Knowledge Enhanced Deep Learning Model for Radiology Text Generation

Kaveri Kale[1], **Pushpak Bhattacharyya**[1],
Aditya Shetty[2], **Milind Gune**[3],
Kush Shrivastava[4], **Rustom Lawyer**[4] and **Spriha Biswas**[4]

[1]Department of Computer Science and Engineering, IIT Bombay, India
[2]Consultant Radiologist, Breach Candy Hospital, India
[3]Consultant Radiologist, Thane, India
[4]Augnito India Private Limited

{kaverikale,pb}@cse.iitb.ac.in,
adityashetty01@gmail.com, dgune@rediffmail.com, {kush.shrivastava, rustom, spriha}@augnito.ai

Abstract

Manual radiology report generation is a time-consuming task. First, radiologists prepare brief notes while carefully examining the imaging report. Then, radiologists or their secretaries create a full-text report that describes the findings by referring to the notes. Automatic radiology report generation is the primary objective of this research. The central part of automatic radiology report generation is generating the finding section (main body of the report) from the radiologists' notes. In this research, we suggest a knowledge graph (KG) enhanced radiology text generator that can provide additional domain-specific information. Our approach uses a KG-BART model to generate a description of clinical findings (referred to as **pathological description**) from radiologists' brief notes. We have constructed a parallel dataset of radiologists' notes and corresponding pathological descriptions to train the KG-BART model. Our findings demonstrate that, compared to the BART-large and T5-large models, the BLEU-2 score of the pathological descriptions generated by our approach is raised by 4% and 9%, and the ROUGE-L score by 2% and 2%, respectively. Our analysis shows that the KG-BART model for radiology text generation outperforms the T5-large model. Furthermore, we apply our proposed radiology text generator for whole radiology report generation.

1 Introduction

Due to the meager ratio of radiologists to patients, radiologists are in high demand. The ratios in the US, China and India are 1:10,000, 1:14,772, and 1:100,000 respectively (Arora, 2014). It leads to a large influx of patients, which keeps radiologists extremely busy and under stress. To boost the effectiveness and productivity of radiologists, several hospitals and diagnostic facilities have established radiology information systems (RIS) and picture archiving and communications systems (PACS) (Honeyman, 1999). Despite this, the current workflow causes a delay in the turnaround time for reports, report inaccuracies, and burnout. Our conversations with radiologists have revealed that many radiologists wish to eliminate the tiresome report-generating process and concentrate on the diagnosis.

The main task in automatic radiology report generation is generating pathological descriptions from radiologists' notes. In Natural Language Processing (NLP), we can look at this as a text generation task. Various neural encoder-decoder models have been proposed to accomplish the text generation goal by learning to map input text to output text. However, the input text alone often provides limited knowledge to generate the desired output. Hence, researchers have considered incorporating external knowledge from domain-specific KGs along with internal knowledge embedded in the input text (Yu et al., 2022). KG-BART (Liu et al., 2021) model incorporates the domain-specific KG in the deep learning model. In our work, we incorporate ultrasound radiology KGs (Kale et al., 2022) in the KG-BART model to generate radiology text from radiologists' input notes. We construct a radiology domain dataset to train the KG-BART model. We obtain grounded KG for input sentences. KG grounding finds the most relevant entities and relations from radiology KG to guide the KG-BART model to better understand the relationships among concepts. It considers the inter-concept relation and significant neighbor entities to generate a more natural and plausible output. Two high-profile radiologists are associated with this research who help us to get domain insights and to create a dataset.

Our contributions in this paper are as follows:

- Parallel dataset of radiologists' notes and corresponding pathological descriptions.

Proceedings of the 19th International Conference on Natural Language Processing (ICON), pages 32 - 42
December 15-18, 2022

- Our work shows that the KG-BART is strong choice for radiology text generation than other state-of-the-art models like T5-base/large, BART-base/large.
- Our KG grounding approach reduces noise (irrelevant neighbor entities) and obtains the most relevant neighbor entities.

