

## Book Chapter

# A Strategy-Driven Semantic Framework for Precision Decision Support in Targeted Medical Fields

Sivan Albagli-Kim<sup>1</sup> and Dizza Beimel<sup>2,3\*</sup>

<sup>1</sup>Department of Mathematics and Computer Science, Seton Hall University, South Orange, NJ 07079, USA

<sup>2</sup>Department of Computer and Information Sciences, Ruppin Academic Center, Emek Hefer 4025000, Israel

<sup>3</sup>Dror (Imri) Aloni Center for Health Informatics, Ruppin Academic Center, Emek Hefer 4025000, Israel

**\*Corresponding Author: Dizza Beimel**, Department of Computer and Information Sciences, Ruppin Academic Center, Emek Hefer 4025000, Israel

Published **May 06, 2025**

This Book Chapter is a republication of an article published by Sivan Albagli-Kim and Dizza Beimel at Applied Sciences in February 2025. (Albagli-Kim, S.; Beimel, D. A Strategy-Driven Semantic Framework for Precision Decision Support in Targeted Medical Fields. Appl. Sci. 2025, 15, 1561. <https://doi.org/10.3390/app15031561>)

**How to cite this book chapter:** Sivan Albagli-Kim, Dizza Beimel. A Strategy-Driven Semantic Framework for Precision Decision Support in Targeted Medical Fields. Top 10 Contributions in Applied Sciences. Hyderabad, India: Academic Reads. 2025.

© The Author(s) 2025. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Author Contributions:** Conceptualization, S.A.-K. and D.B, methodology, D.B. and S.A.-K, validation, S.A.-K. and D.B, formal analysis, S.A.-K. and D.B, investigation, S.A.-K. and D.B; data curation, D.B. and S.AK; writing—original draft preparation, D.B. and S.A.-K, writing—review and editing, D.B. and S.A.-K, visualization, S.A.-K. and D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors declare no funding was received.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abstract

Healthcare 4.0 addresses modernization and digital transformation challenges, such as home-based care and precision treatments, by leveraging advanced technologies to enhance accessibility and efficiency. Semantic technologies, particularly knowledge graphs (KGs), have proven instrumental in representing interconnected medical data and improving clinical decision-support systems. We previously introduced a semantic framework to assist medical experts during patient interactions. Operating iteratively, the framework prompts medical experts with relevant questions based on patient input, progressing toward accurate diagnoses in time-constrained settings. It comprises two components: (a) a KG representing symptoms, diseases, and their relationships, and (b) algorithms that generate questions and prioritize hypotheses—a ranked list of symptom–disease pairs. An earlier extension enriched the KG with a symptom ontology, incorporating hierarchical structures and inheritance relationships to improve accuracy and question-generation capabilities. This paper further extends the framework by introducing strategies tailored to specific medical domains. Strategies integrate domain-specific knowledge and algorithms, refining decision making while maintaining the iterative nature of expert–patient interactions. We demonstrate this approach

using an emergency medicine case study, focusing on life-threatening conditions. The KG is enriched with attributes tailored to emergency contexts and supported by dedicated algorithms. Boolean rules attached to graph edges evaluate to TRUE or FALSE at runtime based on patient-specific data. These enhancements optimize decision making by embedding domain-specific goal-oriented knowledge and inference processes, providing a scalable and adaptable solution for diverse medical contexts.

## Keywords

Knowledge Graph; Semantic Reasoning; Medical Diagnostic; Decision-Support Systems; Semantic Technology

## 1. Introduction

Significant challenges arising from healthcare modernization and digital transformation, such as adopting home-based care, precision treatments, and personalized or remote drug delivery systems, are being addressed through Healthcare 4.0 [1]. This paradigm represents a shift toward innovative solutions that leverage advanced technologies to improve healthcare accessibility and efficiency. In this context, the role of semantic technologies is powered by large datasets and algorithms that warrant examination. Semantic technologies highlight the growing adoption of knowledge graphs (KGs)—graph-based data models valued for their ability to represent intricate knowledge structures [2]. KGs are particularly suited for representing medical data and enhancing the clinical decision-support capabilities of healthcare systems. For instance, Rotmensch et al. [3] demonstrated the potential of health KGs constructed from electronic medical records to improve the organization of complex patient information and enable more accurate diagnostic inferences.

Using KGs is aligned with the framework introduced in our previous works [4,5], designed to assist medical experts in decision making during patient encounters where a medical issue is presented. The framework is iterative by nature, consisting of

a series of iterations. In each iteration, the medical expert is prompted with a question (symptom), and the patient responds. Each iteration brings the expert closer to making a decision, ultimately leading to a diagnosis for the issue raised by the patient. The framework operates within a setting that includes, among other constraints, time limitations. This poses the challenge of helping the medical expert refine their decision within a restricted timeframe.

Our designed and developed semantic technology-based framework comprises two main components: (a) a KG used to represent medical data and the relationships between various elements [6]. Specifically, symptoms and diseases were represented as nodes, and directed edges linked symptom nodes to disease nodes when a symptom characterized a disease; (b) a set of dedicated algorithms designed to infer questions for the medical expert based on the knowledge captured in the KG and the information provided by the patient during the encounter. At the end of the encounter, the framework's final output is a prioritized list of hypotheses based on the likelihood of accurate diagnosis. Each hypothesis consists of a symptom and a disease. In [5], we enriched the KG by extending it with a symptom ontology [7]. This extension included the addition of new symptoms and hierarchical structures, such that the original symptoms in the KG were augmented with hierarchical constructs rooted in them. These structures incorporated new ontological symptoms, with the relationships within the structure defined as ISA edges. This enrichment also required an extension of the algorithms to support inheritance relationships. This enhancement increased the framework's ability to provide more questions to the medical expert and refine the prioritized hypotheses list. In [5], we presented a full implementation of the framework in Python, including evaluation tests to assess the quality of its outputs.

