Manuscript Template

Title
Knowledge Graph Applications in Medical Imaging Analysis: A Scoping Review

Authors
Song Wang,1† Mingquan Lin,2† Tirthankar Ghosal,3 Ying Ding,1 Yifan Peng2*

Affiliations
1. The University of Texas at Austin, US
2. Population Health Sciences, Weill Cornell Medicine, US
3. Institute of Formal and Applied Linguistics, Charles University, Czech Republic

†Co-first author.
*Corresponding author. Email: yip4002@med.cornell.edu

Abstract
Background. There is an increasing trend to represent domain knowledge in structured graphs, which provide efficient knowledge representations for many downstream tasks. Knowledge graphs are widely used to model prior knowledge in the form of nodes and edges to represent semantically connected knowledge entities, which several works have adopted into different medical imaging applications. Methods. We systematically searched over five databases to find relevant articles that applied knowledge graphs to medical imaging analysis. After screening, evaluating, and reviewing the selected articles, we performed a systematic analysis. Results. We looked at four applications in medical imaging analysis, including disease classification, disease localization and segmentation, report generation, and image retrieval. We also identified limitations of current work, such as the limited amount of available annotated data and weak generalizability to other tasks. We further identified the potential future directions according to the identified limitations, including employing semi-supervised frameworks to alleviate the need for annotated data and exploring task-agnostic models to provide better generalizability. Conclusions. We hope that our article will provide the readers with aggregated documentation of the state-of-the-art knowledge graph applications for medical imaging to encourage future research.

MAIN TEXT

1. Introduction
In recent years, incorporating structured domain knowledge into downstream tasks has drawn great research attention from industry and academia (Ji, et al. 2021). This is because domain knowledge provides a proper understanding of a field which can be represented as a knowledge graph that can facilitate efficient inference to empower downstream tasks.
A knowledge graph represents the actual facts in the form of structured graphs, including entities (e.g., realistic objects and general concepts) and the relationships between entities (Ji, et al. 2021). It provides semantically structured information that computers can interpret and promises to build more intelligent systems to solve numerous real-world problems.

Knowledge graphs (viewed as the graph structure) differ from knowledge bases in terms of the involvement of formal semantics for interpretation and inference over facts (Figure 1). Knowledge graphs (KGs) like DBPedia (Auer, et al. 2007), NELL (Carlson, et al. 2010) and Wikidata knowledge base (Vrandečić and Krötzsch 2015) have become instrumental in various machine learning applications, such as information retrieval (Sumithra and Sridhar 2020), information extraction (Fei, et al. 2021, Jaradeh, et al. 2021, Bastos, et al. 2021), question answering (Ma, et al. 2021, Banerjee and Baral 2020), and recommendation (Wang, et al. 2019, Wang, et al. 2020, Xiang, et al. 2021).

![Figure 1. Radiology knowledge graph example: NIH Chest X-ray labels based on RadLex and SNOMED_CT](image)

Within a biomedical setting, researchers can utilize knowledge graphs to tackle various realistic problems, for example, aiding efforts to diagnose patients (Xi, et al. 2021), exploring possible disease treatments (Dai, et al. 2021, Yu, et al. 2021) and identifying associations between biomolecules and diseases (Chilińskiab, Senguptab and Plewczynski 2021). Oftentimes, solutions require a process called representation learning, which is to learn the mappings between the knowledge graphs and low-dimensional graph representations in feature space (Ji, et al. 2021). The representation learning process aims to encode the local as well as the global structure of a knowledge graph, map it to an embedding that can be utilized by algorithms for downstream tasks. Among various knowledge graph applications in biomedicine, medical imaging (e.g., radiography, ultrasound, magnetic resonance imaging, etc.) serves as one of the most significant diagnostic aids that are readily available to physicians (Xie, et al. 2021).

Deep learning methods try to train algorithms to identify abnormal regions and tissue variations in a manner similar to human beings (Goodfellow, Bengio and Courville 2016). Medical record histories, previous diagnoses made by pathologists or radiologists are used...
to train the algorithms. The algorithm learns this with large amounts of data, and after analyzing thousands of iterations of different images and diagnoses, it eventually will learn to make some diagnoses. In the medical imaging analysis domain, knowledge graphs have drawn a lot of research attention. Though there are comprehensive survey papers for knowledge graph applications in biomedical informatics (Nicholson and Greene 2020), we have not seen any literature survey for knowledge graph applications in medical imaging analysis. To bridge this gap, we conducted a systematic review on knowledge graph applications in medical imaging analysis. According to our review, most studies applied knowledge graphs to specific topics such as disease detection, localization, and report generation. In this work, we describe various knowledge graphs applications in medical imaging analysis and then point out future directions that have yet to be explored.

