An ontology for mammography screening recommendation

Cindy Acuña¹, Yasmine Anchén²*, Edelweis Rohrer¹* and Regina Motz¹*

¹Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay
²Facultad de Medicina, Universidad de la República, Av. Gral. Flores 2125, 11800 Montevideo, Uruguay.

Abstract
This article presents the development of a simple Mammography Screening Ontology (MAMO-SCR-Onto) designed to provide mammography application recommendations for the early detection of breast cancer in patients at average risk of the disease. We highlight the simplicity of this ontology that allows, on the one hand, to be a formal communication tool between doctor-patient; on the other, to be an open educational resource for the training of preservice health professionals. The ontology is developed in OWL and accessible with an open licence.

Keywords
mammography screening recommendation, breast cancer early detection, patients with average risk

1. Introduction
Breast cancer is a significant global health concern, underscoring the vital need for effective early detection methods. The American Cancer Society (ACS), a leading cancer-fighting organization, has established comprehensive guidelines for mammography as a crucial test for the early detection of breast cancer in women with an average risk of contracting breast cancer [1]. In spite of other recognized institutions’ recommendations, in the present work we consider ACS’s recommendations as a case study without loss of generality. These recommendations are about the frequency application of mammography as a test for early breast cancer detection based on women’s values in risk factors. On the other hand, ontologies are widely adopted and proven modelling artefacts for conceptualizing various domains, including education and health. They provide a structured framework to describe vocabularies in terms of concepts (or entity types), instances (or individuals), and roles (or relations), as well as assertions or constraints about them [2]. Notably, the W3C standard ontology language OWL, with its formal semantics based on description logics, facilitates the implementation of a comprehensive set of assertions for a given vocabulary, particularly in the health domain [3]. OWL enables the automatic validation of ontology constraints and facilitates the inference of assertions that
may not have been explicitly declared. This capability empowers the ontology to go beyond the explicitly stated information, uncovering implicit knowledge and facilitating more robust reasoning. Reasoners, i.e. algorithmic implementations such as the tableau algorithm, play a pivotal role in this process. By examining an ontology, reasoners ensure consistency by verifying that the specified constraints are met while drawing logical inferences based on the underlying description logic semantics [4].

The present work makes a significant contribution by modelling and implementing a simple owl ontology designed to provide mammography screening application recommendations for the early detection of breast cancer in women at average risk of the disease. The ontology represents the conditions taken into account to consider a woman within the group "women with an average risk of contracting breast cancer" (group of women to whom the guide is directed), and it also represents specific recommendations that these women receive according to their ages, such as the periodicity of the study. These conditions are modelled as constraints, which the reasoner utilizes to infer the appropriate advice by the guidelines established by the American Cancer Society. This ontology captures the intricate relationships between various conditions that influence the decision to perform screening mammography, such as the risk group to which the woman belongs (taking into account confirmed or suspected genetic mutation, personal history, and exposure to chest radiation treatments) and the woman’s age. By formalizing expert knowledge and integrating it into an ontology, we facilitate the development of intelligent decision support systems that can aid healthcare professionals in making informed and personalized screening recommendations. Moreover, this formal specification establishes a standardized framework and is a valuable tool for visualization and reasoning on the conditions. By leveraging the power of ontologies and their reasoning capabilities, we can optimize breast cancer screening protocols and ultimately contribute to improving healthcare outcomes for individuals at average risk.

This paper presents our approach to designing the mammography screening ontology, including selecting relevant concepts, relationships, and axioms. We outline the process of integrating domain-specific knowledge from authoritative sources, such as the ACS recommendations and the NCI thesaurus [5] into the ontology. Furthermore, we discuss the benefits and challenges of using OWL and reasoning techniques to ensure the ontology’s consistency, coherence, and inferential capabilities.

The remainder of this paper is organized as follows: Section 2 provides an overview of related work in the field of ontology-based clinical decision support systems. Section 3 presents the ontology’s key components and knowledge representation. Section 4 offers the evaluation of the ontology. Finally, Section 5 concludes the paper with a summary of the contributions, limitations, and future directions.

