A Semantic AI Conceptual Prototype for the Semantic Cancer Knowledge Framework (SCKF)*

Fatima Zahra Amara1*, Mounir Hemam1 and Meriem Djezzar1

1ICOSI Laboratory, University of Abbes Laghrour, Khenchela, Algeria

Abstract
Cancer, a formidable challenge in modern medicine, claims millions of lives annually, emphasizing the urgency of multidisciplinary approaches for understanding, diagnosis, therapy, and patient care. The power of data and semantic modeling in the age of information offers unprecedented potential to revolutionize oncology research, empower clinicians, engage patients, and drive innovation in cancer care. Semantic modeling, a vital component of AI, enables machines to work with knowledge akin to human understanding. This paper introduces the Semantic Cancer Knowledge Framework (SCKF), uniting healthcare, data science, and semantic web technologies. The SCKF envisions a semantically enriched ecosystem where cancer-related data, knowledge, and expertise converge, fostering collaboration across traditional boundaries. The objective of this research is to leverage semantic modeling and AI to advance cancer research, diagnosis, treatment, and patient care. The Conceptual Prototype for SCKF represents a visionary endeavor poised to transform the landscape of cancer-related knowledge management and decision support.

Keywords
Semantic Modeling, Ontology, Healthcare, Artificial Intelligence, Semantic AI

1. Introduction
Cancer remains one of the most challenging hurdles in modern medicine, necessitating multidisciplinary methods to better understanding, diagnosis, therapy, and patient care. It significantly contributes to global mortality, resulting in approximately 9.3 million deaths each year [1]. Harnessing the power of data and semantic modeling in the age of information gives an unprecedented potential to enhance oncology research, empower clinicians, engage patients, and drive innovation in cancer care.

Cancer, with its diverse manifestations and intricate molecular underpinnings, requires a sophisticated approach to integrate, interpret, and make sense of the vast array of data generated in clinical practice, research laboratories, and academic institutions. Hence, Artificial intelligence is now playing a critical role in the fight against cancer. We can now analyze vast amounts of medical data, such as imaging, genetic tests, and patient medical history records,
using these smart technologies. Early detection is one of the most significant benefits of using artificial intelligence to combat cancer. Current cancer research efforts have resulted in massive collections of cancer-associated data that could be exploited for cancer prediction and early diagnosis [2].

Machines that are semantically disjointed and physically related contribute to a lack of semantic interoperability [3]. The imperative to enable machine-computable reasoning, inferencing, and knowledge discovery for achieving more insightful and meaningful outcomes has led to the development of semantic interoperability [4]. Semantic modeling is a crucial aspect of AI that enables machines to work with knowledge and information in a way that resembles human understanding. It is a procedure that explicitly organizes knowledge; this organized knowledge may then be explored and visualized for a variety of decision-making activities [5, 6]. This research paper delves into the conceptualization and design of the Semantic Cancer Knowledge Framework (SCKF) initiative at the intersection of healthcare, data science, and semantic web technologies. The SCKF aspires to provide a unified, semantically enriched ecosystem where cancer-related data, knowledge, and expertise converge, fostering a collaborative environment that transcends the boundaries of traditional silos.

The goal of this research is to use semantic modeling and AI to advance cancer research, diagnosis, treatment, and patient care. The Conceptual Prototype for SCKF represents a visionary endeavor poised to transform the landscape of cancer-related knowledge management and decision support. It incorporates a cancer-specific ontology, semantically annotates diverse data sources, and constructs a dynamic Knowledge Graph. User-friendly interfaces cater to stakeholders, ensuring data privacy and continuous updates. SCKF aims to revolutionize cancer knowledge management, clinical support, and patient engagement.

The remainder of this paper is organized as follows: Section 2 represents this research background. Section 3 explore similar works. The SCKF Ontology and the data annotation are presented in Section 4. Section 5 demonstrate The System Architect of SCKF. Conclusion and future work outlined in Section 6.

