon-A Collaborative Project of the Radiological Society of North America and the Society of Interventional Radiology. J Vasc Interv Radiol 20:433–435, 2009. https://doi.org/10.1016/j.jvir.2008.10.022

10. Faria SC, Szklaruk J, Kaseb AO, Hassabo HM, Elsayes KM: TNM/Okuda/Barcelona/UNOS/CLIP International Multidisciplinary Classification of Hepatocellular Carcinoma: concepts, perspectives, and radiologic implications. Abdom Imaging 39:1070–1087, 2014. https://doi.org/10.1007/s00261-014-0130-0

11. Dameron O, Roques É, Rubin D, Marquet G, Burgun A (2006) Grading lung tumors using OWL-DL based reasoning,

12. Marquet G, Dameron O, Saikali S, Mosser J, Burgun A (2007) Grading glioma tumors using OWL-DL and NCI thesaurus. In: AMIA Annual Symposium Proceedings, volume 2007, American Medical Informatics Association, pp 508–512. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655830

13. Kleihues P, Sobin LH: World Health Organization classification of tumors. Cancer 88:2887–2887, 2000

14. Keet CM: Open World Assumption. In: Encyclopedia of Systems Biology. Springer, Berlin, 2013, pp 1567–1567

15. Tutac AE, Racoceanu D, Putti T, Xiong W, Leow WK, Cretu V: Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images. 2008 International Conference on BioMedical Engineering and Informatics 2:107–112, 2008. https://doi.org/10.1109/BMEI.2008.166

16. Tutac AE, Cretu VI, Racoceanu D (2010) Spatial representation and reasoning in breast cancer grading ontology. In: 2010 International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC-CONTI), pp 89–94. https://doi.org/10.1109/ICCCYB.2010.5491320

17. Massicano F, Sasso A, Tomaz H, Oleynik M, Nobrega C, Patrão DFC (2015) An Ontology for TNM Clinical Stage Inference. In: ONTOBRAS

18. Franca F, Schulz S, Bronsert P, Novais P, Boeker M (2015) Feasibility of an ontology driven tumor-node-metastasis classifier application: A study on colorectal cancer. In: 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), pp 1–7. https://doi.org/10.1109/INISTA.2015.7276757

19. Zillner S (2009) Towards the Ontology-based Classification of Lymphoma Patients using Semantic Image Annotations. In: SWAT4LS, Citeseer

20. Boeker M, França F, Bronsert P, Schulz S: TNM-O: ontology support for staging of malignant tumours. Journal of Biomedical Semantics 7:64, 2016. https://doi.org/10.1186/s13326-016-0106-9

21. Boeker M, Faria R, Schulz S (2014) A Proposal for an Ontology for the Tumor-Node-Metastasis Classification of Malignant Tumors: a Study on Breast Tumors, et al., Ontologies and Data in Life Sciences (ODLS 2014), IMISEREPORTS Leipzig. http://www.onto-med.de/obml/ws2014/odls2014report.pdf

22. Seneviratne O, Rashid SM, Chari S, McCusker JP, Bennett KP, Hendler JA, McGuinness DL (2018) Knowledge Integration for Disease Characterization: A Breast Cancer Example. arXiv:1807.07991

23. Meriem B, Yamina T, Pathology A (2012) Interpretation breast cancer imaging by using ontology, Cyber Journals: Multidisciplinary Journals in Science and Technology. Journal of Selected Areas in Bioengineering (JSAB). pp 1–6. http://www.cyberjournals.com/Papers/Mar2012/06.pdf

24. Racoceanu D, Capron F: Towards semantic-driven high-content image analysis: An operational instantiation for mitosis detection in digital histopathology. Comput Med Imaging Graph 42:2–15, 2015. https://doi.org/10.1016/j.compmedimag.2014.09.004

25. Möller M., Sonntag D, Erñst P: A Spatio-anatomical Medical Ontology and Automatic Plausibility Checks Berlin: Springer, 2013, pp 41–55. https://doi.org/10.1007/978-3-642-29764-9-3

26. Zillner S, Sonntag D: Image metadata reasoning for improved clinical decision support. Network Modeling Analysis in Health Informatics and Bioinformatics 1:37–46, 2012. https://doi.org/10.1007/s13721-012-0003-9

27. Gimenez F, Xu J, Liu Y, Liu TT, Beaulieu CF, Rubin DL, Napel S (2011) On the Feasibility of Predicting Radiological Observations from Computational Imaging Features of Liver Lesions in CT Scans, in: 2011 First IEEE International Conference on Healthcare

Informatics, Imaging and Systems Biology (HISB), pp 346–350. https://doi.org/10.1109/HISB.2011.37

28. Kurtz C, Depeursinge A, Napel S, Beaulieu CF, Rubin DL: On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. Med Image Anal 18:1082–1100, 2014. https://doi.org/10.1016/j.media.2014.06.009

29. Bulu H, Rubin DL (2015) Java Application Programming Interface (API) for Annotation Imaging Markup (AIM). 1–12. https://sourceforge.net/projects/aimapi/, [Online; accessed 02-March-2015.]

30. Rubin DL: Finding the meaning in images: annotation and image markup. Philos Psychiatry Psychol 18:311–318, 2011. https://doi.org/10.1353/ppp.2011.0045

31. Radlex (2016) Radlex, http://www.radlex.org. [Online; accessed 02-October-2016.]

32. Kokciyan N, Turkay R, Uskudarli S, Yolum P, Bakir B, Acar B: Semantic description of liver ct images: An ontological approach. IEEE J Biomed Health Inform 18:1363–1369, 2014. https://doi.org/10.1109/JBHI.2014.2298880

33. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, et al: The Cancer Imaging Archive (tcia): Maintaining and Operating a Public Information Repository. J Digit Imaging 26:1045–1057, 2013. https://doi.org/10.1007/s10278-013-9622-7

34. Blomqvist E, Seil Sepour A, Presutti V: Ontology Testing - Methodology and Tool. In: Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW12. Springer, Berlin, 2012, pp 216–226

35. Depeursinge A, Kurtz C, Beaulieu C, Napel S, Rubin D: Predicting Visual Semantic Descriptive Terms from Radiological Image Data: Preliminary Results With Liver Lesions in CT. IEEE Trans Med Imaging 33:1669–1676, 2014. https://doi.org/10.1109/TMI.2014.2321347