2. Related work

Over the past two decades, ontologies have emerged as crucial artefacts in supporting the study and research of cancer. The adoption of ontologies as a powerful approach has been evident through the development of numerous ontologies in the biomedical domain [6].

One prominent example of a comprehensive ontology in the field of cancer is the National
Cancer Institute Thesaurus and Ontology (NCIt). Designed as an extensive compilation of terminologies, the NCIt covers various facets of cancer research and care [5]. As a fundamental resource, the NCIt plays a pivotal role in facilitating standardized communication and promoting knowledge sharing among researchers, clinicians, and other healthcare professionals in the field of oncology. However, the significance of the NCIt and other ontologies in the cancer domain extends beyond providing a common vocabulary. These ontologies establish a structured framework that enables collaboration, interoperability, and the integration of diverse data sources. They serve as a foundation for knowledge representation, helping researchers and practitioners to capture and organize complex cancer-related information effectively.

In recent years, there has been increasing interest in using ontologies and knowledge graphs in cancer research, as highlighted by the comprehensive review by Silva et al. [7]. This study examined 141 open articles published between 2012 and 2021. Of the papers analyzed, 18 focus on breast cancer-related ontologies. We can observe that breast cancer ontologies serve several purposes, each addressing different aspects of knowledge representation and integration within the field or supporting decision-based systems. Some of these ontologies prioritize the computational accessibility of breast cancer knowledge to represent the information in a format that computer systems can efficiently process [8, 9, 10, 11]. On the other hand, specific ontologies focus on integrating cancer data from various sources, consolidating the information in a unified database [12, 13]. In some cases, the ontologies are designed to facilitate effective communication between doctors and patients, harmonizing terminology and improving mutual understanding [14] or designed to facilitate database user interfaces, where ontology class and relationship labels allow text annotation within the breast cancer domain [15]. In addition, some ontologies in the breast cancer domain claim to represent causal associations between breast cancer incidence and various risk factors, providing valuable insight into the complex relationships involved [16, 8, 17, 18]. Sherimon et al. [19] describe an architecture of six ontologies (patient ontology, breast cancer, symptoms, risk, lab test and questionnaire) to predict breast cancer. However, we cannot find any of these ontologies accessible. Another application involves the semantic modelling of breast cancer-related drugs using ontological approaches, allowing the inference of possible drug repositioning strategies [20]. However, the most notable applications are those that produce clinical inferences [21, 22, 23].

Regarding the radiology domain, we find Sun et al. work presenting an ontology for training breast radiologists [24]. Focusing on the mammography domain, the GIMI Project developed a fundamental mammographic ontology and an ontology for computerized training in breast radiology [25, 26, 27, 28, 29]. Other works on mammographic ontology are Bulubu et al. [30] describing ontology-based mammography annotation and a case-based retrieval approach for breast masses from the digital mammography archive, Ilyass et al. [31] which annotates images in mammograms according to the concepts of a breast cancer ontology which produces RDF metadata and the more recent ones [32, 33, 34, 35].

However, to the best of our knowledge, no open ontologies provide mammography application recommendations for the early detection of breast cancer in women at average risk of the disease as the simple one that we present in this work.
3. Modelling the mammography screening recommendation ontology

ACS recommendations were based on the quality of the evidence obtained and judgment on the balance of benefits and harms of conducting the study, classifying them into strong recommendations (benefits of conducting the study outweigh the desirable effects) or qualified recommendations (there is clear evidence of benefit but less certainty about the balance of harm/benefit, or on the values and preferences of the patient, allows discussion between doctor and patient to make the decision). Based on the age of women at average risk of breast cancer, the ACS recommends Women at average risk of breast cancer should have screening mammography starting at age 40, following the age range recommendations below:

- Women should begin annual screening between the ages of 40 and 44 as a qualified recommendation.
- Women ages 45 to 54 should have screening mammograms annually as a strong recommendation.
- Women age 55 and older should have screening mammograms biennially or have the opportunity to continue screening annually as a qualified recommendation.