2 Background Concept and Terminology

Traditionally, radiologists either dictate on a voice recorder or write the diagnosis notes (referred to as **radiologist's notes**) on paper. Their secretaries are then given access to the notes. Next, the secretary access a normal report template, which is a scan-specific normal template (referred to as **normal report template**) that corresponds to all normal findings, and creates a preliminary report by altering it in accordance with the measurements and findings that the radiologist reported in a more detailed form (**pathological description**). The radiologist receives the preliminary report once again. The radiologist then reads the report, makes any necessary corrections, and then signs off. Finally, the report is provided to a doctor or patient for potential follow-up care. Table 1 shows the examples of radiologists' notes and corresponding pathological descriptions. The average number of words in radiologists' notes is 15, and the average number of words in pathological descriptions is 26.

Radiologists' Notes	**Pathological Description**
Normal uterus 1 x 5 x 3.4 mm with hypoechoic fibroid 2.3 x 5.6 mm in fundic body region.	Uterus is anteverted and normal in size 1 x 5 x 3.4 mm. Myometrial reflectivity is inhomogeneous and shows a hyperechoic fibroid in fundic body region measuring 2.3 x 5.6 mm.
Liver shows generalized fatty infiltration severe hepatomegaly noted.	Liver severely enlarged and it reveals generalized fatty infiltration.
Cirrhosis with portal hypertension 6 cm.	Liver is small and shrunken and coarse echotexture margin are nodular. portal vein is mildly dilated, it measures 6 cm.

Table 1: Examples of radiologists' notes and corresponding pathological descriptions.

In order to create an accurate diagnostic description from brief notes, domain-specific knowledge will be helpful, given that the knowledge graph can supply relationship information to strengthen the capacity for reasoning and provide adjunct entities to the concept. In our work we use radiology (ultrasound) KGs constructed by Kale et al. (2022). These KGs are constructed for each organ separately. Since all KGs are hierarchical, and the root of each KG is an organ name (e.g., *liver, gallbladder, pancreas*, *etc.*), we have integrated all these KGs into a single KG (reffered as **ultrasound KG**) by adding *upper abdomen* as root entity. First we extract the grounded KGs for each input concept set from ultrasound KG and then we incorporate grounded KGs in KG-BART model to generate pathological descriptions. The **KG grounding** is the process of extracting the subgraphs (referred to as grounded KGs) from domain-specific KG (in our case ultrasound KG). A grounded KG is a subgraph from the KG whose nodes are concepts in the input plus additional nodes. While doing KG grounding we construct two graphs, i) Input-concept graph and ii) Concept-expansion graph. The expansion is due to the KG, supplying related concepts closely related to those in the input. **Input-concept Graph:** It consists of (a) nodes in the KG matching with input concepts and (b) nodes that are along the paths to the root node of the KG, containing these nodes. **Concept-expansion Graph:** It is the input-concept graph plus the relevant children of the nodes in the input-concept graph.

For input-concept graph and concept-expansion graph, we encode the entity representations and their dependency relations using Knowledge Graph Embeddings (KGE) (Choudhary et al., 2021). KGE represents the entities and relations in lower-dimensional vectors that can be efficient for computations.

3 Related Work

With prior knowledge of chest findings, Zhang et al. (2020) created a graph model that could be used in deep learning models. Disease findings are represented in this network as nodes, and related findings are connected closely between them so that they might impact one another during the propagation and aggregation of the graph. To learn specific attributes for each node in the graph, they incorporate this graph into the deep neural network. Features extracted from KG are used for multilabel classification. To improve pre-training language understanding, Zhang et al. (2019) integrates KG instructive items that are contextually aligned. KEPLER uses a pre-trained language understanding model to encode textual descriptions of entities before integrating the goals of knowledge embedding and language modeling (Wang et al., 2021). By including triples from the KG as supplemental words, K-BERT infuses domain information into the models (Liu et al., 2020). In light of these studies, we contend that additional knowledge data can significantly improve the performance of pre-training models used for text generation tasks.