2. Methods

First, we will introduce the background of knowledge graphs. Second, we will describe the data sources and search strategies, including the inclusion and exclusion criteria, and how we selected the articles. At last, we will talk about how we conducted data synthesis and analysis.

a. Knowledge Graphs in General

Recently, knowledge graphs have become a predominant part of many information systems where structured prior knowledge is needed. The concept of graphical knowledge representation can be traced back to 1956 when Richens proposed the idea of Semantic Net (Richens 1956); however, the community realized the importance of his work only belatedly. MYCIN (Shortliffe 1976) features a knowledge base containing about 600 rules and is one of the most well-known rule-based expert systems for medical diagnosis. Many researchers promoted the idea of graph-based knowledge representation aiming to assemble human knowledge. Furthermore, Resource Description Framework (E. Miller 2005) and Web Ontology Language (McGuinness 2004) were released and became the mainstay of Semantic Web.

In 2009, the concept of Linked Data was proposed to build the links between different datasets in the Semantic Web with each other and treat it as one global knowledge graph (Berners-Lee, Bizer and Heath 2009). Subsequently, various ontologies or knowledge bases were published, such as WordNet (G. Miller 1995), YAGO (Suchanek, Kasneci and Weikum 2007), DBpedia (Auer, et al. 2007) and Freebase (Bollacker, et al. 2008), to realize the idea of structured knowledge representation in the form of a graph. In 2012, Google proposed Knowledge Graph (Knowledge Vault) to utilize semantic knowledge in the application of web search, and the concept gained great popularity (Dong, et al. 2014). Google uses the knowledge graph to help identify and disambiguate entities in texts, utilize semantically structured summaries to enrich the search results, and further provide links to related entities in exploratory search (Dong, et al. 2014). Recently, many companies such as Microsoft, Amazon, and Pinterest have started investing massive resources to build knowledge graphs for their commercial applications (Färber 2019, Cui and Shrouty 2020, Dong, et al. 2020).

b. Knowledge Graphs for Medical Imaging Analysis

Machine learning techniques have recently been applied to all stages of radiotherapy, from diagnostic imaging, using image registration for risk
definition, to the automated planning and outcome assessment (Chan, Witztum and Valdes 2020, Shan, et al. 2020). To provide quality healthcare, such as custom medicine or treatment planning refinement, machine learning techniques are able to offer assistive insights. Bringing together artificial-intelligence-driven (AI-driven) radiotherapy and deep-learning-based medical imaging analysis is a promising direction. To provide precision radiotherapy, analyzing medical images and other modalities to derive representative features in a quantitative manner is vital. There tends to be significantly more information underlying in images and other data modalities that cannot be visually perceived in a straightforward way. However, sophisticated algorithms enable us to mine and use these underlying information to improve diagnosis and treatments. There are two main reasons why AI-based radiotherapy is expected to outperform conventional radiotherapy. First, many latent features that cannot be perceived by human readers can be utilized by analyzing radiomics in an automated manner. Moreover, we can mine and infuse priori in a data-driven and end-to-end manner, making AI-based radiotherapy more powerful (Shan, et al. 2020). The radiotherapeutic process can generate a large amount of data on anatomical, metabolic, etc. (Skripacak, et al. 2014). One major challenge is that it can be complicated to extract and present those data in a meaningful yet interpretable manner. Moreover, medical reports are oftentimes written in natural languages, where the report sensitivity and specificity, associated decisions need to be handled appropriately to provide better patient treatments than the current standards. One ideal solution is to synergize knowledge graphs to represent a generalization of prior knowledge from different modalities for treatment planning optimization or patient prognosis improvements (Shan, et al. 2020).

Building a knowledge graph utilizing a patient’s electronic medical records and reports can provide valuable information for disease reasoning and further treatment planning. NLP techniques can be very helpful for identifying and extracting knowledge from text inputs. When given the medical reports, we can use online services such as the Watson Natural Language Understanding platform or the Amazon Comprehend Medical to construct the graph rather than from scratch. Based on these cloud services and other systems, we can distill and query high-quality domain-specific rules, knowledge graphs from unstructured or semi-structured contents extracted from images and data such as medical conditions, medication details (dosage, strength, and frequency), and other data like doctors’ notes, clinical reports, and patient health records (Shan, et al. 2020).