2. Background

2.1. Cancer as a Global Health Challenge

Cancer poses an enduring and formidable global health challenge, affecting millions of lives worldwide\(^1\). This complex group of diseases is characterized by uncontrolled cell growth and can manifest in various forms, often with devastating consequences. Cancer knows no boundaries, transcending geographic, demographic, and socioeconomic barriers [7]. Its impact extends beyond the physical suffering of patients to encompass emotional, financial, and societal burdens. While progress has been made in understanding cancer’s mechanisms and developing treatments, disparities in access to care, prevention, and early detection persist. Moreover, the rising prevalence of risk factors like tobacco use, unhealthy diets, and sedentary lifestyles threatens to exacerbate the cancer burden in many regions. Tackling this global health challenge necessitates collaborative efforts on a global scale, involving governments, healthcare systems,

\(^1\)https://www.who.int/healthtopics/cancer#tab=tab_1
researchers, and communities, to promote prevention, improve access to quality care, and advance innovative therapies, ultimately striving for a world where cancer is no longer a leading cause of suffering and mortality.

There are several significant hurdles to addressing cancer as a global health issue:

• Promoting cancer prevention strategies such as tobacco control, healthy diets, physical activity, and vaccination against certain cancer-causing viruses (e.g., HPV and hepatitis B) remains a challenge, particularly in regions with limited public health infrastructure and education.

• In many parts of the world, especially in low and middle income countries, access to healthcare services, including cancer screening, diagnosis, and treatment, is limited or unequal. This results in delayed diagnosis and inadequate care, which can significantly impact outcomes.

• Early detection through screenings and regular check-ups is crucial for improving cancer outcomes. However, there are disparities in access to screening programs and a lack of awareness about the importance of early detection in many communities.

• Cancer treatments, especially novel therapies and targeted drugs, can be prohibitively expensive. Access to affordable cancer care and medications is a challenge for individuals and healthcare systems, leading to financial hardship for many patients.

• While there have been significant advances in cancer research, there is still much to learn about the complexities of the disease. Funding, collaboration, and access to research findings can be challenges for researchers working towards improved cancer treatments and prevention.

2.2. Role of Semantic AI in Healthcare Cancer

Artificial intelligence in healthcare refers to the use of software or "machine-learning algorithms" to replicate human "cognition" in the study, display, and understanding of complex health and medical care data[9]. In order to anticipate cancer, AI can examine and understand "multi-factor" data from several patient assessments and provide more precise information regarding patient survival, prognosis, and disease progression predictions[10]. Semantic AI provides a foundation for automating end-to-end difficult operations. It employs a variety of machine learning and logic-based methodologies, as well as background knowledge frequently stored in knowledge graphs[11].

Healthcare can reap substantial benefits from data, yet grappling with the extensive and intricate information generated through increased digitization necessitates the implementation of effective strategies for extracting valuable insights. Semantic Artificial Intelligence plays a vital role in healthcare, particularly in areas like cancer care, as it leverages AI and semantic techniques to enhance data analysis, decision-making, and ultimately, patient outcomes.

Artificial Intelligence enabled by semantic technologies presents a wealth of opportunities for enhancing efficiency in healthcare. In this new era, there’s a notable decrease in errors, a

---

3https://www.labiotech.eu/in-depth/cancer-barriers-to-research/
significant acceleration in the generation of advanced data insights, and a newfound freedom for staff to concentrate on delivering exemplary care.

The Semantic AI in healthcare, particularly in the context of cancer:

- Harmonizes diverse data sources for comprehensive patient profiles;
- Extracts insights from text data like clinical notes and research papers;
- Offers personalized treatment recommendations;
- Accelerates drug candidate identification;
- Tailors treatments based on genetics and clinical data;
- Connects eligible patients with trials;
- Assists in knowledge discovery from medical literature;
- Educates and empowers patients;
- Ensures privacy and security of patient information.

Semantic AI in healthcare, particularly in cancer care, revolutionizes the way data is processed, understood, and applied.