4 Dataset Construction

We fetch impressions and corresponding pathological descriptions from the radiology text report corpus to construct a parallel dataset. The radiology text report corpus contains anonymized radiology ultrasound reports that are provided by a company collaborating with us, with due consent of the physicians. We have approximately 10 lac radiology reports, out of which around 1 lac reports are of ultrasound. The radiology report contains the title, history, findings, and impression sections. Each section's content is well-structured despite not being uniform. As a result, we use heuristics like regular expressions and word overlap to identify different sections. Each section's content is tokenized and lower-cased. Impressions are very close to radiologists' notes, but it does not contain patient-specific information like measurement of findings, anatomical location, *etc.* To convert impressions into radiologists' notes, we manually edit impressions by adding patient-specific information like measurements and anatomical locations by referring to pathological descriptions. Examples of impressions, their corresponding pathological descriptions and radiologists' notes prepared using impressions are given in table 2.

4.1 Data Preprocessing

Reports contain free-text clinical narratives. Therefore it has many spelling mistakes and writing mistakes as well. We perform the following preprocessing tasks on a parallel dataset:

- In the corpus, there are a lot of extra spaces and unwanted punctuation marks found. We remove these unwanted characters from the corpus using regular expressions.
- We apply sentence segmentation to identify sentence boundaries between different sentences.
- We use SymSpell[1] library to correct the spellings by applying the unigram and bigram dictionaries. We create dictionaries from the corpus and correct them manually.

Once the data cleaning process done. Our domain experts verify the dataset manually and correct it if necessary. We try to create a dataset because the radiology dataset for ultrasound is not publicly available.

4.2 Concept Extraction

KG-BART model needs input and target dataset for training and validation. We give text input to the KG-BART model in the form of a concept set. **Concept set** is the set of radiological entities extracted from radiologist's notes. For example radiological concepts present in note, *Lesion found in right lobe of liver.* are *lesion, right lobe,* and *liver*. To extract the concepts from radiologists' notes, we use an entity extractor based on the method explained in the paper (Kale et al., 2022). Table 2 shows the examples of radiologists' notes and concept sets extracted from notes along with their corresponding pathological descriptions.

5 Method

Figure 1 shows the architecture of the text-generation model with input/output flow. The main components of our model are KG grounding, KG embeddings, text embedding, encoder, and decoder. This section explains all these components in detail.

5.1 Knowledge Graph Grounding

We extract the small subgraphs, input-concept, and concept-expansion graphs for each input sample from our ultrasound KG.

Algorithm to construct input-concept graph and concept-expansion graph is given below: i) For each input concept set in our dataset, we link input concepts with ultrasound KG entities using entity matcher. Entity matcher implemented using TheFuzz[2] library. ii) To find the appropriate path in the ultrasound KG. First, we find all possible candidate paths from matched entity to the root node. We find the most appropriate top five paths by using ranking based on precision and recall of entities in concept set and entities in all possible candidate paths. We consider all paths which include matched entity which is absent in the already selected top-ranking path. iii) Our algorithm constructs an input-concept

[1] https://symspellpy.readthedocs.io

[2] https://github.com/seatgeek/thefuzz

Impression	Radiologists' Notes	Concept Set	Pathological Description
Bulky retroverted uterus with fundal fibroid.	Bulky retroverted uterus with fundal fibroid 2.3 x 5.6 mm.	uterus, fibroid, bulky, fundal, retroverted, 2.3 x 5.6 mm	Uterus is retroverted and bulky in size. Myometrial reflectivity is inhomogeneous with an illdefined fundal fibroid measuring 2.3 x 5.6 mm noted.
Calculus cholecystitis with multiple large calculi.	Calculus cholecystitis with multiple large calculi within lumen of gallbladder, largest measuring 2.4 mm.	multiple, calculi, calculus, lumen, cholecystitis, enlarged, measuring, 2.4 mm	Gallbladder is distended reveals thick wall. Feature of note is presence of multiple large calculi seen within lumen of gallbladder; largest calculus measures 2.4 mm.
Acute pancreatitis.	Acute pancreatitis.	acute pancreatitis	Pancreas is bulky, reveals reduced reflectivity with increased reflectivity of peripancreatic fat.