Given the relevance of the ACS recommendations, both for the doctor and the patient, the contribution of the ontology described in this section is twofold: (i) conceptualizing the concrete domain by describing those properties or attributes of women which are relevant inputs for applying the recommendation as well as the rules behind it, and (ii) computing a customized recommendation for a set of women instances, which represents a useful tool that aims most doctors and patients quick access to the guide recommendation and to meet the ultimate goal of the earliest prevention of the breast cancer.

Figure 1 depicts the mammography screening recommendation ontology, which has three main classes: Woman, History and Recommendation.

Even though for the scenario described in the present paper, it is important to represent women with an intermediate risk of contracting breast cancer, this group of women is represented by the subclass With intermediate risk, of a more general class Woman, to facilitate a possible extension of the ontology if maybe in the future recommendations for other groups of women were conceptualized (for example, with high or low risk of breast cancer). Classes Age Range 40 to 44, Age Range 45 to 54, and Age gt or eq 55 represent subsets of women with an intermediate risk that have corresponding ages.

As for the class Woman, the class History is modelled as a general class even though in the present work, we focus on the subclass Personal Medical History that represents the set of medical records of women. Properties hasHistory and doesNotHaveHistory connect each woman to her medical records. The property hasHistory connects women to diseases, clinical diagnoses or laboratory test results that they have. The property doesNotHaveHistory connects women to those diseases, clinical diagnoses or important laboratory test results so that the doctor knows that the patient does not have it. It is highly relevant that this be made explicit when evaluating the type of breast cancer risk corresponding to the woman and to be able to make a correct recommendation. The ACS guide clarifies that the recommendations are aimed at women with an average risk of breast cancer; these are those without a personal history of breast cancer,
without suspicion or confirmation of a known genetic mutation that increases the risk of breast cancer and without previous history of chest radiotherapy at a young age. These conditions are represented in our ontology by individuals confirmed or suspected genetic mutation, personal history of breast cancer, radiotherapy to chest, as instances of the class Personal Medical History.

The class Recommendation represents the set of recommendations provided by the guide. Each particular recommendation has three characteristics or dimensions that describe it: (i) the strength of the recommendation given by the level of evidence to issue the recommendation, represented by the class Recommendation Type, which has instances Strong and Qualified (ii) the recommended mammography study, represented by the class Mammography and the instance Screening Mammography\(^1\), and (iii) the frequency with which mammography should be performed, represented by the class Periodicity, with two instances: Annual and Annual or Biaannual. Properties hasRecType, hasMammography and hasPeriodicity connect each recommendation instance to the three dimensions.

Restrictions on classes and properties described above are represented in Table 1 as six groups of Tbox and Rbox axioms. Group (1) represents that classes Woman, History, Recommendation, Recommendation Type, Mammography and Periodicity cannot share instances with each other. The axiom (2) defines the class With intermediate risk as the set of women that meet the three conditions given by the guide to classify them in the category of women that have an average risk of breast cancer. Group (3) defines subclasses Age Range 40 to 44, Age Range 45 to 54 and Age gt or eq 55 as subsets of women with an average risk of breast cancer, based on the age range, represented in the implemented OWL ontology by the data property age. This is why

\(^1\)Note that for the scenario described in this work, only screening mammographies are considered.
Table 1  
Mammography screening recommendation domain Tbox and Rbox.