With the efforts mentioned above, we will have treatment-related feature graphs and knowledge graphs from medical images and medical text data. These two types of graphs are in different domains: one from images and data biologically/clinically informative, and the other is in terms of professional languages directly interpreted. Therefore, we need to bridge these two domains via an across-domain graph transformation. To this end, we can use a graph-based encoder-decoder network, including graph convolution, graph pooling. The encoder will extract the information from a radiomic graph while the decoder will reconstruct a corresponding graph. The bottleneck between the encoder and the decoder will bridge the image and text domains. In the cases of dynamics changes with different reports, one can use a graph-based RNN to learn a dynamic graph mapping.
c. Data Sources and Search Strategies
Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we conducted a comprehensive search of English-language articles published between 2006 and 2021 from five databases. The databases include IEEE Xplore\(^1\), PubMed\(^2\), Arxiv\(^3\), Google Scholar\(^4\), the ACM Digital Library\(^5\). The search strategy is to iteratively search keywords for relevant articles and related citations. As shown in Table 1, the keywords used to retrieve literature included knowledge graph(s) and medical imaging, knowledge graph(s) and medical image(s), graph(s) and medical imaging, graph(s) and medical image(s).

<table>
<thead>
<tr>
<th>Database</th>
<th>Search Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Xplore, PubMed, Arxiv, Google Scholar, ACM Digital Library</td>
<td>knowledge graph medical imaging OR knowledge graph medical image OR knowledge graph medical image OR knowledge graph medical image OR knowledge graphs medical image OR knowledge graphs medical image OR knowledge graphs medical images OR graph medical imaging OR graphs medical imaging OR graph medical image OR graph medical images</td>
</tr>
</tbody>
</table>

Table 1. Search keywords to retrieve literature.

d. Article Selection
After the acquisition of potential articles, we conducted abstract and full text screening. Article exclusion criteria included the following: duplicates; article types such as conference abstract, review, editorial, erratum, letter, note, and comment; unavailable full text; and articles irrelevant to knowledge graph applications in medical imaging analysis. The inclusion criteria for the target publications were as follows: (1) knowledge graphs were used, (2) the aim was to solve medical imaging analysis problems. Two reviewers used the eligibility criteria to screen the articles. During the screening process, all conflicting opinions among reviewers were discussed until we reached a consensus.

e. Data Synthesis and Analysis
From the selected articles, our data synthesis was motivated by an approach to gain insight into how knowledge graphs were applied to different medical imaging analysis tasks and review how knowledge graphs contributed to the medical imaging analysis tasks. We began by examining the general characteristics of the included studies, such as the publication year trend, publication country and the focus medical imaging analysis topic. Furthermore, we investigated the datasets utilized by the included studies, to provide readers with an insight into the available data sources that can be used for different medical imaging analysis topics.

\(^1\) https://ieeexplore.ieee.org/Xplore/home.jsp
\(^2\) https://pubmed.ncbi.nlm.nih.gov/
\(^3\) https://arxiv.org/
\(^4\) https://scholar.google.com/
\(^5\) https://dl.acm.org/
3. Results

a. Identification of Included Studies

We retrieved 780 articles from five databases, of which 728 articles were found to be unique. The titles and abstracts of articles were screened for article filtering and selection. Articles were sorted based on the relevance to applying knowledge graphs to medical imaging analysis, where 609 articles were excluded due to low relevance. We excluded several article types, including the conference abstract, review, editorial, erratum, letter, note, and comment, resulting in excluding 56 articles. We further excluded 38 articles without full text. 25 articles remained for subsequent full-text reviews. During the full-text screening, 4 articles were excluded due to irrelevance to knowledge graph applications in medical imaging analysis. After this full-text screening process, 21 articles were selected to be included in this scoping review. The article selection flowchart is shown in Figure 2.

![Figure 2. The flowchart of the article selection process.](image)

b. Statistical Characteristics of the Included Articles

All the 21 articles included in this work are published from 2006 to 2021, with a noticeable increment in the number of papers published per year (Figure 3). The included publications are across nine countries, with most contributions coming from China (48%) (Figure 4). Among all the included articles, the most common application of knowledge graphs in medical imaging is disease classification (56.5%), followed by disease localization and segmentation (17.4%), report generation (17.4%), image retrieval (8.7%) (Figure 5).
Figure 3. Year trend of reviewed articles

Figure 4. Publication country distributions
c. Disease Classification

One of the most common computer vision tasks is to classify images into appropriate categories (Obaid, Zeebaree and Ahmed 2020). Disease classification is especially of vital importance in medical imaging to assist diagnosis (Xie, et al. 2021). The types and sizes of image datasets are increasing dramatically. Hence often, we need to classify images from unseen classes into the correct categories based on the relationships between the seen and unseen classes. Our world contains millions of visual concepts. Due to its complex and dynamic characteristics, it is impossible to build a large dataset for every concept to ameliorate various computer vision tasks. Prior knowledge is the key to building semantic relationships between classes, which can be of great help, especially when we have limited training data. Knowledge graphs contain rich knowledge, modeling the relationships among classes or concepts. Incorporating disease classification in medical imaging with knowledge graphs has been explored by researchers and has shown promising results (Xie, et al. 2021). Table 1 lists the overview of datasets used by the included articles related to disease classification in this work. The most commonly used datasets are IU X-Ray (Demner-Fushman, et al. 2015), NIH Chest X-Ray/ChestX-Ray 14 (Wang, et al. 2017) and CheXpert (Irvin, et al. 2019), and they all include medical images, associated medical reports, and disease labels.

