In proceedings of 4th Workshop on Knowledge Graphs and Big Data in Conjunction with IEEE BigData 2024

MedReg-KG: KnowledgeGraph for Streamlining Medical Device Regulatory Compliance

Subhankar Chattoraj and Karuna Pande Joshi
Department of Information Systems,
University of Maryland Baltimore County, Baltimore, USA
{*schatto1, karuna.joshi*}*@umbc.edu*

***Abstract*—Healthcare providers are deploying a large number of AI-driven Medical devices to help monitor and medicate patients. For patients with chronic ailments, like diabetes or gastric diseases, usage of these devices becomes part of their daily lifestyle. These medical devices often capture personally identifiable information (PII) and hence are strictly regulated by the Food and Drug Administration (FDA) to ensure the safety and efficacy of the medical device. Medical device regulations are currently available as large textual documents, called Code of Federal Regulations (CFR) Title 21, that cross-reference other documents and so require substantial human effort and cost to parse and comprehend. We have developed a semantically rich framework MedReg-KG to extract the knowledge from the rules and policies for Medical devices and translate it into a machine-processable format that can be reasoned over. By applying Deontic Logic over the policies, we are able to identify the permissions and prohibitions in the regulation policies. This framework was developed using AI/Knowledge extraction techniques and Semantic Web technologies like OWL/RDF and SPARQL. This paper presents our Ontology/Knowledge graph and the Deontic rules integrated into the design. We include the results of our validation against the dataset of Gastroenterology Urology devices and demonstrate the efficiency gained by using our system.**

***Index Terms*—code of federal regulations; compliance; semantic web; medical device; knowledge graph.**

I. Introduction

Artificial Intelligence and machine learning (AI/ML) algorithms have recently gained a lot of attention for their ability to identify and learn patterns automatically from larger datasets. These technologies hold great potential to enhance the efficiency and precision of healthcare delivery, capitalizing on the latest advancements in big data [1], [2]. Following the digitization of healthcare systems, the extensive and continuous data generated during patient care is captured and stored as Electronic Health Record (EHR) data. According to the National Academy of Medicine, the essential functions of EHR include health information and data, decision support, electronic communication and connectivity, patient support, administrative processes, and reporting, as well as population health management [3]. In recent decades, the usage of machine learning (ML) and deep learning (DL) has significantly advanced various applications, including communicable disease diagnosis [4], resource allocation through task prediction [5], patient diagnosis [6]–[8], length-of-stay prediction [9], cancer diagnosis, mortality estimation [10] from EHR data, medical images [11] [12]. Knowledge Graphs have been widely adopted to enhance data insights and complement EHR modeling.

Nevertheless, medical AI/ML introduces new challenges for society and, in particular, for current regulators like those in the U.S. These evolving technologies are driving the need to reassess existing procedures for medical device approval and post-market monitoring systems. In the early 20th century, the U.S. Food and Drug Administration (FDA) was tasked with the vital role of ensuring that drugs were safe and effective before they could enter the market [13]. Later, in 1976, amendments to the Federal Food, Drug, and Cosmetic Act expanded the FDA's oversight to include the safety of medical devices [14]. These devices fall under the jurisdiction of the FDA's Center for Devices and Radiological Health (CDRH). According to the Federal Food, Drug, and Cosmetic Act, a device can be defined as an instrument, implement, apparatus, machine, implant, or in vitro reagent, among other forms. FDA's definition of ML as a system that improves performance on a given task through training. While early AI systems primarily relied on decision rules, modern AI research emphasizes ML techniques. The FDA, responsible for ensuring the "safety and effectiveness" of medical devices, regulates the market entry of devices in the U.S. [15]. Advocacy groups concerned with consumer protection call for stricter regulatory control, while businesses, many physicians, and certain patient advocacy groups argue that current regulations stifle innovation and restrict patient access to new devices [16].

Medical devices are generally categorized by matching their description to the relevant sections in Title 21 of the Code of Federal Regulations (CFR), specifically Parts 862-892. The FDA has classified and outlined more than 1,700 distinct device types, organizing them into 18 medical specialty "panels," such as Cardiovascular or Ear, Nose, and Throat devices. These panels correspond to Parts 862 through 892 in the CFR. For each device classified by the FDA, the CFR provides a general overview that includes its intended use, the device's classification (Class I, II, or III), and details regarding marketing requirements [17]. Since 2015, regulations for device manufacturers have grown by 64%, with a total of 13,485 regulations in place by 2022 [18]. For Class III devices following the Pre-Market Approval (PMA) pathway, the approval process takes, on average, more than eight months. The estimated cost to bring a premarket 510(k) product from

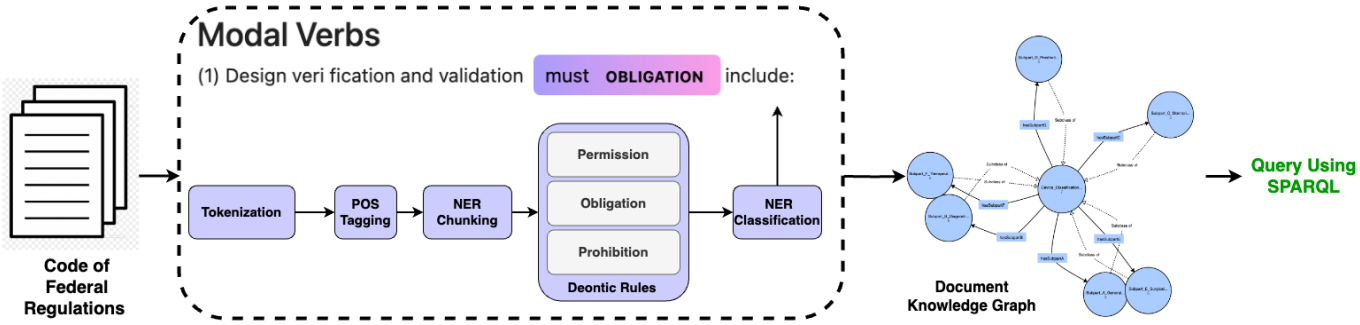

Figure 1. Overview of MedReg-KG, knowledge graph based architecture for automating medical device regulations.

concept to clearance is approximately 31 million, with 24 million of that amount allocated to FDA-related activities [19]. These documents are typically not in a format that can be easily processed by machines, necessitating considerable human effort to interpret and understand. As a result, device manufacturers face significant costs in complying with the FDA's regulations throughout the approval process.