<table>
<thead>
<tr>
<th>Axiom</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woman ⊓ History ⊑⊥ Woman ⊓ Recommendation ⊑⊥ Woman ⊓ Mammography ⊑⊥ Woman ⊓ RecommendationType ⊑⊥ Woman ⊓ Periodicity ⊑⊥ History ⊓ Recommendation ⊑⊥ (1) History ⊓ Mammography ⊑⊥ History ⊓ RecommendationType ⊑⊥ History ⊓ Periodicity ⊑⊥ Recommendation ⊓ Mammography ⊑⊥ Recommendation ⊓ RecommendationType ⊑⊥ Mammography ⊓ RecommendationType ⊑⊥ Mammography ⊓ Periodicity ⊑⊥</td>
<td>Pairwise class disjointness.</td>
</tr>
<tr>
<td>(2) ∃doesNotHaveHistory.[Confirmed or suspected genetic mutation] ⊓ ∃doesNotHaveHistory.[Personal history of breast cancer] ⊓ ∃doesNotHaveHistory.[Radiotherapy to the chest]</td>
<td>Women with intermediate-risk are those do not meet a set of conditions.</td>
</tr>
<tr>
<td>Age ≥ 55 ⊑ Woman ⊓ ∃hasHistory.[With intermediate risk] ⊑⊥ Age ≥ 55 ⊓ ∃hasRecommendation.[55 Age range recommendation]</td>
<td>Each age subclass has corresponding age restriction and recommendation</td>
</tr>
<tr>
<td>Age ≥ 40 ⊑ Woman ⊓ ∃hasHistory.[With intermediate risk] ⊑⊥ Age ≥ 40 ⊓ ∃hasRecommendation.[40 to 44 Age range recommendation]</td>
<td></td>
</tr>
<tr>
<td>Age ≥ 45 ⊑ Woman ⊓ ∃hasHistory.[With intermediate risk] ⊑⊥ Age ≥ 45 ⊓ ∃hasRecommendation.[45 to 54 Age range recommendation]</td>
<td></td>
</tr>
<tr>
<td>∃hasHistory ⊑ ⊤ ⊓ Woman ⊑ ⊤ ⊓ ∃hasHistory ⊑ ⊤ ⊓ Woman ⊓ ∃doesNotHaveHistory.[Age gt or eq 55] ⊓ ∃hasRecommendation.[55 Age range recommendation]</td>
<td>Properties domain and range</td>
</tr>
<tr>
<td>∃hasHistory ⊑ ⊤ ⊓ Woman ⊓ ∃doesNotHaveHistory.[Age gt or eq 55] ⊓ ∃hasRecommendation.[55 Age range recommendation]</td>
<td></td>
</tr>
<tr>
<td>(4) hasMammography ⊓ ∃Recommendation ⊑⊥ hasMammography ⊓ ∃hasRecommendation.[Mammography] ⊓ ∃hasPeriodicity.[Periodicity] ⊓ ∃isRecommended.[Mammography ⊓ Periodicity ⊓ RecommendationType]</td>
<td></td>
</tr>
<tr>
<td>(5) Dis(hasHistory, doesNotHaveHistory)</td>
<td>A woman cannot have a clinical situation and does not have it at the same time</td>
</tr>
<tr>
<td>hasRecommendation o hasRecType ⊑⊥ isRecommended</td>
<td></td>
</tr>
<tr>
<td>(6) hasRecType ⊑⊥ hasMammography ⊓ isRecommended</td>
<td></td>
</tr>
<tr>
<td>hasRecType ⊑⊥ hasPeriodicity ⊓ isRecommended</td>
<td>A woman is recommended a strong or qualified recommendation, a kind of mammography and a periodicity.</td>
</tr>
</tbody>
</table>

the guide sets a different recommendation for each group of women, which is represented in the ontology as restrictions on instances of classes Age Range 40 to 44, Age Range 45 to 54 and Age gt or eq 55 that must have the corresponding recommendation. Axioms (2) and (3) are used by the reasoner to classify instances of the class Woman into categories given by descriptions of classes With intermediate risk, Age Range 40 to 44, Age Range 45 to 54 and Age gt or eq 55, entailing the corresponding recommendation. Group (4) describes restrictions on domains and ranges of properties. (5) and (6) describe restrictions on roles (Rbox axioms). The axiom (5) represents that a woman cannot be connected to the same medical record by both properties hasHistory and doesNotHaveHistory. In (6), the property isRecommended is defined as the superproperty of role chains hasRecommendation o hasRecType,
hasRecommendation o hasMammography and hasRecommendation o Periodicity. Besides inferring if the recommendation is for 40 to 44, 45 to 54 or ≥ 55 age range, for each woman, the reasoner directly entails if the recommendation is strong or qualified, the kind of mammography and its periodicity.