3. Related Work

In the realm of cancer research and healthcare informatics in semantic field, a plethora of innovative approaches and solutions have emerged to tackle the multifaceted challenges posed by the detection, treatment, and management of various types of cancer. This section provides an overview of pertinent research endeavors that have significantly contributed to this field.

In [12] authors developed an ontological model based on the decision tree algorithm for predicting breast cancer. The researchers extracted rules from the decision tree algorithm that distinguish between malignant and benign breast cancer patients and implemented these rules in the ontological reasoner using the Semantic Web Rule Language (SWRL). The results showed that the ontological model achieved a high prediction accuracy of 97.10%. This approach combines machine learning and ontological reasoning to provide reliable predictions for breast cancer detection.

This paper [13] is the introduction of a privacy-preserving dashboard for F.A.I.R (Findable, Accessible, Interoperable, Reusable) head and neck cancer data. The paper addresses the challenges of reusing real-world clinical data by proposing a federated learning approach, specifically the Personal Health Train (PHT), which allows for the distribution of models to data centers instead of centralizing datasets. The paper also emphasizes the importance of making data semantically interoperable using ontologies and knowledge representation.

The authors of [14] the development of a framework called DIGGER, which analyzes mined logical rules to uncover meaningful insights in the context of lung cancer treatments. DIGGER incorporates the semantics of Knowledge Graphs (KGs) and uses logical rules to identify missing information, errors in data, and potential violations of clinical guidelines.

The paper [15] aim to the development of a framework for automatically generating a disease-symptom knowledge graph integrated with standardized ontologies. The framework utilizes reliable online medical resources and disease named entity recognition (DNER) models to construct the knowledge graph. The integrated knowledge graph provides a base for intelligent
expert advisor healthcare systems, enabling disease prediction and symptom checking for both normal users and medical professionals.

The researcher in [16] proposes the implementation of a clinically meaningful use case for federated learning in the context of head and neck cancer. The authors focused on preparing the data in a FAIR (Findable, Accessible, Interoperable, and Reusable) manner and developing a visual data exploration dashboard. They also developed a prognostic model for survival using both clinical and image-based features, validated it through an internal-external validation procedure, and achieved this without the need for exchanging individual-level patient data.

The authors of [17] developed a knowledge graph (KG) and ontology-based approach for cancer diagnosis and biomarker discovery. The KG integrates domain-specific knowledge from scientific literature and other external sources, allowing for more accurate and comprehensive analysis of cancer-related data. The paper also proposes a BERT-based information extraction method for enriching the KG with valuable information from scientific articles.

The study [18] makes significant contributions in the field of lung cancer knowledge graph construction and classification. It leverages Latent Dirichlet Allocation (LDA) for topic modeling and classification of lung cancer articles, optimizing topic numbers based on coherence metrics. A novel PMI_2 weight is introduced for weighted knowledge graph construction, enhancing it with four graph neural network (GNN) algorithms and a PMI_2 + link method for improved classification. The study rigorously evaluates the classification performance using GNN algorithms and builds a comprehensive knowledge graph of lung cancer using Neo4j.

Authors in [19] developed the Operational Ontology for Oncology (O3). O3 is a professional society-based, consensus-driven approach that aims to address standardization gaps in the field of radiation oncology. It focuses on standardizing data and methods, improving communication and documentation, and increasing interoperability of data across different datasets. The goal of O3 is to facilitate clinical practice improvement, research, and the aggregation and learning from real-world data.

The paper [20] propose and implement a fuzzy ontology for breast cancer diagnosis and knowledge representation. The paper introduces a framework that uses fuzzy logic to reduce vagueness in the crisp breast cancer ontology. The resulting fuzzy ontology is evaluated and shown to effectively reduce vagueness, providing a valuable resource for computational reasoning and knowledge-based systems in breast cancer detection and diagnosis.