There are five articles included in this review that explored binary disease classification incorporated with knowledge graphs. Xie et al. constructed a knowledge-based collaborative sub-model for the task of nodule classification. They proposed to fine-tune three pre-trained ResNet-50 networks using three types of image patches. The three pre-trained ResNet-50 networks were respectively used to characterize the nodules’ overall appearance, voxel, and shape heterogeneity (Xie, et al. 2019). In this way, a knowledge-based collaborative model was introduced to incorporate the multi-view information for separating the benign nodules from the malignant ones using very limited data. Yu et al. aimed to facilitate the process of pneumonia diagnosis (Yu, Wang and Zhang 2021). A graph-based feature reconstruction module was employed that takes the produced image features from a trained convolutional neural network (CNN) as input. The resulting combined features will be fed to a one-layer graph neural network (GNN) to classify chest X-ray images into two classes: normal and pneumonia. According to Chen et al., most existing work manually built a population graph for structural...
information aggregation where the relationship between nodes was represented by the graph adjacency matrix (Chen, et al. 2021). Chen et al. automatically constructed the population graph and further utilized the fusion of multimodal information, which improved the diagnostic accuracy for Autism Spectrum Disorder and breast cancer. Specifically, they proposed an encoder that can select the appropriate phenotypic measures in an automatic manner in terms of their spatial distribution. They further computed the edge weights between nodes utilizing a mechanism which is aware of the text similarity. Liu et al. claimed to outperform previous work on the Mammogram mass classification task (Liu, et al. 2021). To model the intrinsic geometric and semantic relations of ipsilateral views, they proposed a Bipartite Graph Convolutional Network. The asymmetric visual information of bilateral views was widely adopted in clinical practice to assist the diagnosis process of lesions. To model the structural similarities of bilateral views, an Inception Graph Convolutional Network was further proposed. The representations learned from the constructed graphs were capable of multi-view reasoning, since there was a systematical propagation of the multi-view information through nodes (Liu, et al. 2021). Fu et al. pointed out that most existing methods only focus on the image modality while ignoring or not fully leveraging information from other modalities (Fu, et al. 2021). They proposed to exploit the inter-category relationships in the 7-point visual category checklist (7PC) for Melanoma diagnosis. Specifically, they proposed to use a graph-based relational module to leverage inter-categorical and inter-modal relations. The dermoscopy visual structure details were further prioritized by representing the features in a graph network (Fu, et al. 2021). Another category embedding learning module was also employed to capture the specialized representations for each category and support the graph-based relational module.

Six articles explored multi-label classifications in medical imaging. Zhang et al. constructed a disease finding knowledge graph and utilized it to tackle the disease classification task (Zhang, et al. 2020). Incorporating a knowledge graph with the disease classification task allowed for dedicated feature learning for each disease finding (Zhang, et al. 2020). Similarly, Hou et al. modeled the correlations among disease labels by employing the graph convolutional network (GCN). They further pre-trained the disease label embeddings on the radiology reports. A transformer-based encoder was employed to fuse the semantic features along with the encoded image features to initialize the graph features (Hou, Zhao and Hu 2021). To have a better graph representation capability, they mined additional medical terms from radiology reports and these newly mined terms were added to the graph serving as auxiliary nodes without changing the actual output space size.

However, Zhou et al. pointed out that the developing a robust automated diagnosis system could be hindered by the fact that the lesions can have inconsistent appearances and high complexities in chest X-rays (Zhou, et al. 2021). They proposed one promising approach to address this issue, which is to attend to the abnormal regions and exploit relevant prior information (Zhou, et al. 2021). To have a better thoracic disease identification performance, especially for those whose lesions rarely appear on both sides symmetrically, one contrastive network was proposed to learn the intra-attentive abnormal features between the left and right lung. They further utilized an inter-contrastive attention module to acquire the abnormal attention map. Specifically, they compared the query scan with multiple
anchor scans where no lesions were present. After the features were weighted using the intra- and inter-contrastive attention scores, the radiology graph was further constructed for graph reasoning in a dual-weighted manner in addition to the basic visual-spatial convolution (Zhou, et al. 2021). Following the same direction, Agu et al. noted that most existing methods solely used the chest X-ray images for classification, but they failed to utilize the underlying anatomical information that can be really helpful (Agu, et al. 2021). They utilized a GCN which enables their model to learn the anatomical region relationships and label dependencies in the chest X-ray images. They further created an anatomical region adjacency matrix based on the correlation of the labels across different regions. Combining this with a detection module, they proposed a multi-label chest X-ray classifier that can classify image findings and localize them to their anatomical regions (Agu, et al. 2021).