Table 2: Samples from dataset constructed using radiology report corpus. The first column shows the impressions extracted from the radiology report, and the last column shows the pathological description corresponding to the impression fetched from the radiology report. The second column shows the radiologists' notes prepared by adding patient-specific information to the impression. The third column shows the concepts extracted from radiologists' notes. The final training dataset contains only concept set (as input) and pathological description (as target) columns.

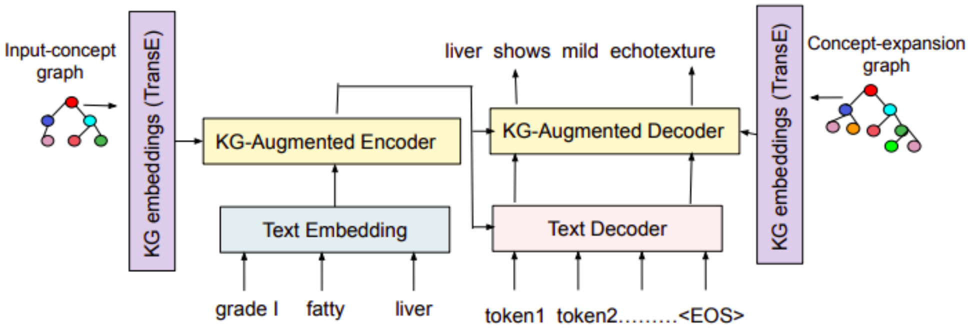

Figure 1: Our model architecture with input as radiologist's dictation and output as pathological description.

Total Samples	Train Samples	Test Samples	Validation Samples
6860	6000	430	430

Table 3: Statistics of the parallel dataset. Training dataset contains concept sets and corresponding pathological descriptions.

and concept-expansion graphs containing all paths that we have selected using a ranking algorithm and neighbor nodes which are the default properties of node present in path. Since ultrasound KG is hierarchical KG where if the node is **finding**, then its parent is the **anatomical location**, and its children are **properties** of findings. Hence, even if some information is missing in the input, we can get it from the input-concept or concept-expansion graphs.

Algorithm 1 gives the pseudocode to construct input-concept and concept-expansion graphs. Adding one-hop, two-hop, or n-hop neighbors of concept nodes adds irrelevant nodes in the expanded graph, which leads to noise. Our approach reduces the noise and obtains the most relevant neighbor nodes. Instead of passing these graphs (as it is for training), the model represents it in vector form. KG embedding module produces the vector representation for input-concept and concept-expansion graphs. Example of the input-concept and concept-expansion graphs is shown in figure 2.

5.2 KG Embeddings

The ultrasound KG is represented in low dimensional vector space using KGE. For simplicity and concreteness, in this work, we primarily consider TransE (Bordes et al., 2013) model due to their state-of-the-art performance. To implement the TransE model for KG embeddings, we use the open-source OpenKE[3] tool. Ultrasound KG contains 860 nodes and 1016 triples. Triples are divided into train and validation sets. The training triple set contains 900 triples, and the validation triple set contains 116 triples.

[3] https://github.com/thunlp/OpenKE

Algorithm 1: Construct input-concept and concept-expansion graphs.