Our generic framework provides a general solution for the iterative nature of patient–expert interactions. In the current work, we advance the framework by addressing specific medical domains. The iterative nature persists in most domains, but

domain-specific information can be added to enable the medical expert to make more accurate decisions.

To this end, we reinterpreted the term *Strategy* to describe an extension of the framework incorporating a layer of knowledge and algorithms tailored to a specific medical domain. The following are sample strategies:

- Healthy lifestyle strategy promotes long-term wellness by incorporating attributes related to nutrition, physical activity, and preventive measures into the KG. By leveraging patient-specific data, the framework can suggest personalized recommendations associated with the patient's health goals and lifestyle.
- Mental Health Support Strategy integrates mental health assessments and decision-making tools to assist in diagnosing and treating psychological conditions.
- Emergency strategy prioritizes rapid and accurate decision making under time-critical conditions. It incorporates attributes such as symptom severity into the KG, enabling medical experts to identify potential diagnoses quickly.

Each proposed strategy retains the iterative nature of the medical decision-making process while incorporating distinct goals and additional knowledge layers. These layers are implemented by adding new attributes to the KG nodes and edges, supported by algorithms that refine the framework's generated hypotheses. The structure of the hypotheses may also change according to the needs of the specific medical domain.

This paper presents the concept of strategy through a case study of *emergency strategy*. This strategy simulates an emergency room environment, focusing on promptly identifying emergency conditions despite time-constrained communication between patients and medical experts. The goal of this strategy is to identify life-threatening diseases and save lives.

To determine the essential characteristics of the strategy, we conducted interviews with two physicians, which revealed the following key insights: (1) physicians often identify additional

symptoms during physical examinations or through abnormal vital signs not initially reported by patients; (2) personal patient data, such as age, gender, pre-existing conditions, and medications, play a critical role in the diagnostic process; (3) diseases are categorized as life-threatening or non-life-threatening based on symptoms and patient information; and (4) the strategy prioritizes the prompt elimination of life-threatening conditions as a primary objective.

This enhancement enriches the KG with semantic conditional edges reflecting patient-specific indicators, such as age, gender, and pre-existing conditions, while refining the framework's algorithms to prioritize excluding life-threatening diseases. The hypotheses are generated through an inference process that identifies symptoms to either confirm or rule out critical conditions. Simulating an emergency room environment, this enhancement enables the framework to focus on rapidly identifying life-threatening conditions in time-sensitive settings. Building on these findings, the architecture of our framework has been enhanced to provide medical experts with a prioritized list of hypotheses focusing on life-threatening diseases. This enhancement involved enriching the KG with conditional edges that incorporate patient-specific indicators, such as age, gender, and pre-existing conditions, and refining the framework's algorithms to emphasize the exclusion of life-threatening conditions. The hypotheses are generated through an inference process to identify symptoms that confirm or rule out critical diseases. By simulating emergency room scenarios, this enhancement enables the framework to prioritize the rapid identification of critical conditions in time-sensitive settings.

This strategy-based approach is both innovative and practical. By preserving the core iterative process and integrating domain-specific strategies, the framework remains adaptable across diverse medical domains while retaining its foundational strengths. The simplicity of the KG structure, combined with its ability to incorporate new attributes, enhances both the scalability and flexibility.

The paper is organized as follows: Section 2 discusses knowledge representation and reviews studies that utilize KGs for healthcare applications. Section 3 details our framework, and Section 4 describes the framework extension via tailored strategies for enhanced decision making. Section 5 describes in detail the case study of emergency strategy. Finally, Section 6 concludes with a summary of contributions and suggestions for future work.

## 2. Background and Prior Work

Knowledge representation (KR) plays a pivotal role in enabling entities to predict outcomes, form structured perceptions of the world, establish foundations for intelligent reasoning, facilitate efficient computations, and serve as a medium for human expression [8]. KR methods are extensively applied across diverse domains, including expert systems, artificial intelligence, and robotics. Many applications in these areas rely on KR methodologies for representation and reasoning [9] and are applied in various fields such as education, science, engineering, finance, and healthcare [10].

Prominent KR methodologies include knowledge graphs, ontologies, and semantic technologies. Knowledge graphs (KGs), also referred to as semantic graphs, encode relationships between entities within graph structures, providing semantically structured data that support various applications, such as question answering, recommendation systems [11], and information retrieval [12]. KGs are instrumental in advancing intelligent systems by enabling machines to process and utilize interconnected information effectively. Ontologies, defined as explicit machine-interpretable specifications of conceptualizations, describe entities within a domain, their attributes, and their interrelationships [13]. They provide a shared vocabulary for humans and machines, facilitating consistent information sharing, systematic analysis, and reuse of domain knowledge [14]. Ontologies are usually formulated in two main languages: RDF and OWL [15]. Semantic technologies extract meaning from data by integrating and managing diverse information streams. With ontologies often represented as

graphs, advanced graph algorithms can infer insights by modeling domain knowledge and data relationships.

In the context of healthcare, KR methods such as KGs and ontologies play a crucial role in representing complex medical relationships across multiple dimensions, including disease diagnosis, monitoring, and treatment, enabling precise decision making and inference within clinical frameworks [16]. Recent advancements in KR include integrating machine learning techniques with KGs, enabling dynamic updates, and improving inference capabilities. Additionally, KR serves as a cornerstone for explainable AI [17] by offering interpretable domain knowledge models.

Organizing medical knowledge with KR methodologies defines rules and relationships across various contexts, providing essential decision support in monitoring scenarios. This approach helps medical staff interpret data effectively, leading to informed decisions that enhance patient care [18]. In addition, diagnosing diseases relies on identifying specific illnesses through established classifications, supporting accurate and reliable diagnoses [19]. A methodology for constructing a graph of disease–symptom relationships, using data from 273 K patient records, is presented in [3].