4. Validation of the ontology.

The model and restrictions described in the previous section were implemented in the standard W3C ontology language OWL2 by using the Protégé editor. Part of the vocabulary of the implemented ontology was reused from the standard vocabulary of the NCI Thesaurus, e.g. Woman, Mammography and History [36]. The ontology consistency and inferences were verified by using the Hermit reasoner [37].

Figure 2: Reasoner entailments for a woman with intermediate risk and age range 40 to 44.

Figure 3: Reasoner entailments for a woman with intermediate risk and age range ≥ 55.

The ontology was tested with six different instances of the class Woman. Three of them meet conditions to be classified as with intermediate risk of breast cancer, each one within a different age range, one woman meets conditions to be classified as with intermediate risk but does not fall into any of the three age range, and the other two women do not meet conditions to be
classified as with intermediate risk of breast cancer. Figures 2 to 7 show inferences made by the reasoner for each Woman instance. The ontology is available in https://bioportal.bioontology.org/ontologies/MAMO-SCR-ONTO and has the https://creativecommons.org/licenses/by-sa/4.0/ license.

**Figure 4:** Reasoner entailments for a woman that does not meet all conditions for intermediate risk.

**Figure 5:** Reasoner entailments for a woman that does not meet any condition for intermediate risk.

**Figure 6:** Recommendation entailed for a woman with intermediate risk and age range 45 to 54.
5. Conclusions and future work

This paper presents a simple Mammography Screening Ontology (MAMO-SCR-Onto) designed to provide mammography application recommendations for the early detection of breast cancer in patients at average risk of the disease. Risk stratification allows for personalized recommendations regarding the frequency, the recommendation strength, and the need for screening mammography tailored to an individual’s risk profile. In this sense, MAMO-SCR-Onto is only applied to patients with average disease risk, where the risk is defined based on three conditions of the woman’s clinical history. However, the risk could be obtained by applying different evaluation risk models outside the current ontology. This is not necessarily seen as a limitation of the MAMO-SCR-Onto ontology, in the sense that its design aimed to focus the recommendation for exclusively medium-risk patients while maintaining its simplicity to be a helpful tool in doctor-patient communication, and in the training of preservice health professionals.

As a main conclusion, from the point of view of final users, the advantage of using the MAMO-SCR-Onto is that medical professionals and patients can access guideline recommendations in a non-verbose presentation. Notably, the ontology makes it explicit whether these recommendations are classified as strong or qualified, facilitating physician and patient to question the benefits and harms of conducting the study and subsequently assisting decision-making by the patient. Ultimately, this comprehensive ontology enables individuals to make well-informed decisions regarding their healthcare attention.

From the point of view of ontology engineers, another conclusion of our work is regarding ontology maintainability and reusability quality attributes. The MAMO-SCR-Onto represents the conditions that women must meet to receive specific recommendations, established in the ACS guidelines. However, using the ACS guide is only a showcase proposal since it is no difficulty to change ACS guide for any other one. The recommendations in general terms established by the guidelines are concordant, and variations occur when setting the age range in the recommendations by group and the frequency of screening, these constraints are very well identified in the ontology, and they are straightforward the change.

As a future work, the ontology could be extended for women at high risk of breast cancer, representing the recommendations provided for them. More precise information about the
corresponding type of risk assessment could be included, such as being based on a statistical model of breast cancer risk assessment, for example, the Tyrer-Cuzick, Clause and Gail models, that the physician at the time of the consultation can assess the risk of breast cancer of the woman with greater precision and in turn provide the relevant recommendations [43, 44, 45].

References


date accessed July 2023.