We have created a novel semantically rich framework called MedReg-KG to extract knowledge from the rules and policies governing medical devices in a machine-readable format. MedReg-KG can be queried and reasoned over to identify complex rules relevant to various devices. MedReg-KG was developed using AI/Knowledge Management techniques and Semantic Web technologies, including deontic logic, OWL/RDF, and SPARQL. In this paper, we provide a detailed explanation of our approach and present the results of our validation. Our design aims to automate the pre-market analysis for 1,700 distinct types of devices regulated by the FDA, categorized into 18 medical specialty "panels". In Fig. 1, the overview of MedReg-KG, knowledge graph based architecture for automating medical device regulation is illustrated. MedReg-KG was validated using Part 876 *Gastroenterology Urology devices* [20]. Fig. 2 illustrates the proposed architecture. The key contributions of this paper are as follows:

- The MedReg-KG for medical device compliance is designed to reduce manual effort, lower costs, shorten approval times, and accelerate time-to-market by enabling automated analysis of Title 21 of the CFR Parts 862-892.
- To the best of our knowledge, our approach towards machine-processable compliance knowledge graph is first of it's kind that captures knowledge by identifying key terms, rules, topic summaries, relationships between various terms, semantically related terminologies, deontic expressions, and cross-referenced facts and regulations.

In this paper, Section II illustrated the related work. Section III introduces our methodology for constructing a semantically enriched, machine-readable compliance knowledge graph. Section IV outlines the experimental evaluation and validation of our approach. Finally, Section VI presents the conclusion and outlines potential directions for future research.

II. Related Work

A. Medical Device Regulation and Risk Classification

The Code of Federal Regulations (CFR) is a key collection of documents maintained by the U.S. Executive Branch [21]. It provides the public with a comprehensive repository of all regulations established by the president and federal agencies [22]. These regulations are critical to the government's operations and communication, offering necessary guidelines and standards across a variety of functions. Specifically, Title 21, Parts 800-1050, of the CFR covers all medical devices, with Parts 800-861 addressing cross-cutting device regulations and Parts 862-1050 focusing on device-specific requirements [23].

FDA classifies devices as different classes based on risk category [24]. The Class I devices are classified as the lowest risk and include products such as bandages, crutches, and tongue depressors [25]. Class II devices, like electrocardiographs, contact lens solutions, hearing aids, and orthopedic drills, are determined as moderate risk [26], while the Class III devices are categorized as the highest potential risk, include items such as implantable pacemakers, stents, heart valves, and HIV diagnostic tests [27]. Many Class I devices and certain Class II devices are exempt from premarket review and regulations. For exempt devices, companies are not required to seek FDA review or clearance but only need to notify the FDA of their intent to market these products. Class II devices, which present a moderate risk, are reviewed via the 510(k) premarket notification process. Class III devices, including life-sustaining or implantable devices, generally undergo the more stringent PMA process [28], which is the most rigorous process mandated by the FDA for devices. These devices require clinical evidence to support their application. However, if a Class III device only presents minor modifications from an existing, approved device (or "predicate" device), it may not need to undergo the strict PMA process. In such cases, the sponsor can petition the FDA for reassignment of the device using a 513(g) application. These devices can generally be approved through the less rigorous 510(k) process [29]. The FDA has classified and delineated over 1,700 distinct types of devices, arranging them within the CFR into 18 medical specialty "panels," which include categories like Cardiovascular, Orthopedic, and Radiology devices [17]. These devices

can generally be approved through the less rigorous 510(k) process [29]. The FDA has classified and delineated over 1,700 distinct types of devices, arranging them within the CFR into 18 medical specialty "panels," including categories like Cardiovascular, Orthopedic, and Radiology devices [17].

The classification and exemption process, outlined in Fig 2, requires a thorough review of devices based on their specific categories. For sponsors seeking clearance, understanding the relevant CFR codes is crucial to ensuring compliance with regulatory guidelines. However, this process is highly time-consuming and remains inaccessible to machines, as it is available only in textual form. This necessitates significant manual effort to interpret the rules and constraints, leading to higher costs and extended time-to-market.

B. Federal Regulation using Knowledge Graph

Previous research has highlighted the application of semantically rich knowledge graphs in fields such as drug discovery, predictive modeling, and healthcare compliance [30]–[35]. These studies have demonstrated how knowledge graphs can incorporate machine-processable rules to automate the monitoring of data operations, transfers, and sharing [36].

One example includes a knowledge graph designed for HIPAA-compliant cloud services, organized into key categories like privacy rules, security rules, and stakeholders [37], with additional classes like Health Information and detailed sub-classes [38]. Another study focused on COVID-related privacy and security regulations using HIPAA as a foundation to streamline compliance with patient record access [39]. Other research combined the Semantic Web with Ethereum Blockchain to enforce data protection [40], while encryption schemes have also been proposed to enhance healthcare security [41]. Many prior works have introduced knowledge graph-based methods for implementing data protection policies [42], [43]–[45]. Our recent publication details automating medical device regulations using a semantically rich knowledge graph [46].