Recent advancements in clinical decision-support systems have increasingly utilized KGs to improve diagnostic accuracy and deliver personalized care. For instance, constructing and evaluating causal KGs for diabetic nephropathy has enhanced clinical decision making by modeling intricate causal relationships within patient data [20]. Similarly, integrating KGs with large language models has been investigated to support real-time emergency decision making in critical care scenarios [21].

Clinical KGs incorporating proteomics data have also demonstrated potential in personalized medicine, enabling more precise diagnostics and customized treatments [22]. Moreover, enriching KGs with clinical narratives through natural language processing (NLP), named entity recognition (NER), and



biomedical ontologies has improved the extraction and organization of critical healthcare information [23].

These advancements highlight the transformative role of KGs in augmenting clinical decision-support systems, especially when combined with semantic technologies and patient-specific data. The authors of [3] focused on the automated construction of a KG from vast datasets, while the authors of [20,21] focused on using KGs for improved data modeling and real-time application, and those of [22,23] focused on enhancing data extraction and representation for improved personalization and organization.

Our research focuses on leveraging the KG to assist medical experts in refining and expediting their iterative decision-making process with patients. This paper introduces a novel approach to extend our framework by incorporating domain-specific knowledge into the KG. This is achieved by enriching the nodes and edges with attributes tailored to specific medical contexts and supported by dedicated algorithms. Some attributes are defined as Boolean rules attached to the graph's edges, which evaluate as TRUE or FALSE at runtime based on the patient's data.

### **3. The Framework**

The framework developed in our previous studies [4,5] supports collaborative decision making between a medical expert and a patient through an iterative exchange of questions and answers about symptoms and potential diseases. It assists the expert by suggesting relevant questions (e.g., “Does the patient exhibit a particular symptom?”) and refining the decision-making process based on the patient's responses. The output is a ranked list of hypotheses, each linking a specific disease to a related symptom. Thus, symptoms, diseases, and hypotheses form the framework's core elements.

The framework employs a knowledge graph (KG), which has become increasingly popular due to its ability to naturally represent interconnected data [24,25]. The KG's framework

consists of nodes representing symptoms and diseases, with edges (labeled symptomOf) linking symptoms to diseases. Leveraging the KG and the patient’s initial input, an inference process—driven by interactive algorithms—generates context-specific questions for the medical expert.

The framework operates in two stages: (1) a pre-processing phase during initialization and (2) a dynamic processing phase activated with each new patient interaction. These stages are described in the following sections.

### 3.1. Pre-Processing Phase

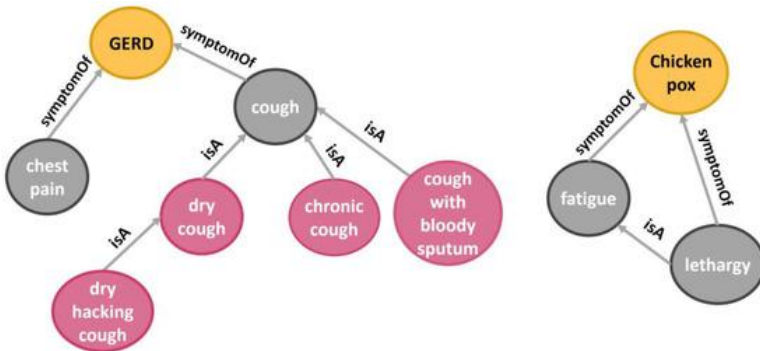
This phase involves constructing the framework’s KG, including creating nodes for symptoms and diseases and edges linking symptoms to diseases. In our previous studies [4], we demonstrated the KG’s construction using Neo4j Graph Database, Version 5 (<https://neo4j.com/>, accessed on 8 October 2024), with raw data from Kaggle [26]. The dataset comprises patient records, each associated with a diagnosed disease and reported symptoms, covering 41 diseases and 130 unique symptoms. While some symptoms are linked to a single disease, others are associated with multiple conditions.

Additionally, this phase enriches the KG by incorporating hierarchical structures. These structures consist of a root symptom node (a KG symptom) and additional ontology symptom nodes connected via ISA edges [5]. Semantic knowledge from the SYMP ontology [7] was used to create these structures, enhancing the KG’s capacity to generate a broader range of recommended questions for medical experts [27].

The final step in this phase involves identifying clusters of diseases—called *communities*—that share similar symptoms. For this, we applied the Louvain hierarchical clustering algorithm [28] to the KG, a process defined as “Algorithm 1” in [4].

Figure 1 presents a Neo4j screenshot illustrating a subgraph of the enriched KG. The example highlights the cough KG symptom node linked to the GERD disease node via

a symptomOf edge. Additionally, cough serves as the root of a hierarchical structure comprising ontology symptom nodes (e.g., dry cough), each connected to its parent node via an ISA edge.



**Figure 1:** An example of integrating a hierarchical tree of symptoms into the KG. Disease nodes are represented in yellow, KG symptom nodes in gray, and ontology nodes in red.

### 3.2. Processing Phase

The processing phase begins whenever a new interaction between a medical expert and a patient starts, with the patient presenting evidence symptoms. During this interaction (named “Algorithm 2” in [4]), the framework executes inference algorithms that utilize the identified communities to determine which diseases are compatible with the patient’s symptoms. Specifically, Algorithm 2 identifies the most probable diseases that align with the evidence symptoms. Next, “Algorithm 3” [4] iteratively, as needed, suggests to the medical expert questions (i.e., symptoms) that point toward the community most likely to include the patient’s disease. Finally, the processing phase concludes with “Algorithm 4” [4], which infers and outputs a ranked list of hypotheses (recall, ordered pairs of a disease and an indicative symptom) that the patient might have.