According to Sekuboyina et al., learning to map images to binary labels made it a challenging task to take advantage of auxiliary information (e.g., annotation uncertainty, label dependencies, etc.) (Sekuboyina, et al. 2021). A multimodal knowledge graph was constructed using chest X-ray images and their labels. They approached the task of multi-label disease classification in a link prediction manner. They claimed that they added additional nodes and relations to incorporate auxiliary information into the graph (Agu, et al. 2021). Similarly, Chen et al. noted that given the fact that graph data featured high complexity, most previous works failed to fully use such valuable graph-structured information, but solely focused on learning to classify the input into binary labels (Chen, et al. 2020). As a result, they proposed to explicitly explore the graph structure information, such as the pathology dependencies, for the classification task. They introduced the pathology word embeddings and multi-layer graph information propagation to generalize the relationships between pathologies into a set of classifier scores. The flexible integration into the image feature embedding module and the adaptive recalibration of multi-label outputs with these scores were made possible during the training process (Chen, et al. 2020).

Since 2020, knowledge graphs have also been explored in COVID-19 related research and shown noticeable performance improvements. Zheng et al. pointed out that current deep learning methods suffered from data adequacy issues and that multimodal information should be considered together to make accurate inferences (Zheng, et al. 2021). To solve this, they proposed a multimodal graph attention embedding mechanism to assist diagnosing COVID-19. Their method learned the relational embeddings in a constituted knowledge graph and at the same time, improved the classifier through the medical knowledge attention mechanism (Zheng, et al. 2021). According to Mudiyanselage et al., the poor performance for unseen data in COVID-19 classification can result from the limited correlation between the pre-trained model and a specific imaging domain (e.g., X-ray), and the possibility of overfitting (Mudiyanselage, et al. 2021). They proposed that the relational knowledge between data instances can be exploited through graph representations, and further utilized through graph convolutions (Mudiyanselage, et al. 2021).

To summarize, the aforementioned work utilized different types of knowledge graphs, and they incorporated the knowledge graphs with the disease classification

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>Task</th>
<th>Dataset</th>
<th>Dataset Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Xie, et al. 2019)</td>
<td>2019</td>
<td>Binary Classification</td>
<td>LIDC-IDRI (Armato III, et al. 2011)</td>
<td>1,018 chest CT scans with lung nodules present. CT scans were obtained from seven institutions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CheXpert (Irvin, et al. 2019)</td>
<td>224,316 chest radiographs of 65,240 patients, with 14 common disease labels.</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Task Type</td>
<td>Dataset Description</td>
<td>2017)</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Hou, Zhao and Hu 2021)</td>
<td>2021</td>
<td>Multi-label Classification</td>
<td>IU X-Ray (Demner-Fushman, et al. 2015) 3,955 radiology reports, 7470 chest X-ray images.</td>
<td>3,955 radiology reports, 7470 chest X-ray images.</td>
</tr>
<tr>
<td>(Zhen, et al. 2021)</td>
<td>2021</td>
<td>Binary Classification</td>
<td>COVID-19 multi-modal dataset 1,393 doctor–patient dialogues and 3706 images (347 X-</td>
<td>1,393 doctor–patient dialogues and 3706 images (347 X-</td>
</tr>
</tbody>
</table>
2021) & Multi-label Classification & 7PC (Kawahara, et al. 2019) & 1,011 lesion cases, and report comprehensive results. \\

| (Fu, et al. 2021) | 2021 | Multi-label Classification | 7PC (Kawahara, et al. 2019) | 1,011 lesion cases, and report comprehensive results. |

Table 2. Overview of datasets used in the disease classification articles.

d. Disease Localization and Segmentation

In medical imaging, disease localization and segmentation are useful for clinical diagnosis, disease assessment, and treatment planning (Sharma and Aggarwal 2010). Previous supervised methods suffered from the lack of finely annotated data, and weakly-supervised methods often generated inaccurate or incomplete regions (Qi, et al. 2021). When taking into account the anatomical region relationships and the correlations between images, complementary information can be obtained to improve the disease localization accuracy. This also aligns with the clinical practice in the medical domain: usually to train a radiologist to analyze X-ray images, they read many X-ray images, compare the differences between different images and even the differences between different regions of the same image. (Qi, et al. 2021). Infusing knowledge graphs into the system offers the potential for more accurate localizations and segmentations. Table 2 lists the overview of included articles and the datasets that are used. The most commonly used datasets for this task are CheXpert (Irvin, et al. 2019) and NIH Chest X-Ray/ChestX-Ray 14 (Wang, et al. 2017). Both datasets include medical images and associated medical reports, and disease labels.