C. Semantic Web

Semantic information derived from knowledge graphs can significantly improve search results in semantic-aware question-answering (QA) systems. IBM's Watson, a QA system utilizing multiple knowledge bases like YAGO and DBpedia, was designed to outperform human experts in the Jeopardy game show, showcasing the potential of KGs in such applications [47]. Semantic parsing-based QA systems operate by converting natural language questions into logical forms that represent the full meaning of the query. These logical representations are then used to create structured queries (such as SPARQL) to search knowledge bases for answers. They use Freebase to create a rough mapping between phrases and predicates, applying all relevant predicates, including neighboring ones and those generated through bridging operations, to form an accurate query and retrieve the correct answer [48].

Information retrieval-based QA systems, on the other hand, focus on automatically converting natural language questions into structured queries, from which they extract candidate answers from a knowledge base. They then analyze the features of both the question and candidate answers to rank them and identify the correct answer. In recent work, linguistic features like question words, verbs, focus, and topics are extracted to transform a question into a feature graph [49]. DL in natural language processing, many researchers have enhanced the performance of traditional QA methods by integrating deep learning techniques. In a DL-based framework, multi-column convolutional neural networks (MCCNNs) for information retrieval were employed to eliminate the need for manually crafted features and rules [50]. A scoring layer is used to rank candidate answers based on the representations of both questions and potential answers, yielding more accurate results.

D. Deontic Logic in Knowledge Graph

Deontic logic parser extracts key terms containing all the coverage and exclusion keywords based on three types of modalities: Permissions, Obligations, Prohibitions. In earlier work deontic logic extracted from insurance policies to standardized structure for policy formatting. The extracted rules was categorized into coverages (permissions) and exclusions (prohibitions) using deontic expressions [51]. In another work, Semantic Web, Deontic Logic, and Natural Language Processing (NLP) was utilized to reason over publicly available policies from seven different cyber insurance providers established by the United States Federal Trade Commission (FTC). Deontic expressions, including permissions and prohibitions, were leveraged to extract the relevant policy coverages and exclusions [52]. In another study, deontic logic was applied to classify the entire rule set into Permissions or Obligations. The framework details, along with results from analyzing a dataset of 3,000 privacy policies for GDPR compliance, are provided. The framework utilizes a BiLSTM multi-class classification approach combined with a BERT-based extractive summarizer. We assessed the framework's performance by measuring the context similarity between the summarized GDPR guidelines and the privacy policies of web service providers [36].

E. Adverse event analysis

FDA analysts invest a considerable amount of time searching for appropriate documents before they can access relevant information. In a recent study, CDRH developed the Semantic Search and Retrieval Framework (SARF), which aimed to expedite the process of locating documents [53]. In recently published work, RDF was utilized to detect errors like syntax errors [54]and logical inconsistencies, while inconsistency in Japanese medical devices was studied using SPARQL [55]. In another study, knowledge graph embedding techniques were utilized to design and train a customized Deep Neural Network (DNN) for predicting Adverse Drug Reactions. The knowledge graph comprised of drugs, ADRs, target proteins, indications, pathways, and genes entities [56] while a similar approach for classifying Adverse Drug Reactions (ADRs) was also

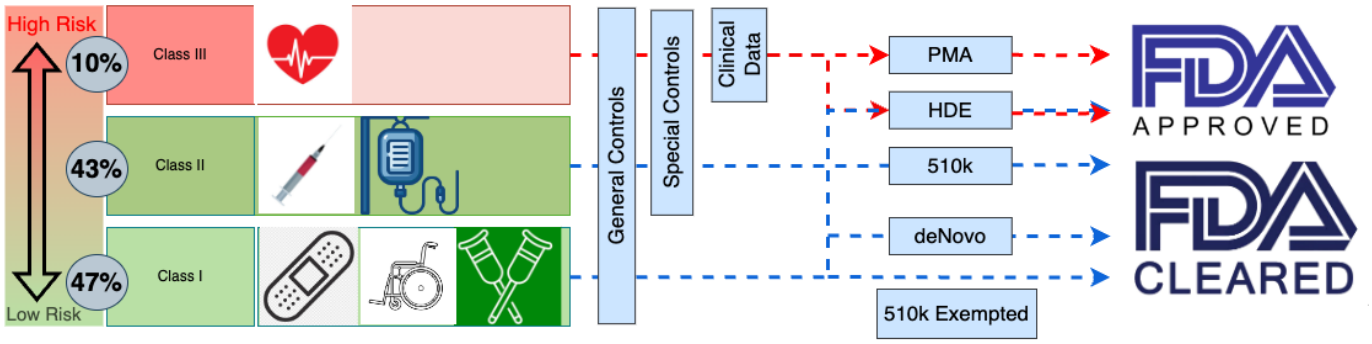

Figure 2. Overview of the FDA's medical device classification

proposed using knowledge graph, which leverages machine-readable interlinked representations of knowledge graph representing uniform heterogeneous data [57].

III. Methodology

In this section, we provide a detailed explanation of our MedReg-KG framework, including the structure of the CFR and key instances within it to automate the query analysis of Title 21, Parts 862-892, which covers 1,700 unique medical devices categorized by risk and indicates whether they require 510(k) notifications or pre-market approval [13]. In the Fig. 1 presents the overall system architecture.