The entire framework was implemented in Python, and we conducted a series of tests to evaluate its output and effectiveness [5].

## 4. Extending the Framework: Tailored Strategies for Enhanced Decision Making

KR methodologies are essential for creating structured frameworks to organize and represent medical data. They enable healthcare professionals to effectively store and leverage critical information, such as risk factors, treatments, symptoms, and other patient-specific details [29].

Our generic framework addresses the iterative nature of patient–expert interactions across various medical domains. To enhance its precision, we introduce the concept of a strategy—a targeted extension designed to address specific medical decision-making needs and enable more accurate decisions in specialized medical contexts.

Strategies tailor the iterative framework to specific healthcare scenarios while preserving its core structure. Each strategy is defined by its goal and a set of characteristics, formulated as KG attributes and algorithm-based rules.

The modularity of the KG model simplifies the integration of new strategies. The KG, composed of nodes and edges, allows attributes to be easily appended or adjusted. This flexibility enables the KG to incorporate domain-specific attributes such as patient-specific factors, risk thresholds, or clinical priorities. An adapted set of algorithms further supports each strategy to ensure effective implementation.

In addition to the strategies proposed in the Introduction, other potential strategies include the following:

- Chronic disease management: focused on conditions like diabetes or hypertension, incorporating attributes for long-term monitoring and personalized treatment plans.
- Rehabilitation management: supporting post-surgical or injury recovery by tracking progress and suggesting therapy adjustments.

Integrating tailored strategies transforms the framework into a robust adaptable tool for medical decision making across diverse healthcare scenarios. This modularity enables

1. Targeted interventions: addressing specific healthcare challenges to enhance diagnostic precision and improve patient outcomes.
2. Personalized care: incorporating patient-specific attributes to generate customized contextually relevant hypotheses.
3. Scalability: facilitating seamless integration of new strategies without disrupting the fundamental framework.

Section 5 illustrates the strategy concept through a detailed case study of the emergency strategy.

## **5. The Emergency Strategy Case Study: Needs Analysis and Formalization**

### **5.1. Motivation for the Emergency Strategy**

The emergency strategy aims to prioritize identifying life-threatening conditions within time constraints. To define the characteristics of this strategy, we interviewed two physicians and gathered the following key insights:

1. Physicians identify additional symptoms beyond those patients report, often through physical examinations or abnormal vital signs (e.g., abnormal blood pressure).
2. Personal patient information—such as age, gender, pre-existing conditions, and medications—is critical to the diagnostic process.
3. Diseases, based on symptoms and patient data, are classified into two categories: life-threatening and non-life-threatening.
4. The strategy's primary objective is first to rule out life-threatening conditions.

The following subsections detail the characteristics of the strategy and its integration into our framework, including additional KG representations and new algorithms.

## 5.2. Emergency Strategy Overview

To implement the insights gathered, we performed two primary actions: (a) enhancing the KG to incorporate emergency-specific attributes and (b) modifying the processing phase to align with the emergency strategy.

### 5.2.1. KG Enhancement

The KG enhancement process consists of two key steps, executed during the pre-processing phase:

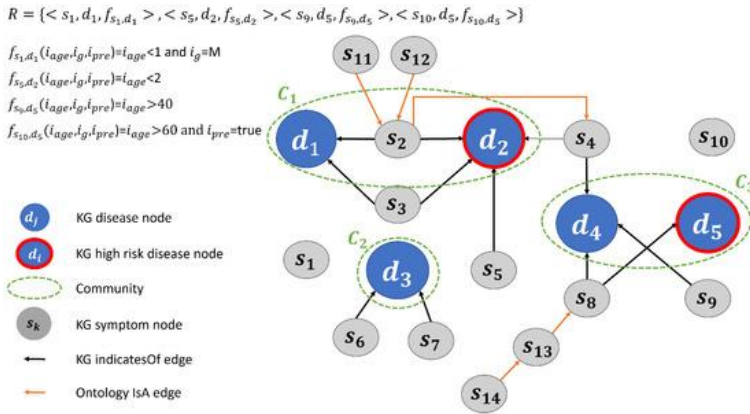
Step A: Adding a risk attribute to diseases:

A Boolean attribute named *risk?* is introduced for each disease in the KG, indicating whether the disease requires immediate attention to rule out life-threatening conditions.

Step B: Incorporating patient indicators as Boolean rules:

Three key indicators—age, gender, and pre-existing conditions—are integrated into the KG. To capture the influence of these indicators on life-threatening conditions, a new edge type, *conditional SymptomOf edges*, is defined. These edges include an attribute formulated as a logical rule involving one or more indicators connected with AND/OR operators.

For instance, if a symptom, *s1*, indicates a life-threatening disease, *d1*, in male patients over the age of 60, the rule governing the relationship between *s1* and *d1* would be “age > 60 AND gender = M”. This rule is represented using a Boolean function, where if the function evaluates to true, the edge *s1->d1* exists; otherwise, it does not. This implies that the edge is conditional, hence the term *conditional edge*. Refer to Figure 2 for the enhanced knowledge graph, which includes rules defining the *conditional SymptomOf edges*.



**Figure 2:** The enhanced knowledge graph.

**Step C: Categorizing SymptomOf edges:**

The KG distinguishes between two types of edges:

- Unconditional edges represent static relationships universally applicable to all patients.
- Conditional edges reflect relationships dependent on patient-specific indicators, dynamically included in the patient’s graph during runtime if the logical conditions are met.

This dynamic adjustment enables the KG to adapt to individual patient profiles, enhancing precision and personalization in the diagnostic process.

Figure 2 illustrates an enhanced KG including two diseases (d2, d5) marked as high-risk and conditional edges (e.g., the edge  $s_5 \rightarrow d_3$ , defined by the rule  $age < 2$ ).