The authors [21] developed a knowledge-based bladder cancer treatment infrastructure (BCTECI) that integrates research on patient conditions and bio-structural levels. The infrastructure incorporates health information standards to ensure interoperability and uses semantic concepts to classify and segment randomized controlled trials (RCTs). It also supports treatment protocol development and provides tailored evidence-based feedback to guide treatment decisions for oncologists, patients, and caregivers. The BCTECI aims to improve the effectiveness of bladder cancer treatments, reduce complications, and accelerate the inclusion of RCT evidence in clinical decision-making.

Table1 presents a concise overview of several research papers in the field of cancer research and healthcare informatics in the semantic domain. Each paper is listed with its reference, highlighting its main contribution, the technologies employed, and, where relevant, the specific cancer domain it addresses.
4. Ontology and Data Annotation

The most important part of the Semantic Web is the ontology, which facilitates semantic interoperability by serving as a repository for data and knowledge about objects and their kinds [22]. SCKF-Onto, a cancer-specific ontology, is a structured representation of concepts, entities, and their relationships in the domain of cancer knowledge and management. It is designed to organize, categorize, and semantically link various elements related to cancer research, diagnosis, treatment, and patient care.

Figure 1: A Part of SCKF-Onto

Here is a description of key components and features of the SCKF-Onto and Figure 1 depicts a part of the developed ontology under Protege⁴:

1. **Cancer Types:** The ontology defines a taxonomy of cancer types, encompassing a wide range of malignancies such as breast cancer, lung cancer, prostate cancer, and many

⁴https://protege.stanford.edu/
2. **Clinical Data**: Concepts related to patient-specific clinical information, including medical history, diagnostic reports, treatment records, symptoms, and outcomes.

3. **Genomic Data**: Entities representing genetic information, mutations, biomarkers, and molecular profiles relevant to cancer.

4. **Treatment Modalities**: Categories for different cancer treatment approaches, including surgery, chemotherapy, radiation therapy, immunotherapy, and targeted therapy.

5. **Biomarkers**: Molecular markers associated with cancer diagnosis, prognosis, and treatment, such as oncogenes, tumor suppressor genes, and specific protein markers.

6. **Healthcare Professionals**: Representation of professionals involved in cancer care, such as clinicians, oncologists, nurses, and researchers.

7. **Patients**: Entities for patients, including their medical histories, diagnoses, treatment preferences, and reported outcomes.

8. **Research Papers**: Resources representing academic publications, research articles, and studies related to cancer research.

9. **Cancer Staging**: A structured representation of cancer staging systems, tumor size, lymph node involvement, and metastasis information.


11. **Drug Database**: A database of cancer-related drugs, their mechanisms of action, dosages, and side effects, aiding in treatment decision-making.

The SCKF ontology provides a structured framework for organizing, retrieving, and analyzing diverse data sources in the field of cancer, fostering collaboration, advancing research, and ultimately improving patient outcomes. Its use of semantic modeling enhances the understanding of complex cancer-related data and facilitates the development of predictive models and personalized treatment strategies. The ontology serves as a foundational element in the SCKF, enabling the framework to fulfill its mission of transforming cancer knowledge management and decision support.

5. **Semantic Cancer Knowledge Framework Conceptual prototype**

The Semantic Cancer Knowledge Framework (SCKF) is a conceptual framework for managing cancer-related knowledge that is structured and interrelated. This framework provides as a core structure for organizing and utilizing oncology-related information, data, and insights. SCKF provides a complete and integrated approach to cancer knowledge management, analysis, and utilization. By offering an organized and integrated knowledge environment, it anticipates to improve cancer research, clinical decision support, and patient engagement. SCKF contributes to the progress of cancer research and the improvement of cancer prevention, diagnosis, and treatment results.
5.1. SCKF Architecture

The Semantic Cancer Knowledge Framework (SCKF) architecture is intended to assist in the management and application of cancer-related knowledge through semantic modeling, data integration, and user interfaces adapted to different stakeholders.