Peng et al. identified the fissure region of interest using lung anatomy prior knowledge, then isolated the plate-like structures from clutters utilizing an oriented derivative of stick filter for lobar fissure verification. Finally, to segment lung lobes, they completed the incomplete fissure surface employing a surface fitting model (Peng, et al. 2021). Qi et al. noted that one reason for incomplete localization regions was neglecting the anatomical region relationships within each image and the inter-image relationships (Qi, et al. 2021). Hence, they proposed to model the inter-image relationships by comparing multiple images in an inter-image graph and to model the intra-image relationships by comparing different regions in an intra-image graph. These cross-image and cross-region relationships were used as the contextual and compensating knowledge and were incorporated for disease localizations. Through ablation study, they showed that the model employing the intra-image and inter-image prior knowledge outperformed the localization accuracy of the baseline model by 0.08, 0.11, and 0.1 when the intersection over union (IoU) threshold was 0.3, 0.5, and 0.7. Zhao et al. also noted that most weakly-supervised disease localization methods failed to consider the chest X-ray image characteristics (e.g., the highly structural attributes) (Zhao, Qi and Li 2021). They used a very similar method to Qi et al. (Qi, et al. 2021), which
integrated the intra-image anatomical structural knowledge and inter-image knowledge into one unified framework.

In summary, the aforementioned articles incorporated knowledge graphs with disease localization and segmentation using two approaches: (1) embed visual features to pre-constructed prior knowledge region graph (Qi, et al. 2021, Peng, et al. 2021, Zhao, Qi and Li 2021); (2) use images as graph nodes (Qi, et al. 2021, Zhao, Qi and Li 2021). These articles identified the importance of prior knowledge, proposed to infuse prior knowledge into the disease localization and segmentation tasks in the form of knowledge graphs, and showed that the prior knowledge did bring drastic performance improvements.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>Method</th>
<th>Dataset</th>
<th>Dataset Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Peng, et al. 2021)</td>
<td>2021</td>
<td>Fissure verification, Surface fitting</td>
<td>LOBe and Lung Analysis 2011 (LOLA11) (Lassen, et al. 2011)</td>
<td>A dataset of chest CT scans with varying abnormalities for which reference standards of lung and lobe segmentations have been established.</td>
</tr>
</tbody>
</table>

Table 3. Overview of datasets used in the disease localization and segmentation articles.

e. Report Generation

Natural language captioning aims to summarize visual information in one sentence or generate one topic-related paragraph (Li, et al. 2019). Medical report generation translates the medical images to human-readable medical reports, which requires an increased capability to cover accurate abnormal terminologies, understand the medical domain knowledge and describe the findings at a semantic-coherent and fine-grained level that should satisfy both medical commonsense and logic (Li, et al. 2020). Outstanding challenges associated with automatic medical report generation lie in successfully detecting visual groundings and incorporating medical domain knowledge. To write a medical image report, radiologists will first check a patient’s images, carefully inspect the abnormal regions to identify the findings, then describe the abnormal findings in detail based on prior medical
experiences and medical knowledge. Only employing the global images as input and training the language model with the dataset’s corpora alone cannot provide the underlying prior knowledge vital for accurate reporting. Several works infused knowledge graphs into report generation and showed the performance gain on the quality of generated reports. The datasets used by the included articles related to report generation are listed in Table 3, and the most used dataset is IU X-Ray (Demner-Fushman, et al. 2015).

Zhang et al. constructed a graph embedding module on multiple disease findings as the prior knowledge to assist report generation. The incorporation of a knowledge graph allowed for dedicated feature learning for each disease finding and the relationship modeling (Zhang, et al. 2020). The knowledge graph module improved the baseline SentSAT model (Yuan, et al. 2019) on nearly all report generation evaluation metrics, especially 0.036 improvements on the CIDEr metric. Li et al. noted the significant challenges towards bridging visual and linguistic modalities; hence they proposed to encode visual features as an abnormality knowledge graph, which incorporated the visual features with prior medical knowledge, and was then used to guide the report template retrieval-paraphrase process or used for disease classification (Li, et al. 2019). Similarly, Liu et al. noted that visual and textual data biases remained a challenge for data-driven report generation systems, so in addition to the disease-tag attended visual features and the disease-attended textual features, they proposed to explore the prior knowledge from a pre-defined medical knowledge graph guided by attended-image features, and further adaptively distill the knowledge for report generation (Liu, et al. 2021). Through ablation study, they show that removing the prior knowledge graph module from the proposed model will cause a significant drop in all evaluation metrics, especially 0.66 drop on CIDEr, 0.34 drop on BLEU-1. Li et al. pointed out that previous methods suffered from the deviation that taught models to generate inessential sentences regularly. Therefore, inspired by Generative Pre-Training, they proposed to guide medical knowledge transfer and learning through a medical graph encoder, by integrating internal visual feature fusion and external medical linguistic information (Li, et al. 2020).