Input : CS: Concept Set
G(V, E) : Knowledge Graph
Output : Input-concept and concept-expansion graph

Find all candidate paths in G(V, E) that includes the node with input concept
path-dict -> initialize
for *each path in possible candidate-paths* **do**
    Precision $= \frac{CS \cap \text{All entities in path}}{\text{No. of concepts in CS}}$
    Recall $= \frac{CS \cap \text{All entities in path}}{\text{No. of nodes in path}}$
    F-score $= \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$
    add path-dict -> (path:F-score)
end
Sort path-dict in descending order of F-score
Get top 5 paths
for *each path in top-5-paths* **do**
    if *len(set(CS) - set(path)) > 0* **then**
        Add all triplets from that path in input-concept graph triplet set
        Add all triplets from that path in concept-expansion graph triplet set
        for *each node in path* **do**
            Find all neighbors of *node* with default-property relation
            Add all triples of form (neighbor, DefaultPropertyOf, *node*) in concept-expansion graph triplet set
        end
    end
end
Save input-concept graph triples set (input-concept graph) in csv file.
Save concept-expansion graph triples set (concept-expansion graph) in csv file.

5.3 Text Embedding

The input embeddings are made of two separate embeddings, 1) Token embeddings and ii) Position embeddings. To get the final text embedding we add the vectors of token embeddings and position embeddings.

5.3.1 Token Embeddings

Tokens are nothing but a word or a part of a word. The textual encoder uses the vocabulary offered by large-cnn BART, and the token embedding is consistent with BART. Using a trainable lookup table, we transform each token in the input concept-set into an embedding vector.

In order to create these token embeddings, a method called BART tokenizer is used to tokenize the text. The encoder, decoder, and language modeling head (Press and Wolf, 2016) all share the embedding parameters. Due to the permutation-invariance of the attention layers, BART learns positional embeddings for absolute token positions and adds them to the token embeddings (Vaswani et al., 2017; Devlin et al., 2018).

5.3.2 Positional Embeddings

Position embeddings represents the position of the word within that sentence that is encoded into a vector. We must introduce some information about the relative or absolute location of the tokens in the sequence because our model lacks recurrence and convolution and hence cannot use the sequence's order. To do this, we augment the token embeddings at the base of the encoder and decoder stacks with positional embeddings. The text embeddings are the sum of the token embeddings and the positional embeddings.

5.4 Encoder

The encoder uses two modalities- text, and KG to condition the generation. According to Figure 1, the KG enhanced encoder layer sits above the text embedding layer and is intended to enhance the text representation by taking the KG structure into account. We use a graph attention layer to incorporate graph representations into the input encoding process. It uses explicit relations to help the model learn intra-concept relations more effectively. Formally, the grounded KG embedding, as well as the text embeddings, are combined by the KG-augmented encoder to update the text token representation. Our self-attention layer and fully-connected layer with residuals make up the stack of m transformer blocks that make up our bidirectional KG-augmented encoder.

5.5 Decoder

The decoder uses the text embedding module at the bottom layer to encode the text. Similar to encoder, decoder contains KG-augmented decoder layer. It incorporates a concept-expansion graph to get input concepts' missing information and context. The decoder of our model is also a multi-layer transformer. Our decoder is auto-regressive and unidirectional. We skip over a detailed explanation of these modules because our textual transformers are the same as those used in BART (Lewis et al., 2019) and (Vaswani et al., 2017).

6 Experimental Setup

The model input consists of the concept set and KG encoding for the input-concept and concept-expansion graphs. The output is the TARGET statement, i.e., pathological description. We use above

	BLEU Score				ROUGE Score								
					1-gram			3-gram			L-gram		
	1-gram	**2-gram**	**3-gram**	**4-gram**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
T5-base	0.81	0.74	0.68	0.63	0.87	0.88	0.87	0.76	0.77	0.76	0.84	0.85	0.84
T5-large	0.85	0.80	0.75	0.72	0.93	0.88	0.90	0.84	0.8	0.81	0.92	0.87	0.89
BART-base	0.86	0.82	0.78	0.75	0.93	0.90	0.91	0.84	0.82	0.83	0.91	0.89	0.90
BART-large	<u>0.89</u>	<u>0.85</u>	<u>0.84</u>	<u>0.81</u>	<u>0.93</u>	<u>0.92</u>	<u>0.92</u>	<u>0.86</u>	<u>0.86</u>	<u>0.86</u>	<u>0.93</u>	<u>0.92</u>	<u>0.92</u>
KG-BART	**0.93**	**0.89**	**0.86**	**0.83**	**0.96**	**0.96**	**0.95**	**0.89**	**0.89**	**0.88**	**0.94**	**0.94**	**0.93**