The visual schema (Figure 2) represents the core components of the Semantic Cancer Knowledge framework (SCKF) and their relationships. The Clinical Interface, Patient Interface, Cancer Research Interface, Data Integration Hub, Semantic Annotation Engine, Knowledge Graph, Clinical Decision Support, Patient Support Interface, and Data Analyst & Predictive Model are all integral parts of the platform, working together to achieve its objectives in cancer research, diagnosis, treatment, and patient care.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Main Contribution</th>
<th>Used Technologies</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>[12]</td>
<td>Ontological model for breast cancer prediction using decision tree algorithm and SWRL</td>
<td>Decision tree algorithm, SWRL, ontological reasoning</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>[13]</td>
<td>Privacy-preserving dashboard for head and neck cancer data with a focus on federated learning</td>
<td>Federated learning, data privacy, ontologies</td>
<td>Head and neck cancer</td>
</tr>
<tr>
<td>[15]</td>
<td>Development of a disease-symptom knowledge graph integrated with standardized ontologies</td>
<td>Disease Named Entity Recognition (DNER), ontologies</td>
<td>Disease-symptom, general</td>
</tr>
<tr>
<td>[16]</td>
<td>Implementation of clinically meaningful federated learning for head and neck cancer data</td>
<td>Federated learning, FAIR data principles</td>
<td>Head and neck cancer</td>
</tr>
<tr>
<td>[17]</td>
<td>Knowledge graph and ontology-based approach for cancer diagnosis and biomarker discovery</td>
<td>Knowledge Graphs, ontologies, BERT-based information extraction</td>
<td>Cancer, general</td>
</tr>
<tr>
<td>[18]</td>
<td>Contribution to lung cancer knowledge graph construction and classification</td>
<td>Latent Dirichlet Allocation (LDA), Graph Neural Networks (GNN), PMI_2 weight</td>
<td>Lung cancer</td>
</tr>
<tr>
<td>[19]</td>
<td>Development of the Operational Ontology for Oncology (O3) to address standardization gaps</td>
<td>Ontologies, data standardization</td>
<td>Radiation oncology, general</td>
</tr>
<tr>
<td>[20]</td>
<td>Implementation of a fuzzy ontology for enhancing breast cancer diagnosis and knowledge representation</td>
<td>Fuzzy logic, ontology, breast cancer ontology</td>
<td>Breast cancer</td>
</tr>
</tbody>
</table>

Table 1
Cancer Research Overview in the Semantic Domain.
Figure 2: Overview of the architecture of the SCKF
Following gives an outline of the SCKF architecture’s essential components and their interactions:

5.1.1. User Interfaces

- **Clinical Interface**
  This interface serves as a portal for clinicians to interact with the SCKF. Clinicians can input patient data and access personalized treatment recommendations based on semantic analysis and the latest research. It allows them to make informed decisions about patient care.

- **Patient Interface**
  The patient interface is designed to empower patients by providing them with information about their cancer diagnosis, treatment options, and support resources. It allows patients to actively participate in their care, make informed decisions, and find clinical trials or support groups.

- **Cancer Research Interface**
  Researchers can use this interface to access the SCKF’s knowledge graph for hypothesis generation, literature review, and exploration of research areas. It provides a platform for discovering hidden connections between genes and cancer types, accessing relevant research papers, and collaborating with other researchers.

- **Clinical Decision Support**
  This module integrates with healthcare systems to provide real-time clinical decision support based on the semantic model. It assists clinicians in selecting personalized treatment options and monitoring patient progress, enhancing the quality of patient care.

- **Patient Support Interface**
  The patient support interface complements the patient interface by providing additional support resources, connecting patients with support groups, and helping them navigate their cancer journey. It aims to improve the overall well-being of patients beyond medical treatment.

5.1.2. Semantic Annotation Engine

This engine is responsible for automatically annotating incoming data with semantic tags from the SCKF-Onto ontology. It adds structured metadata to various data sources, such as patient records or research papers, enabling the system to understand and link them based on their semantic meaning.