**5.2.2. Processing Phase Enhancement**

The processing phase has been expanded with new algorithms to support the emergency strategy. The input process now includes a broader range of evidence, encompassing not only symptoms reported by the patient but also vital signs (e.g., blood pressure) and additional symptoms identified by the medical expert during the examination (e.g., rigid abdomen). These inputs, along with

the patient’s indicators—specifically age, gender, and pre-existing conditions—are incorporated into the patients’ KG instance and categorized as evidence symptoms.

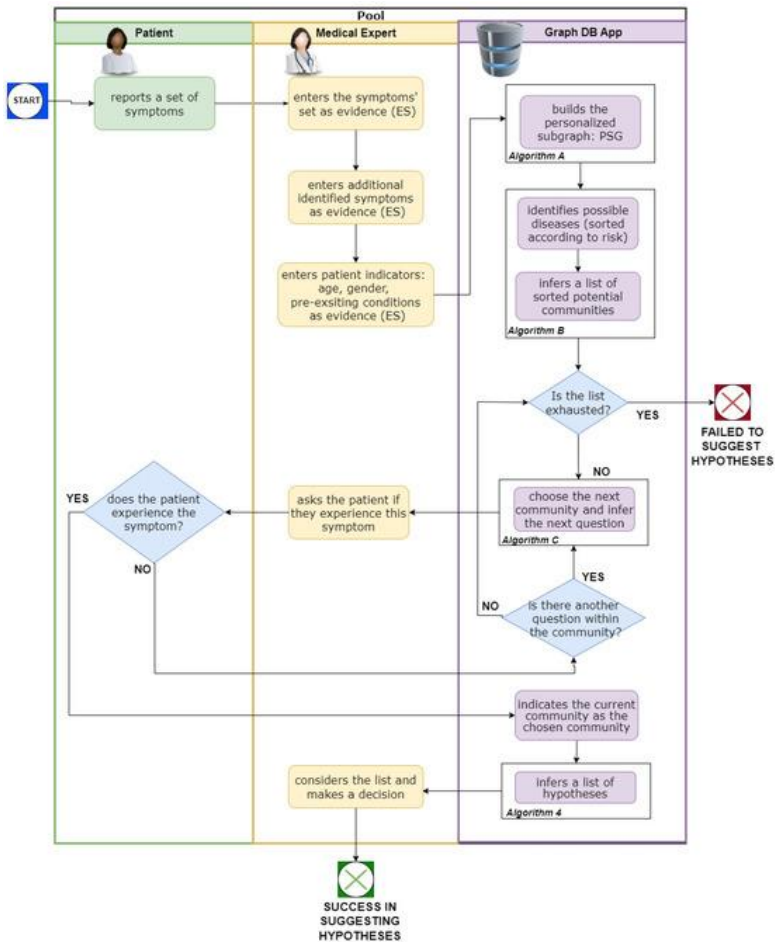
A new algorithm performs logical inference during runtime to handle the conditional edges introduced in the KG (See “Algorithm A” in Section 5.3.3). For each conditional edge, if its logical rule evaluates to true, it is dynamically added to the patient’s KG instance, ensuring that it reflects the patient’s specific profile. The output of this stage is named personalized subgraph or PSG in short (see Section 5.3.3 for more details).

With the patient’s PSG prepared, the disease identification and community inference processes proceed similarly to the original framework, with a slight modification to Algorithms 2 and 3: diseases are now prioritized in two stages: First, diseases marked with *risk? = true* are ranked, followed by all other diseases (see “Algorithm 2” in Section 5.3.3). Community rankings are subsequently determined based on disease scores, calculated by the number of evidence symptoms linked to a disease, including conditional edges evaluated as true (see “Algorithm 3” in Section 5.3.3).

Algorithm 4 remains as described in [4,5]. The framework identifies a symptom for each community to confirm or rule out a life-threatening condition. The inference process concludes with a ranked list of hypotheses, prioritizing any identified life-threatening diseases at the top.

Figure 3 illustrates the stages of the processing phase within the emergency strategy.





**Figure 3:** The interactions within the framework among the patient, the medical expert, and the KG during the processing phase in the emergency strategy.

### 5.3. Formalizing the Emergency Strategy

In this section, we provide a formal description of how the strategy aligns with the KG, which includes the refined KG process and is supported by the algorithms.

### 5.3.1. Knowledge Graph and Pre-Processing Formalizing

Refining the KG includes two main steps, as explained earlier. Both steps are implemented in the framework's pre-processing phase, as they do not involve the patient and remain consistent across patients.

**A:** Identify the diseases with high risk and add a Boolean attribute that recognizes them in the graph:

Let  $D$  be the set of nodes representing the diseases in the KG. For every disease  $d \in D$ , add a Boolean attribute named *risk?* with the default value *false*.

Let  $D_{risk} \subseteq D$  be the set of diseases with high risk. For each disease  $d \in D_{risk}$ , set *risk?* to *true*.

**B:** Incorporate the indicators of age, gender, and pre-existing conditions into the KG: This step translates a set of rules  $R$  into conditional edges  $E_C$  EC in the processing step. Each rule  $r \in R$  represented by a tuple  $\langle s, d, f_{s,d}(i_{age}, i_g, i_{pre}) \rangle$  where  $s$  is a symptom,  $d$  is a disease, and  $f_{s,d}$  is a Boolean function that receives three personal indicators and returns *true* if  $s$  indicates  $d$  according to the patient indicators.

The set of conditional edges  $E_C$  are defined as follows:

$E_C = \{(s, d) | f_{s,d}(i_{age}, i_g, i_{pre}) = true\}$ . These edges will be evaluated during the processing step when a patient arrives.

### 5.3.2. Framework-Specific Terminology

Table 1 presents the terminology that we use to describe the algorithms.