A. Building MedReg-KG CFR -Title 21 from Parts 862-892

The knowledge graph was constructed in accordance with the structural guidelines and provenance detailed in Title 21 of the CFR, specifically Parts 862-892. This section of the CFR covers 18 medical specialties, each divided into multiple subparts that include general provisions and five device categories: diagnostic, monitoring, prosthetic, surgical, and therapeutic devices. There are 1,700 distinct devices across these categories. The knowledge graph captures key attributes such as device classification, title, volume, citation, section, and part number. The primary nodes in the graph represent different subparts of devices, as shown in Figure 3, which illustrates how devices are classified under these subparts. Each device node is detailed with attributes including identification, classification, device name, and references to relevant CFR regulations. The knowledge graph reflects the hierarchical organization of the regulations, starting from general provisions (Device_Classification_Panel) and progressively focusing on specific device categories and individual devices within each category.

The classes "Manufacturer" and "Device" allow query sponsors to seek information about a device's classification (Class I, II, or III) and its premarket notification (510(k)) or approval requirements. Queries are structured around subparts such as Subpart_A_General_Provision, Subpart_B_Diagnostic_Device, Subpart_C_Monitoring_Device, Subpart_D_Prosthetic_Device, Subpart_E_Surgical_Device, and Subpart_F_Therapeutic_Device (refer to Figure 3). The MedReg-KG offers a comprehensive and organized representation of the regulatory framework for medical devices, simplifying navigation, compliance evaluation, and understanding of FDA requirements. It also supports reasoning through SPARQL [58] queries, enabling connections between entities and rules at the most detailed subsections of the CFR. Developed with OWL [59] in the Protégé tool [60], this methodology ensures effective exploration and analysis of CFR regulatory information while facilitating advanced querying and inference.

B. Population of Knowledge Graph

Title 21 of the CFR covers Parts 862-892, which are made up of various sections, subsections, and sentences detailing FDA information and regulatory guidelines. Instances for the ontology were extracted from Title 21 and incorporated into the appropriate classes using the Protégé tool. Likewise, data properties and object properties of these instances were also added from Title 21 of the CFR via Protégé. Some key policy statements, which were instrumental in populating MedReg-KG, are highlighted below.

- 876.1080 Gastroenterology-urology accessories to a biopsy instrument [61]
  Identification: *A gastroenterology-urology accessory to a biopsy instrument is an accessory used to remove a specimen of tissue for microscopic examination by cutting or aspiration. This generic type of device includes a syringe for specimen aspiration and a biopsy channel adaptor. This device does not include accessories to biopsy instruments used in other medical specialty areas.*
  Classification: Class I *(general controls).*
  The device is ***exempt*** *from the premarket notification procedures in subpart E of part 807 of this chapter subject to the limitations in § 876.9.*
- 876.1075 Gastroenterology-urology biopsy instrument [62]
  Identification: *A gastroenterology-urology biopsy instrument is a device used to remove, by cutting or aspiration, a specimen of tissue for microscopic examination. This generic type of device includes the biopsy punch, gastrointestinal mechanical biopsy instrument, suction biopsy instrument, gastro-urology biopsy needle and needle set, and nonelectric biopsy forceps. This section does*

not apply to biopsy instruments that have specialized uses in other medical specialty areas and that are covered by classification regulations in other parts of the device classification regulations.
Classification: Class II *(performance standards).*

Algorithm 1 Extract Rules using Deontic Logic from CFR

Input: A list of texts `texts` to analyze for modal verbs.
Output: A dictionary of modal verbs categorized as `permissions_deontic` (PD), `obligations_deontic` (OD), and `prohibitions_deontic` (PDO).
Step 1: Load NLP Model
Load the Spacy Model
Step 2: Define Modal Categories
Define the sets of modal verbs:
`permissions_deontic = {''can'', ''may'', ''could'', ''might''}`
`obligations_deontic = {''should'', ''shall'', ''must'', ''Ought to'', ''Have to'', ''Need to'', ''Are required to'', ''Are obligated to''}`
`prohibitions_deontic = {''Must not'',''Should not'', ''Shall not'', ''Cannot'',''Are not allowed to'', ''Are forbidden to''}`
Step 3: Process Each Text
for each `text` in `texts` **do**
    Tokenize text: `doc = nlp(text)`
    Initialize an empty dictionary, `extracted_modals`:
    `extracted_modals = {}`
    `extracted_modals["PD"] = []`
    `extracted_modals["OD"] = []`
    `extracted_modals["PDO"] = []`
    Step 4: Identify and Categorize Modals
    for each `token` in `doc` **do**
        if `token` in `PD` **then**
            Add *token* to `extracted_modals["PD"]`
        else if `token` in `OD` **then**
            Add `token` to `extracted_modals["OD"]`
        else if *token* in *PDO* **then**
            Add *token* to `extracted_modals["PDO"]`
        end if
    end for
    Step 5: Output Results
    the corresponding `extracted_modals`.
end for

C. Deontic Logic: Rules Identification and Analysis in MedReg-KG

In this phase, we categorize the extracted elements, such as key term definitions and rules found in the sections, into fundamental deontic expressions using modal logic [63]. These terms and rules outlined in Title 21 of the CFR Parts 862-892 establish the permission, obligations, and prohibitions for key stakeholders, including federal agencies, organizations, and researchers. Our framework allows reasoning over these deontic rules to answer questions like "*Is the submission of the original design and verification of a device important?*", "*whether the device enhances reader performance as intended*". The responses to such inquiries should clearly define the three deontic expressions. Deontic logic includes three main modalities:

- Permissions: "can", "may", "could", "might". These describe the rights or authorizations granted to an entity.
- Obligations: "should", "shall", "must", "Ought to", "Have to", "Need to", "Are required to", "Are obligated to". These refer to mandatory actions that an entity is required to perform.
- Prohibitions: "Must not", "Shall not", "Cannot","Are not allowed to", "Are forbidden to". These outline actions that are explicitly forbidden.