Table 4: BLEU and ROUGE score of generated pathological descriptions by T5-base/large, BART-base/large, and KG-BART models vs. gold standard pathological descriptions. The best results are in bold font, and the second best is underlined.

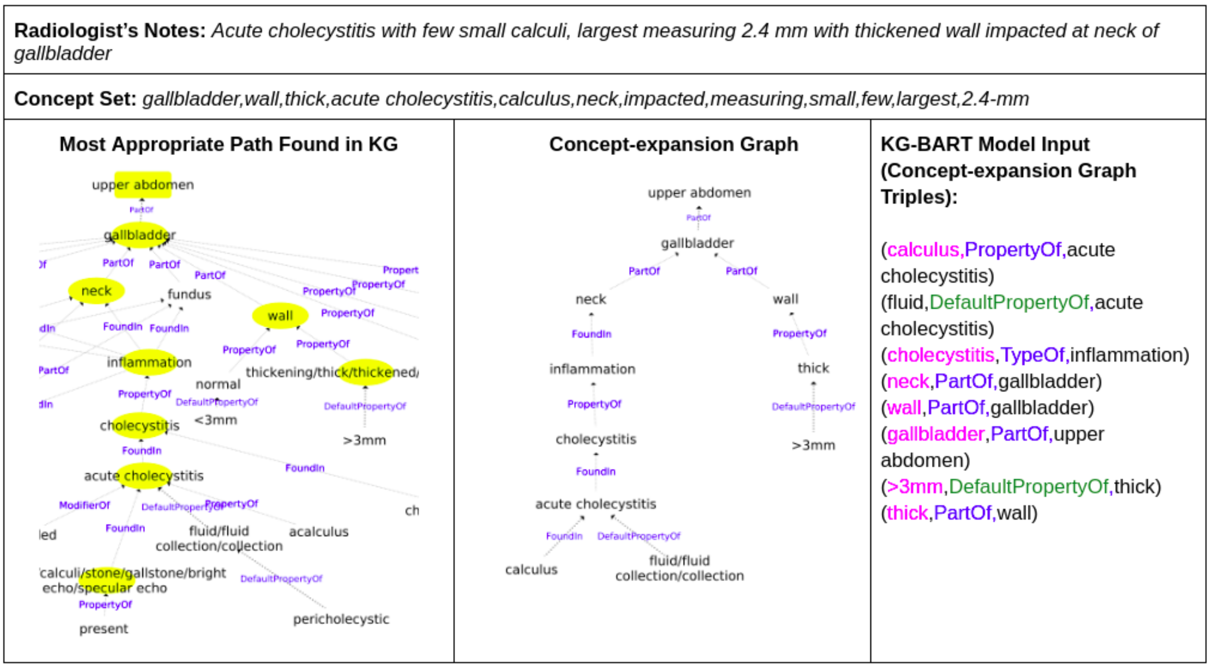

Figure 2: Left hand side graph is the snapshot of ultrasound KG that we are using for training. Nodes highlighted in yellow shows the concepts from the concept set that matches with the KG entities. Right hand side graph is the concept-expansion graph constructed for input concept set.

mentioned constructed dataset to train our model. Table 3 shows the statistics of constructed dataset.