**Table 1:** The extended algorithms' terminology.

Term	Definition
$D$	The set of disease nodes
$D_{risk}$	Let $D_{risk} \subseteq D$ be the set of high-risk diseases
$S$	The set of symptom nodes
$ES$	The set of evidence symptoms (i.e., the symptoms indicated by the patient)
$PI$	The patient's personalized indicators
$C$	The set of communities
$ c $	The size of a single community $c \in C$ Defined by the number of diseases that belong to $c$
$Risk(c)$	Defined by the number of diseases in $D_{risk}$ + the number of evidence symptoms indicating a disease in $D_{risk}$
$LinD(c)$	The local-in-degree of a given $c \in C$ . Defined by the number of edges that point to diseases of $c$ , by ES; hence, it is the sum of $R^c(s,c)$ , for each $s \in ES$ and the given $c$
$PD$ 's communities	The set of communities $c \in C$ with a positive $LinD(c)$ , hence, a community in which at least one edge from $s \in ES$ points to $c$
$R^d(d)$	The disease's symptoms rank Defined by the number of symptoms the patient has that indicate $D$
$R_s(s,c)$	$R^c(s,c) - (\sum_{c \neq c' \in PD} R^c(s,c'))$ Defined by the number of edges from symptom $s$ to community $c$ minus the number of edges from $s$ to some other community $c'$ The outcome indicates how this symptom characterizes $c$
$C_s$	Community symptom Defines a symptom indicating a high number of diseases in the community $c$ and a low number out of $c$ ; hence, given a community $c$ , it is the symptom $s$ with the highest $R^s(s,c)$
$SCM$	symptom community matrix ( $SCM$ ) represents the associations between groups of diseases and the various symptoms (see [4] for a formal definition)

### 5.3.3. The Refined Framework Algorithms

We describe the additional algorithms developed in our framework to support the emergency strategy.

Algorithm 1 builds the personalized subgraph from KG by adding the patient’s personal information (the indicators). Algorithm 2 incorporates the patient’s symptoms into the personalized sub-graph and uses inference to generate a ranked list of potential diseases. This list is then used as input for Algorithm 3, which identifies the most probable community and a relevant question, and for Algorithm 4 in [4,5], which generates a set of hypotheses ranked by their urgency. Recall that each hypothesis is a pair consisting of a disease and a symptom indicating it.

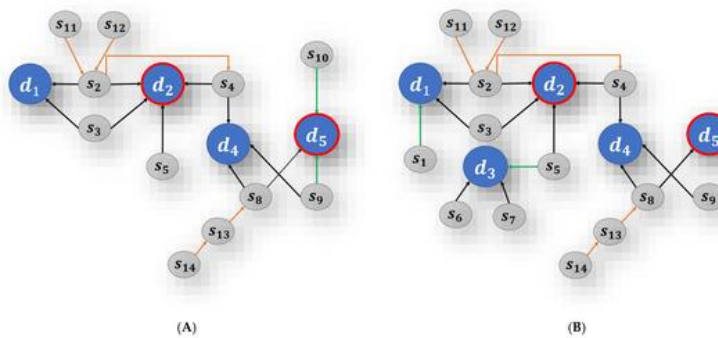
<b>Algorithm 1:</b> personalized sub-graph
<b>Input:</b> knowledge graph $KG = (DUS, E)$ , PI, ES, R
<b>Output:</b> personalized sub-graph PKG
Algorithm:
<ol style="list-style-type: none"> <li>0. Let PKG be KG.</li> <li>1. For every <math>s \in ES</math>: <ol style="list-style-type: none"> <li>a. For every <math>r \in R</math> that contains <math>s</math>, that is <math>r = \langle s, d, f_{s,d} \rangle</math>: <ol style="list-style-type: none"> <li>b. If <math>f_{s,d}(PI) = true</math>, add the edge <math>(s,d)</math> to PKG.</li> </ol> </li> </ol> </li> <li>2. Return PKG.</li> </ol>
<b>Algorithm 2:</b> identify possible diseases
<b>Input:</b> PKG, ES, C
<b>Output:</b> possible diseases, sorted according to their risk
Algorithm:
<ol style="list-style-type: none"> <li>1. Let <math>PD \leftarrow \{\}</math>.</li> <li>2. Let <math>C' \subseteq C</math> be the set of communities having positive LinD.</li> <li>3. Sort <math>C'</math> in a non-decreasing order according to their Risk (primary) and then according to their LinD (secondary).</li> </ol>

3.1. Let $c$ be the community in the order:
3.1.1 Go over the diseases in $c$ .
First, go over the diseases $d$ with $\text{risk}^d = \text{true}$ . Sort them according to their $R^d(d)$ (in a decreasing order) and add them in that order into PD.
Then, add the rest of the diseases in $c$ , sorted (in a decreasing order) according to their $R^d(d)$ .
4. Return PD.
<b>Algorithm 3:</b> find the most probable community
<b>Input:</b> PD
<b>Output:</b> $c_s$ and the community it indicates (presented as a question to the domain expert), or null if it does not exist.
Algorithm:
1. Let $C$ be the list of PD's communities, sorted according to the order of PD.
2. Let $c \in C$ be the current community in the order.
3. For every symptom $s \notin ES$ in $SCM(\_, c)$ , calculate $R^S(s, c)$ .
4. Let $s' = \text{argmax}_{s' \notin ES} R^S(s', c) > 0$ . If $R^S(s', c) > 0$ , return $s'$ (i.e., $c_s$ ) and $c$ .
Otherwise, return to step 2.
5. Return null.

## 5.4. Simplified Example

We illustrate two distinct scenarios involving patients who exhibit the same symptoms ( $s_1, s_5, s_9$ , and  $s_{10}$ ) but differ in their indicators, such as age and medical history. These differences lead to unique results in the generated knowledge graphs.