Once we extracted all statements containing deontic expressions, we stored each as an instance in the class of our ontology. The detailed algorithm is represented in section III-B. The results of this phase are presented in Section IV.

IV. Experimental Validation

The design of MedReg-KG has been validated using a use case and one of our domain expert collaborators in medical device regulation.

A. Use case

In the following subsections, we outline two specific use cases of our system. Numerous other comparable use cases can be identified, depending on the application of the device.

1) Case 1: In a scenario where a device manufacturer is preparing a 510(k) notification to the FDA, they need to verify the classification and regulatory details of their device. The initial step involves checking if the device falls within a specific classification category, as predefined by the FDA. The SPARQL query assists in this process by providing essential information about device classifications, including the types and attributes associated with various devices. By querying this data, the manufacturer can confirm whether their device aligns with the existing classifications and regulatory standards, ensuring accurate submission of the 510(k) notification.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
    ↪ syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
    ↪ schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#
    ↪ >
PREFIX fdadc: <http://www.semanticweb.org/
    ↪ medicalresearch/FDADeviceClassification#
    ↪ >

SELECT ?subject ?predicate ?object
WHERE {
  ?subject ?predicate ?object.
}
```

The result of the query provide crucial information about the Vendor_ID, Manufacturer, Device_ID, Identification, subparts, Section_Name, and Cite the rule associated with it.

2) Case 2: In this use case, a medical device manufacturer is preparing a detailed submission for FDA approval and needs to ensure that their device classification aligns with existing FDA predicate devices. The SPARQL query retrieves information about devices, their descriptions, and classifications from an ontology of FDA device classifications. For

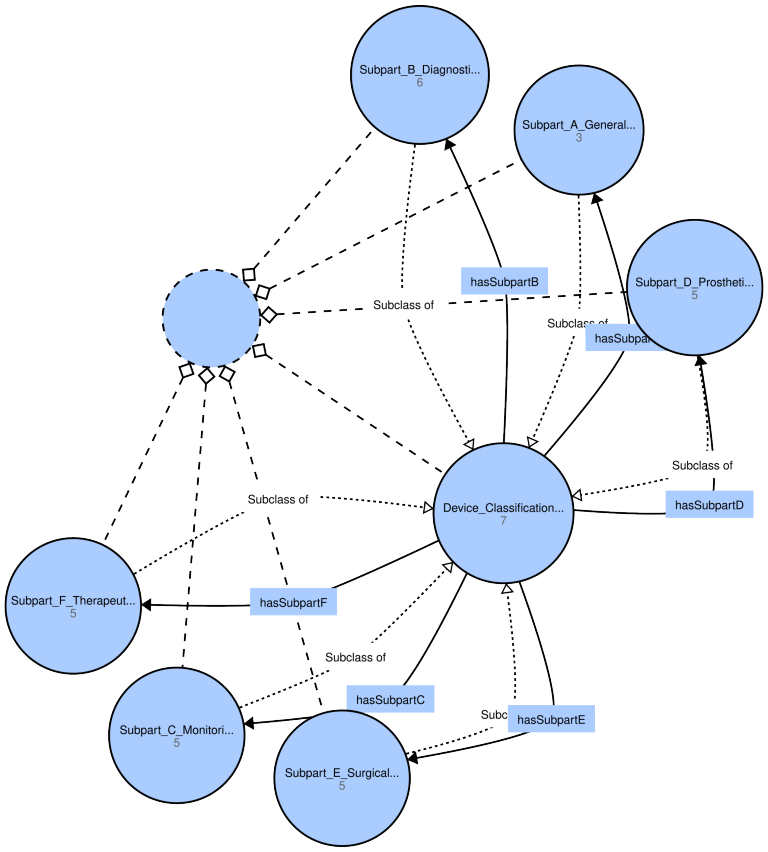

Figure 3. MedReg-KG: CFR -Title 21 medical device KG top-Level Classes

instance, if the query results indicate that a device described as *Gastroenterology-urology fiberoptic retractor* falls under "Class I (premarket approval)," the manufacturer can use this information to confirm the correct regulatory pathway. This ensures that the device is categorized properly according to FDA standards, facilitating accurate and compliant submission of documentation for approval.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
    ↪ syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
    ↪ schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#
    ↪ >
PREFIX fdadc: <http://www.semanticweb.org/
    ↪ medicalresearch/FDADeviceClassification#
    ↪ >

SELECT DISTINCT ?device ?description ?
    ↪ classification
WHERE {
  ?device rdf:type fdadc:Device .
  ?device fdadc:Device_Description ?description
      ↪  .
  ?device fdadc:ControlledBy/fdadc:
      ↪ ClassifiesDevice/fdadc:hasSubpartE ?
      ↪ subpart .

  ?subpart fdadc:Section_Name ?sectionName ;
         fdadc:Classification ?classification .

  FILTER(STR(?description) = STR(?sectionName))
}
```

The SPARQL query the class "Subpart_E_Surgical_Devices" and retrieves *Class I (general controls)* and the *Exempted from the premarket notification procedures* is retrieved.

B. Rule-Based Decision using Deontic Expressions

Through our framework, we extracted deontic expressions found within Title 21 of the CFR, Parts 862-892, and categorized each statement based on its deontic type. The system effectively identified permissions, obligations, and prohibitions embedded in these regulations. Below are some of the several deontic rules from Title 21 of the CFR that our system successfully extracted:

- Permission: "The pH electrode is at the end of a flexible lead which may be inserted into the esophagus or stomach

through the patient's mouth. The device may include an integral gastrointestinal tube." [CFR Title 21, Volume 8 Sec. 876.1400 Stomach pH electrode].