To summarize, the aforementioned articles incorporated knowledge graphs with report generation using one common approach: embed visual features to pre-constructed prior knowledge abnormality graph (Li, et al. 2019, Zhang, et al. 2020, Liu, et al. 2021, Li, et al. 2020). Knowledge graphs utilized in the included articles modeled the relationships of disease findings and bridged multiple modalities by embedding visual representations. Through ablation studies, all included articles demonstrated the effectiveness of incorporating knowledge graphs with report generation.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>Method</th>
<th>Dataset</th>
<th>Dataset Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Li, et al. 2019)</td>
<td>2019</td>
<td>Graph Transformer</td>
<td>CX-CHR dataset</td>
<td>Private dataset. 35,609 patients, 45,598 images and corresponding reports.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IU X-Ray</td>
<td>3,955 radiology reports, 7,470 chest X-ray images.</td>
</tr>
</tbody>
</table>
Table 4. Overview of datasets used in the report generation articles.

<table>
<thead>
<tr>
<th>Article</th>
<th>Year</th>
<th>Method</th>
<th>Data Description</th>
</tr>
</thead>
</table>

f. Image Retrieval

Automated image retrieval systems show enormous potential in medical applications (Qayyum, et al. 2017). It can be beneficial for the clinical decision-making process to extract similar images that share common aspects (i.e., modality, anatomic region, disease, etc.). This allows for extracting similar images with similar diagnoses, and also allows for finding similar images but with different diagnoses. In medical domain, doctors can adopt image retrieval systems to retrieve images with known pathologies that are similar to the anchor image, and further to assist the diagnosis process. In addition to doctors, medical researchers, lecturers and even students can extract relevant images using visual retrieval methods for their teaching and research. Primitive features (i.e., color, texture, etc.) are still the dominant features used by most image retrieval systems for image presentation purpose (Hwang, Lee and Choi 2012). No medical knowledge was used in this process; hence there exists the domain gap if we want to apply the systems to medical domain. This loss of information can be reduced by incorporating prior knowledge and other sources of knowledge (Putzu, Piras and Giacinto 2020). Table 4 lists the overview of datasets used in the image retrieval articles included in this review.

Lacoste et al. presented their medical image retrieval method incorporating medical prior through a fusion framework (Lacoste, et al. 2006). The text knowledge infused was from the Unified Medical Language System (UMLS) sources. They learned semantic features from examples and further to derive the visual knowledge. UMLS concepts enabled the communication between visual and
textual information, allowing a higher level systematic medical data standardization. Racoceanu et al. employed similar methods by using global indexing to access image modality and local indexing to access local semantic features to fuse the textual and visual knowledge into image retrieval (Racoceanu, et al. 2006).

In summary, these two included articles incorporated knowledge graphs with image retrieval using one common approach: represent images and texts in UMLS graphs (Lacoste, et al. 2006, Racoceanu, et al. 2006). The introduced knowledge graphs facilitated the communication between multiple modalities and benefitted the image retrieval task.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>Method</th>
<th>Dataset</th>
<th>Dataset Info</th>
</tr>
</thead>
</table>

Table 5. Overview of datasets used in the image retrieval articles.

4. Discussion

After the article selection phase, there were 21 articles selected and included in this survey. This relatively small number of articles related to knowledge graph applications in medical imaging analysis may suggest the cross-disciplinary gaps and lack of collaborations. In this section, we identified the limitations of the included articles and suggested the potential future directions.

a. Disease Classification

Most articles share some common limitations in this review. The datasets are still too small in size (on average, 129,788 images, ranging from 450-384,580) to provide results that are more convincing. The construction of graphs and the reconstruction of features are important aspects of most of the works. However, the graphs were constructed on a given dataset, making the extension to other domains inconvenient and challenging. For example, some graphs were designed as components of the proposed model for diagnosing chest diseases, which would not work for a brain tumor diagnosis task. If researchers want to transfer the method to handle a problem in another domain, building a new graph using a similar approach would be necessary. Also, an encoding component pre-trained towards a specific task (multi-label classification) could result in representations that do not generalize well across tasks. Furthermore, global classifications can be unreliable even when the predicted label is correct. The classifier might predict the correct label but for a wrong reason at an irrelevant spot.