**Scenario 1: A 75-year-old man** The first scenario involves a 75-year-old man with no prior health conditions. Despite having the same symptoms as the second patient, his age and lack of medical history significantly impact the resulting knowledge graph. The patient’s age serves as a key personal indicator, influencing the identification of possible diseases. After processing this information, the resulting graph, referred to as PKG1, is generated. PKG1 reflects the relationships between the symptoms and potential conditions most relevant to this patient’s age group. Figure 4A shows this graph, providing a visual representation of how the patient’s characteristics, like age, affect the input graph for the analysis.



**Figure 4:** (A) PKG1—the graph for the 75-year-old man, (B) PKG2—the graph for the 9-month-old baby.

**Scenario 2: A 9-month-old baby** The second scenario involves a 9-month-old baby, also with no prior health conditions. While this patient shares the same symptoms (s1, s5, s9, and s10) as the 75-year-old man, the baby’s age significantly alters the resulting knowledge graph. In this case, age is a critical factor that redefines the potential conditions and diseases linked to the symptoms. The resulting graph, PKG2, is generated after inputting the baby’s indicators. PKG2 highlights the diseases and medical considerations that are more relevant to infants, reflecting the distinct medical risks associated with early childhood. Figure 4B presents this graph, showing how the patient’s age impacts the analysis and identifying relevant health concerns.

In both scenarios, despite the identical symptoms, the personal indicators—age and health history—lead to different interpretations and analyses, as evidenced by the differing graphs (PKG1 and PKG2). These graphs demonstrate the importance of considering individual patient characteristics when diagnosing and identifying potential conditions.

It is important to note that these two scenarios produce different graphs, meaning the algorithms process different inputs and generate distinct hypotheses for each patient. This variation in the inputs directly affects the communities and risks that are evaluated, ultimately influencing the hypotheses generated by the algorithms.

In the first scenario, the 75-year-old man with no prior health conditions, the graph (PKG1) only includes the communities C1 and C3. The algorithm analyzes these two communities and calculates the associated risks. The risk levels of each community—denoted as  $\text{Risk}(C1)$  and  $\text{Risk}(C3)$ —are compared. Since  $\text{Risk}(C3) \geq \text{Risk}(C1)$ , the algorithm determines that the disease most likely associated with community C3 should be prioritized for exclusion. In this case, disease **d5** is ruled out first.

In the second scenario, involving the 9-month-old baby, the graph (PKG2) includes a broader range of communities, as the personal indicator—age—results in a wider set of communities being considered. The algorithm considers all relevant communities: C1, C2, and C3. After evaluating the risks associated with each community, the algorithm compares them and finds that  $\text{Risk}(C1) \geq \text{Risk}(C2) = \text{Risk}(C3)$ . Since  $\text{Risk}(C1) \geq \text{Risk}(C2)$ , the disease associated with community C1 is ruled out first. In this case, disease **d2** is the first to be excluded.

### **Key differences in the scenarios:**

**Graph construction:** The primary difference between the two scenarios is the composition of the graph. In the first scenario, only two communities (C1 and C3) are involved, while in the second scenario, all communities are considered.

**Risk calculation:** The risk values of each community differ in the two scenarios due to the varying personal indicators (age, health history). These differences lead to distinct risk assessments, influencing which diseases are ruled out first.

**Disease exclusion:** The first disease to be excluded also differs between the two scenarios. In the first scenario, disease d5 is ruled out first, while disease d2 is excluded first in the second scenario. This is a direct result of the distinct graphs and the unique risk levels for each scenario.

In summary, while the symptoms are identical in both scenarios, the differing personal indicators (age, health history) lead to different graphs, which in turn cause the algorithms to process distinct inputs and generate unique hypotheses.

## 6. Discussion and Future Work

Knowledge representation and reasoning techniques assist doctors in decision making and inferring new information about diseases from previously represented data [30].

The current work presented in this paper introduces a new concept, which we have termed “strategy.” A strategy extends the existing framework designed to address specific needs in various medical domains. On the one hand, the strategy continues to implement the framework’s core functionality, which supports the iterative nature of interactions between the patient and the medical expert. On the other hand, it enables adding a layer representing a specific medical domain through attributes added to the KG, supported by dedicated algorithms.

In the case study we presented, some attributes are formulated as Boolean rules attached to the graph’s edges. During runtime, an instance of the KG is constructed for the patient interacting with the medical expert. Each edge associated with a Boolean rule undergoes evaluation, yielding either TRUE or FALSE based on the patient’s data. If the rule evaluates to TRUE, the edge is included in the patient’s graph; otherwise, it is excluded.



Building upon the foundational framework, the strategy concept extends its capabilities to address domain-specific needs in diverse medical contexts. Strategies leverage the KG's flexibility to introduce specialized attributes and relationships tailored to specific healthcare challenges. Each strategy is designed to enhance the framework's capacity for precise context-aware decision support while preserving the iterative nature of the decision-making process.

The emergency strategy is a case study focusing on rapidly identifying life-threatening diseases. Building on the work by Hornig et al. [31], who demonstrated the effectiveness of machine learning in sepsis clinical decision support, our approach integrates an emergency strategy layer into the KG-based decision-making framework. This enhancement prioritizes life-threatening conditions, improving decision-making efficiency and patient outcomes in urgent scenarios.

We advanced clinical decision making by developing a KG enriched with conditional edges informed by patient-specific data, enabling real-time adaptation to individual profiles. Additionally, we refined existing algorithms to seamlessly incorporate the emergency strategy, creating a diagnostic process that is both precise and responsive to critical clinical demands. These advancements establish a more adaptable framework for emergency contexts, providing a strong foundation for future innovations in medical diagnostics.