- Obligation. "The device design should ensure that the EGG signal is distinguishable from background noise that may interfere with the true gastric myoelectric signal." [CFR Title 21, Volume 8 Sec. 876.1735 Electrogastrography system].
- Prohibition: "The summary should not only include the percentage of patients in which a polyp was correctly identified by capsule endoscopy, but also the percent of patients in which the capsule either missed or falsely identified a polyp with respect to the clinically acceptable alternative structural imaging method." [CFR Title 21, Volume 8 Sec. 876.1330 Colon capsule endoscopy system].

V. Discussion

The categorization of medical devices by risk is based on the 1,700 distinct device types specified in CFR Title-21. These regulations, which govern medical devices, are currently presented in comprehensive textual formats, often containing references to other related documents. As of now, there are no specific regulatory standards in place for implementing AI in healthcare software or device applications. However, in April 2018, FDA Commissioner Scott Gottlieb highlighted the potential of AI in healthcare, stating, "AI holds enormous promise for the future of medicine, and we're actively developing a new regulatory framework to promote innovation in this space and support the use of AI-based technologies." Despite these efforts, under the existing regulatory system, medical devices and software applications are still evaluated by the FDA under traditional pathways (e.g., 510(k), PMA, or De Novo). Recently, the FDA released a discussion paper outlining a proposed regulatory framework for modifications to AI/ML-based software as a medical device (SaMD) that may require premarket review.

Given the complexity of these regulations, understanding and navigating them can be a labor-intensive process. Consequently, device manufacturers often face significant costs during the regulatory submission process as they work to meet the extensive rules and requirements set forth by the FDA. In this work, we have developed MedReg-KG to automate the representation of the 1,700 device types from CFR Title-21, which are categorized into 18 medical specialty panels. MedReg-KG is validated using SPARQL queries, as described in section IV-A. The FDA's first step is to classify a device based on its risk level into Class I, II, or III, as shown in Figure. 2. In section IV-A1, the initial process involves determining if a device belongs to a particular classification category, as defined by the FDA. SPARQL queries help streamline this process by providing critical information about device classifications, including the attributes and types of various devices. Class III devices, which carry the highest risk to patients, require the most stringent review process by the FDA. Pre-market approval (PMA) involves submitting clinical data to support their safety and effectiveness. Additionally, Class III devices are vital for maintaining or supporting human life, preventing deterioration of health, or addressing potentially severe risks of illness or injury. In section IV-A2, a manufacturer preparing a submission for FDA approval must ensure that their device's classification aligns with existing FDA-approved predicate devices. The SPARQL query output provides details on whether the device is categorized as *Class I (general controls)* and whether it is *exempt from premarket notification requirements*.

AI-based models are increasingly gaining clearance for use in diagnostic applications. Regulated by the FDA's CDRH, a total of 108 and 139 AI/ML-based medical devices were approved in 2023 and 2022, respectively, representing 35% of all medical devices approved to date [48]. Since 2015, regulations for medical device manufacturers have increased by 64%, with 13,485 regulations in place by 2022 [18]. The approval process can be cumbersome, and approval times may vary depending on the FDA's workload. MedReg-KG aims to alleviate the burden by translating CFR rules into a machine-readable format that supports reasoning and streamlines compliance. As with previous technological advancements, the FDA must evolve its regulatory frameworks to keep pace with AI innovations.

VI. Conclusion and Future Work

Regulatory documents for medical devices, like CFR Title 21, are traditionally managed as extensive text-based files, requiring substantial manual analysis due to their complexity and volume. This process is both time-intensive and costly. In this paper, we introduce MedReg-KG a novel framework that leverages knowledge representation and Semantic Web technologies to automate pre-market processes governed by FDA medical device policies, particularly those outlined in CFR Title 21. MedReg-KG categorizes devices by risk level and determines whether they require 510(k) notification or pre-market approval. The validation of this research focuses specifically on CFR Title 21 Part 870, within "*Subpart H - Medical devices Part 876 Gastroenterology-Urology Devices*."

This research aims to create an efficient and automated Question and Answer (QnA) system to help manufacturers and regulators minimize the human effort and costs involved in navigating medical device regulations. In collaboration with domain experts, we are continuously validating and enhancing the design of our knowledge graph to actively populate with all 1,700 approved devices.

VII. Acknowledgement

We extend our gratitude to Dr. Andrea Iorga for her assistance in the expert validation of our knowledge graph design. This work was partially funded by the NSF under award number 2310844, IUCRC Phase II UMBC: Center for Accelerated Real-Time Analytics (CARTA).

REFERENCES

[1] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature Biomedical Engineering*, vol. 2, pp. 719 – 731, 2018.

[2] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," *Stroke and Vascular Neurology*, vol. 2, pp. 230 – 243, 2017.

[3] E. Kim, S. M. Rubinstein, K. T. Nead, A. P. Wojcieszynski, P. E. Gabriel, and J. L. Warner, "The evolving use of electronic health records (ehr) for research." *Seminars in radiation oncology*, vol. 29 4, pp. 354–361, 2019.

[4] S. Chattoraj, B. Reddy, M. Tadepalli, and P. Putha, "Comparing deep learning models for tuberculosis detection: A retrospective study of digital vs. analog chest radiographs," *Indian Journal of Tuberculosis*, 2024.