5.1.3. Data Integration Hub

The data integration hub is the interface responsible for collecting, harmonizing, and integrating diverse cancer-related data sources. It acts as a central repository for electronic health records (EHRs), genomics data, imaging data, clinical trial information, and more, making the data accessible for analysis and decision support.
5.1.4. Knowledge Graph

The knowledge graph visually represents the relationships and dependencies between different cancer-related concepts. It plays a central role in the system, enabling intuitive exploration of cancer knowledge, supporting semantic searches, and facilitating data analysis and predictive modeling.

5.1.5. Data Analyst and Predictive Model

Data analysts use the semantically enriched dataset and the knowledge graph to create predictive models for cancer outcomes. These models aid in early detection, treatment planning, and research. They play a critical role in leveraging data to improve cancer care and research.

The predictive model in the Semantic Cancer Knowledge Framework (SCKF) plays a crucial role in enhancing the platform’s capabilities for improving cancer research, clinical decision support, and patient care. The predictive model serves various key functions, including: Early Detection and Diagnosis, Risk Assessment, Treatment Personalization, Optimization of Treatment Plans, Outcome Prediction, Research Support and Data-Driven Insights.

Figure 3: SCKF Predictive Model

The predictive model in the SCKF is a powerful tool that leverages advanced technologies. It employs deep learning to analyze medical images, extracting invaluable insights to aid in
early cancer detection and treatment assessment. Through natural language processing, it comprehends and analyzes textual information from research papers and clinical notes, ensuring a comprehensive understanding of the latest research findings and patient records. Furthermore, reinforcement learning continually optimizes treatment strategies, adapting to individual patient responses and outcomes, thereby maximizing the effectiveness of care. The overarching role of this predictive model is to revolutionize cancer care by offering precision, evidence-based decision support, placing the patient at the center of the treatment journey, and ultimately advancing the field of oncology for the betterment of patient outcomes. Figure 3 depicts and details the SCKF predictive model.

5.2. SCKF Use Case Diagram

The use case diagram in the context of the Semantic Cancer Knowledge Framework (SCKF) serves as a visual representation of the system’s functionalities from the perspective of its users or stakeholders. It outlines the primary interactions and use cases within the framework.

Use Case Diagram illustrate the interactions between different actors (such as clinicians, researchers, data analysts, and patients) and the system. Show how each actor interacts with the system’s functionalities (Figure 4).

The use case diagram gives a clear and comprehensive overview of the SCKF’s major features and user interactions, assisting in ensuring that the system is tailored to satisfy the unique demands of physicians, researchers, data analysts, and patients. It is an extremely useful communication and planning tool for system design and development.

6. Conclusion and Perspective

In the relentless battle against cancer, the Semantic Cancer Knowledge Framework (SCKF) emerges as a beacon of hope, transforming the landscape of cancer knowledge management and decision support. The fusion of advanced technologies, semantic modeling, and artificial intelligence within the SCKF represents a visionary endeavor aimed at conquering the multifaceted challenges posed by this disease.

The SCKF serves as a bridge between data, knowledge, and expertise, fostering collaboration that transcends the limitations of traditional silos. Its design, underpinned by a cancer-specific ontology, semantic data annotation, and a dynamic Knowledge Graph, is set to reshape the way we understand, diagnose, treat, and care for cancer patients. Through its user interfaces, the SCKF places clinicians, researchers, data analysts, and patients at the center of a patient-centric ecosystem. It respects the paramount importance of data privacy and ensures that its knowledge is continuously updated with the latest research findings.

While this research paper introduces a conceptual prototype of the Semantic Cancer Knowledge Framework (SCKF) to illustrate the potential of our approach, it’s crucial to emphasize that the SCKF presented here is at a conceptual stage. However, the future perspectives outlined are based on the premise that these ideas could be implemented in practical, real-world applications of the SCKF. As the framework advances, these perspectives could serve as a roadmap for its continued development and impact in the field of cancer knowledge management.
Figure 4: Use Case Diagram for the SCKF

References


[14] D. Purohit, Mining symbolic rules to explain lung cancer treatments (????).