For the future direction, one can consider a semi-supervised learning framework to reduce the demand for annotated data. Also, we can think of considering more sophisticated graph structures, which model more detailed disease relationships in the future. Other approaches can be explored to better incorporate visual and semantic features. It is worth exploring a task-agnostic representation learning framework for better generalizability. Another future research direction is t
combine the encoding and embedding modules resulting in a fully end-to-end formulation.

b. **Disease Localization and Segmentation**

The included articles regarding disease localization and segmentation in this review did not consider the label uncertainty, which is worth exploring to improve the performance. The reported results showed that small targets (e.g., Atelectasis, Effusion, Nodule) were more challenging to localize due to their relatively smaller size. Algorithms applicable to the localization of small targets (e.g., Atelectasis, Effusion, Nodule) are worth exploring.

Existing work like Peng, et al. heavily depends on other tasks like airway segmentation. However, the segmentation is an arduous task as it is highly sensitive to the image quality (Peng, et al. 2021). They segmented pulmonary fissures using lung anatomy knowledge, which is time-consuming. There is a possibility that parts of fissures can be undetected due to the poorly segmented airways. To conclude, a more fine-grained fissure detection and lung lobe segmentation method will be ideal to pursue.

c. **Report Generation**

Currently, most work applying knowledge graphs into report generation uses visual features for graph feature initializations. It is worth exploring different fusion methods to combine knowledge graphs with multimodal features. We can also encode and decode general knowledge for report generation tasks by exploring a general captioning framework guided by the auxiliary signals.

d. **Image Retrieval**

In this review, two included articles regarding image retrieval used global and local indexing to infuse additional visual, textual, and knowledge graph features into image retrieval. One can further use appropriate clustering methods to take advantage of other fusion schemas. It is also worth exploring the visual filtering based on the local information from the semantic local indexing module to distill visual features for better performance.

5. **Conclusions**

This review discussed the current work on knowledge graph applications in medical imaging analysis and identified the limitations and future directions. We looked at the proven success of applying knowledge graphs into four medical imaging tasks: disease classification, disease localization and segmentation, report generation, and image retrieval. We identified the limitations due to limited annotated data for some supervised tasks and weak generalizability. We also identified potential future directions, for example, employing semi-supervised framework, exploring different fusion methods, exploring task-agnostic models that may improve the opportunities for better performance.

**Ethical Approval:** N/A

**Data Availability:** N/A

**Funding:** This work was supported by the National Library of Medicine under Award No. 4R00LM013001 and was also supported by Amazon Machine Learning Research Award 2020.

**Author contributions:** Song Wang and Mingquan Lin designed the study, performed a literature search, systematic analysis, and drafted the manuscript. Tirthankar Ghosal, Ying Ding, and Yifan
Peng provided intellectual input to the study and edited the manuscript. Song Wang, Mingquan Lin contributed equally to this work.

**Conflict of Interest:** The authors declare no competing interests.

**Acknowledgments:** This work was supported by the National Library of Medicine under Award No. 4R00LM013001 and was also supported by Amazon Machine Learning Research Award 2020.

**Supplementary Materials:** N/A

---

**References**


Dong, Xin Luna, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, et al. 2020. "AutoKnow: self-driving knowledge collection for products of thousands of types."
Dong, Xin, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. "Knowledge vault: a web-scale approach to probabilistic knowledge fusion." The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA.


Fei, Hao, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. 2021. "Enriching contextualized language model from knowledge graph for biomedical information extraction." Briefings in Bioinformatics.

Fu, Xiaohang, Lei Bi, Ashnil Kumar, Michael Fulham, and Jinman Kim. 2021. "Graph-Based Intercategory and Intermodality Network for Multilabel Classification and Melanoma Diagnosis of Skin Lesions in Dermoscopy and Clinical Images."


Martino, Adriana Di, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco Xavier Castellanos, Kaat Alaerts, Jason Anderson, Michal Assaf, Susan Bookheimer, and Mirella Dapretto. 2014. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism."


Qi, Baolian, Gangming Zhao, Xin Wei, Chaowei Fang, Chengwei Pan, Jinpeng Li, Huiguang He, and Licheng Jiao. 2021. "GRE: Graph-Embedded Regularized Network for Weakly-Supervised Disease Localization in X-ray images."


Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *MICCAI.*


Vrandečić, Denny, and Markus Krötzsch. 2015. "Wikidata: a free collaborative knowledgebase." *Communications of the ACM.*


Yang, Xingyi, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, and Pengtao Xie. 2020. "COVID-CT- Dataset: A CT Scan Dataset about COVID-19."


Zhao, Gangming, Baolian Qi, and Jinping Li. 2021. "Cross Chest Graph for Disease Diagnosis with Structural Relational Reasoning."