[5] S. R. Blake, N. Das, M. Tadepalli, B. Reddy, A. Singh, R. Agrawal, S. Chattoraj, D. Shah, and P. Putha, "Using artificial intelligence to stratify normal versus abnormal chest x-rays: External validation of a deep learning algorithm at east kent hospitals university nhs foundation trust," *Diagnostics*, vol. 13, 2023.

[6] S. Pratiher and S. Chattoraj, "Diving deep onto discriminative ensemble of histological hashing & class-specific manifold learning for multi-class breast carcinoma taxonomy," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1025–1029, 2018.

[7] S. Pratiher, S. Chattoraj, D. Nawn, M. Pal, R. R. Paul, H. Konik, and J. Chatterjee, "A multi-scale context aggregation enriched mlp-mixer model for oral cancer screening from oral sub-epithelial connective tissues," *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 1323–1327, 2022.

[8] S. Pratiher, S. Chattoraj, S. K. Agarwal, and S. Bhattacharya, "Grading tumor malignancy via deep bidirectional lstm on graph manifold encoded histopathological image," *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 674–681, 2018.

[9] A. Govindarajan, A. Govindarajan, S. Tanamala, S. Chattoraj, B. Reddy, R. Agrawal, D. Iyer, A. Srivastava, P. Kumar, and P. Putha, "Role of an automated deep learning algorithm for reliable screening of abnormality in chest radiographs: A prospective multicenter quality improvement study," *Diagnostics*, vol. 12, 2022.

[10] J. Feng, J. Lee, Z. Vesoulis, and F. Li, "Predicting mortality risk for preterm infants using deep learning models with time-series vital sign data," *NPJ Digital Medicine*, vol. 4, 2021.

[11] S. Pratiher and S. Chattoraj, "Diving deep onto discriminative ensemble of histological hashing & class-specific manifold learning for multi-class breast carcinoma taxonomy," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1025–1029, 2018.

[12] D. Nawn, S. Pratiher, S. Chattoraj, D. Chakraborty, M. Pal, R. R. Paul, S. Dutta, and J. Chatterjee, "Multifractal alterations in oral sub-epithelial connective tissue during progression of pre-cancer and cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 152–162, 2020.

[13] *U.S. Food and Drug Administration.*, 2020 (accessed February, 2023). [Online]. Available: http://www.fda.gov/AboutFDA/WhatWeDo/History/default.html

[14] A. Nolan, "Federal authority to regulate the compounding of human drugs," 2013.

[15] *US Food and Drug Administration Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)*, 2020 (accessed August, 2023). [Online]. Available: https://www.fda.gov/media/122535/download

[16] J. McCarthy, "What is artificial intelligence," 1998.

[17] *U.S. Food and Drug Administration. Title 21 of the Code of Federal Regulations (CFR), Parts 862-892*, (accessed November 08, 2023). [Online]. Available: https://www.fda.gov/medical-devices/classify-your-medical-device/device-classification-panels

[18] *MDR throughout the product lifecycle*, (accessed December 19, 2023). [Online]. Available: https://www.elsevier.com/industry/medical-device-regulation

[19] J. A. Johnson, "Fda regulation of medical devices," 2012.

[20] *U.S. Food and Drug Administration CFR - Code of Federal Regulations Title 21 PART 876 Gastroenterology Urology devices*, (accessed December 12, 2023). [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=876

[21] *Electronic Code of Federal Regulation*, 2020 (accessed December 16, 2023). [Online]. Available: https://www.ecfr.gov

[22] S. Mesner, "Medical device technology: Does federal regulation of this new frontier preempt the consumer's state common law claims arising from injuries related to defective medical devices?" *The Journal of Law and Health*, vol. 7, p. 253, 1993.

[23] *An Introduction to FDA's Regulation of Medical Devices*, 2020 (accessed March 18, 2024). [Online]. Available: https://www.fda.gov/media/123602/download

[24] *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback.*, (accessed March 24, 2024). [Online]. Available: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion/Paper.pdf

[25] *Learn if a Medical Device Has Been Cleared by FDA for Marketing*, (accessed March 19, 2024). [Online]. Available: https://www.fda.gov/medical-devices/consumers-medical-devices/learn-if-medical-device-has-been-cleared-fda-marketing

[26] *U.S. Food and Drug Administration. CFR-Code of Federal Regulations Title 21 vol. 8. Sec 878.5020 Nonabsorbable polyamide surgical suture*, 2003 (accessed January 12, 2024). [Online]. Available: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?FR=878.5020

[27] *U.S. Food and Drug Administration. CFR-Code of Federal Regulations Title 21 vol. 8 Sec 870.3610 Implantable pacemaker pulse generator*, (accessed March 11, 2024). [Online]. Available: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=870.3610

[28] *PMA Review Process*, (accessed June 18, 2024). [Online]. Available: https://www.fda.gov/medical-devices/premarket-approval-pma/pma-review-process

[29] *U.S. Food and Drug Administration. FDA and Industry Procedures for Section 513(g) Requests for Information Under the Federal Food Drug and Cosmetic Act: Guidance for Industry and Food and Drug Administration Staff.*, (accessed October 16, 2023). [Online]. Available: http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm209851.pdf

[30] Y. Li, B. Qian, X. Zhang, and H. Liu, "Graph neural network-based diagnosis prediction," *Big data*, 2020.

[31] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "Kame: Knowledge-based attention model for diagnosis prediction in healthcare," *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018.

[32] S. Sang, Z. Yang, X. Liu, L. Wang, H. Lin, J. Wang, and M. Dumontier, "Gredel: A knowledge graph embedding based method for drug discovery from biomedical literatures," *IEEE Access*, vol. 7, pp. 8404–8415, 2019.