Future efforts will focus on validating the emergency strategy with real-world clinical data to assess its practical effectiveness in supporting healthcare professionals. Additionally, we plan to refine the logic for conditional edges by incorporating continuous data, enabling more granular inferences and improved diagnostic accuracy.

Expanding on this work, we aim to integrate machine learning models capable of dynamically updating the KG with incoming data, enhancing the framework's adaptability to evolving clinical practices and diverse patient populations.

The strategy concept offers potential for scalability across other medical domains requiring rapid and precise decision making, such as cardiology or oncology. Furthermore, addressing ethical considerations, such as mitigating bias and ensuring data privacy, will be a critical focus as we integrate machine learning into the framework. Collaborations with clinical practitioners and institutions will also play a pivotal role in co-developing and validating these innovations in real-world scenarios.

## References

1. Gupta A, Singh A. Healthcare 4.0: Recent advancements and futuristic research directions. *Wirel. Pers. Commun.* 2023; 129: 933–952.
2. Webber RJ, Eifrem E. *Graph Databases: New Opportunities for Connected Data*. Middlesex County: O'Reilly Media, Inc. 2015.
3. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a Health Knowledge Graph from Electronic Medical Records. *Sci. Rep.* 2017; 7: 59-94.
4. Albagli-Kim S, Beimel D. Knowledge graph-based framework for decision making process with limited interaction. *Mathematics.* 2022; 10: 39-81.
5. Beimel D, Albagli-Kim S. Enhancing Medical Decision Making: A Semantic Technology-Based Framework for Efficient Diagnosis Inference. *Mathematics.* 2024; 12: 502.
6. Rajabi E, Kafaie S. Knowledge graphs and explainable AI in healthcare. *Information.* 2022; 13: 459.
7. Symptom Ontology (SYMP), Ontology Lookup Service (OLS). Available online at: <https://www.ebi.ac.uk/ols/ontologies/symp>
8. Davis R, Shrobe H, Szolovits P. What is a Knowledge Representation? *AI Mag.* 1993; 14: 17–33.
9. Praveenkumar T, Sabhrish B, Saimurugan M, Ramachandran KI. Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox. *Measurement.* 2018; 114: 233–242.
10. Abu-Salih B. Domain-specific knowledge graphs: A survey. *J. Netw. Comput. Appl.* 2021; 185: 103076.

11. Guo Q, Zhuang F, Qin C, Zhu H, Xie X, Xiong H, He Q. A survey on knowledge graph-based recommender systems, *IEEE Trans. Knowl. Data Eng.* 2020; 34: 3549–3568.
12. Dietz L, Kotov A, Meij E. Utilizing knowledge graphs for text-centric information retrieval. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, MI, USA. 2018; 8–12: 1387–1390.
13. Gruber TR. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *Int. J. Hum.-Comput. Stud.* 1995; 43: 907–928.
14. Noy NF, McGuinness DL. *Ontology Development 101: A Guide to Creating Your First Ontology*; Stanford Medical Informatics Tech Report SMI-2001-0880. Stanford: Stanford Knowledge Systems Laboratory. 2001.
15. Chandrasegaran SK, Ramani K, Sriram RD, Horváth I, Bernard A, et al. The evolution, challenges, and future of knowledge representation in product design systems. *Comput.-Aided Des.* 2013; 45: 204–228.
16. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, et al. Development and validation of a deep learning algorithm for the detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016; 316: 2402–2410.
17. Borrego-Díaz J, Galán Páez J. Knowledge representation for explainable artificial intelligence. *Complex Intell. Syst.* 2022; 8: 1579–1601.
18. Sohn S, Larson DW, Habermann EB, Naessens JM, Alabbad JY, et al. Detection of clinically important colorectal surgical site infection using Bayesian network. *J. Surg. Res.* 2017; 209: 168–173.
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017; 542: 115–118.
20. Lyu K, Tian Y, Shang Y, Zhou T, Yang Z, et al. Causal knowledge graph construction and evaluation for clinical decision support of diabetic nephropathy. *J. Biomed. Inform.* 2023; 139: 104-298.
21. Chen M, Tao Z, Tang W, Qin T, Yang R, et al. Enhancing emergency decision-making with knowledge graphs and

- large language models. *Int. J. Disaster Risk Reduct.* 2024; 113: 104-804.
22. Santos A, Colaço AR, Nielsen AB, Niu L, Geyer PE, et al. Clinical knowledge graph integrates proteomics data into clinical decision-making. *bioRxiv.* 2020.
  23. Thukral A, Dhiman S, Meher R, Bedi P. Knowledge graph enrichment from clinical narratives using NLP, NER, and biomedical ontologies for healthcare applications. *Int. J. Inf. Technol.* 2023; 15: 53–65.
  24. Besta M, Blach N, Kubicek A, Gerstenberger R, Podstawski M, et al. Graph of thoughts: Solving elaborate problems with large language models. *Proc. AAAI Conf. Artif. Intell.* 2024; 38: 17682–17690.
  25. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* 2001; 29: 2724–2743.
  26. Kaggle. Available online at: <https://www.kaggle.com>
  27. Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, et al. Understanding the performance of knowledge graph embeddings in drug discovery. *Artif. Intell. Life Sci.* 2022, 2, 100036.
  28. Lu H, Halappanavar M, Kalyanaraman A. Parallel heuristics for scalable community detection. *Parallel Comput.* 2015; 47: 19–37.
  29. Tian S, Yang W, Le Grange JM, Wang P, Huang W, et al. Smart healthcare: Making medical care more intelligent. *Glob. Health J.* 2019; 3: 62–65.
  30. Lezcano-Valverde JM, Salazar F, León L, Toledano E, Jover JA, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Sci. Rep.* 2017; 7: 10189.
  31. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE.* 2017; 12: 0174708.