[33] S. Sang, Z. Yang, L. Wang, X. Liu, H. Lin, and J. Wang, "Sematyp: a knowledge graph based literature mining method for drug discovery," *BMC Bioinformatics*, vol. 19, 2018.

[34] R. Walid, K. Pande, and L. Elluri, "Secure and privacy-compliant data sharing: An essential framework for healthcare organizations," 2024.

[35] D. young Kim, L. Elluri, and K. P. Joshi, "Trusted compliance enforcement framework for sharing health big data," *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4715–4724, 2021.

[36] K. P. Joshi, L. Elluri, and A. Nagar, "An integrated knowledge graph to automate cloud data compliance," *IEEE Access*, vol. 8, pp. 148 541–148 555, 2020.

[37] D. young Kim and K. P. Joshi, "A semantically rich knowledge graph to automate hipaa regulations for cloud health it services," *2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 7–12, 2021.

[38] K. P. Joshi, Y. Yesha, and T. W. Finin, "An ontology for a hipaa compliant cloud service," 2016.

[39] L. Elluri, A. Piplai, A. Kotal, A. Joshi, and K. P. Joshi, "A policy-driven approach to secure extraction of covid-19 data from research papers," *Frontiers in Big Data*, vol. 4, 2021.

[40] A. Mahindrakar and K. P. Joshi, "Automating gdpr compliance using policy integrated blockchain," *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on*

High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 86–93, 2020.
[41] J. K. . C. S. Walid, R., "Automating gdpr compliance using policy integrated blockchain," *Sci Rep 14, 7147*, 2024.
[42] K. P. Joshi, A. Gupta, S. Mittal, C. Pearce, A. Joshi, and T. W. Finin, "2016 ieee international conference on big data (big data) semantic approach to automating management of big data privacy policies."
[43] L. Elluri, S. S. L. Chukkapalli, K. P. Joshi, T. Finin, and A. Joshi, "A bert based approach to measure web services policies compliance with gdpr," *IEEE Access*, vol. 9, pp. 148 004–148 016, 2021.
[44] L. Elluri, K. Pande Joshi, and A. Kotal, "Measuring semantic similarity across eu gdpr regulation and cloud privacy policies," in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 3963–3978.
[45] K. Joshi, K. Pande Joshi, and S. Mittal, "A semantic approach for automating knowledge in policies of cyber insurance services," in *2019 IEEE International Conference on Web Services (ICWS)*, 2019, pp. 33–40.
[46] S. Chattoraj, R. Walid, and K. P. Joshi, "Semantically rich approach to automating regulations of medical devices," *2024 IEEE International Conference on Digital Health (ICDH)*, pp. 132–137, 2024.
[47] D. A. Ferrucci, E. W. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. M. Prager, N. Schlaefer, and C. Welty, "Building watson: An overview of the deepqa project," *AI Mag.*, vol. 31, pp. 59–79, 2010.
[48] J. Berant, A. K. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Conference on Empirical Methods in Natural Language Processing*, 2013.
[49] X. Yao and B. V. Durme, "Information extraction over structured data: Question answering with freebase," in *Annual Meeting of the Association for Computational Linguistics*, 2014.
[50] L. Dong, F. Wei, M. Zhou, and K. Xu, "Question answering over freebase with multi-column convolutional neural networks," in *Annual Meeting of the Association for Computational Linguistics*, 2015.
[51] K. Sane, K. P. Joshi, and S. Mittal, "Semantically rich framework to automate cyber insurance services," *IEEE Transactions on Services Computing*, vol. 16, pp. 588–599, 2021.
[52] K. Joshi, K. P. Joshi, and S. Mittal, "A semantic approach for automating knowledge in policies of cyber insurance services," *2019 IEEE International Conference on Web Services (ICWS)*, pp. 33–40, 2019.
[53] L. L. Singh, S. D. Sudarsan, R. P. Jetley, B. Fitzgerald, and M. G. Milanova, "Semantic search tool for adverse event reports of medical devices."
[54] J. Tao, L. Ding, and D. McGuinness, "Instance data evaluation for semantic web-based knowledge management systems," in *2009 42nd Hawaii International Conference on System Sciences*, 2009, pp. 1–10.
[55] A. Yagahara, T. Tanikawa, K. Ogasawara, and H. Yokoi, "Integration of japanese medical device adverse event terminologies," *Studies in health technology and informatics*, vol. 245, p. 1345, 2017.
[56] P. Joshi, V. Masilamani, and A. Mukherjee, "A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network," *Journal of biomedical informatics*, p. 104122, 2022.
[57] E. Muñoz, V. Nováček, and P.-Y. Vandenbussche, "Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models," *Briefings in Bioinformatics*, vol. 20, no. 1, pp. 190–202, 08 2017.
[58] *W3C. 2013. SPARQL 1.1 Overview.*, (accessed December 16, 2023). [Online]. Available: http://www.w3.org/TR/sparql11-overview/
[59] *Owl Web Ontology Language*, (accessed October 28, 2023). [Online]. Available: http://WWW.W3.ORG/TR/OWL-FEATURES/
[60] M. A. Musen, "The protégé project: a look back and a look forward," *AI matters*, vol. 1 4, pp. 4–12, 2015.
[61] *Gastroenterology-urology accessories to a biopsy instrument*, (accessed March 18, 2024). [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=876.1080
[62] *Gastroenterology-urology biopsy instrument*, (accessed March 26, 2024). [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=876.1075
[63] C. Allen, U. Nodelman, and E. N. Zalta, "Stanford encyclopedia of philosophy: a dynamic reference work," *2003 Joint Conference on Digital Libraries, 2003. Proceedings.*, pp. 383–, 2003.