6.1 Retraining Setup

We have implemented our own algorithm for KG-grounding task. We use pre-trained KG-BART[4] model which was trained for commonsense reasoning on ConceptNet KG and commonsense dataset. We fine tune this model on radiology text dataset that we have constructed. We use byte-pair encoding for tokenization with a maximum length of 32 for the encoder and 64 for the decoder. We set learning rate to 0.00001 and used AdamW with 1 = 0.9, 2 = 0.98 for optimization. We set the batch size to 32. We trained the KG-BART for 15 epochs, and the gradients are accumulated every 6 steps. We apply dropout with a probability 0.1 to avoid over-fitting. We use beam search with beam size 5 and length penalty with factor 0.6 while inferencing. The training time took 7 hrs on a single NVIDIA GeForce GTX 1080 Ti GPU with 11 GB GDDR5X memory.

[4] https://github.com/yeliu918/KG-BART

7 Baseline and Evaluation

We compare the performance of KG-BART model with T5-base/large (Raffel et al., 2020) and BART-base/large (Lewis et al., 2019) state-of-the-art pre-trained conditional text generation models. Following other conventional generation tasks, we use several widely-used automatic metrics to automatically assess the performance, such as BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004), which mainly focus on measuring n-gram similarities. Table 4 shows the BLEU score and ROUGE score of generated pathological descriptions by KG-BART and T5-base/large and BART-base/large models vs. gold standard pathological descriptions.

8 Qualitative Analysis

KG-BART model performs better on unseen data. Sentence formation of the KG-BART model is better than T5 and BART models when input is in abstract form and random in sequence. As shown in table 5, the output generated for input one by the BART-large model does not consider the extra part of notes which does not occur in the training set. Also, in most cases like example two, the KG-BART model correctly identifies the finding location since KG-BART gets the hierarchical anatomical location information from the KG.

9 Radiology Report Generation Using KG-BART Radiology Text Generator

Radiology report generation includes following main tasks:

- Generate pathological description from radiologists' notes.
- Replace appropriate normal sentences (referred as **normal description**) in normal report template with generated pathological descriptions.

For the first task we use our proposed radiology text generator to generate pathological description from radiologists' notes. This section gives the details of second task; how to replace generated pathological description in normal report template to generate whole report. Our domain experts provide us normal report templates. For example, Male Abdomen Pelvis Ultrasound Normal Report, Female Abdomen Ultrasound Normal Report *etc.* According to patient's gender and scan procedure we provide an appropriate normal report template to the system. System finds the appropriate normal sentences to replace with generated pathological description and replace it. As we discussed with hospitals, radiologists, physicians, *etc.*, they are happy to provide impression by themselves to generate whole report. We add impressions provided by radiologists in impression section and generates the whole report.

9.1 Replace Appropriate Normal Sentences with Generated Pathological Descriptions

We create a parallel corpus for the radiologists' notes and the corresponding normal descriptions. Table 6 shows the samples from the parallel corpus of radiologists' notes and normal descriptions.

We consider following input radiologists' notes: *'Chronic pancreatitis.', 'Cholecystitis with 3 mm gallbladder calculus in lumen.'* and *'Grade ii fatty liver.'* and their corresponding generated pathological descriptions by our radiology text generator, *'Pancreas is slightly small, reveals thin inhomogenous parenchyma. the pancreatic duct is dilated.', 'Gallbladder is distended reveals wall thickening. feature of note is presence of a calculus measuring 3 mm noted in lumen of gallbladder.'* and *'Liver shows moderate increase in echogenicity.'* respectively.

- **Step 1:** We look up similar radiologists' notes to the input radiologist's notes in a parallel corpus shown in the table 6. We utilize the BLEU score to determine the match. Then, we retrieve the appropriate normal description for the matching sample. For example, for input notes, *'chronic pancreatitis'* algorithm gives matched radiologists' notes *'chronic pancreatitis'*. Similarly, for input notes, *'cholecystitis with 3 mm gallbladder calculus in lumen'* algorithm gives corresponding matched notes *'cholecystitis with gallbladder calculus in lumen.'* and for input notes *'grade ii fatty liver'* algorithm gives corresponding matched notes *'Fatty liver'*.
- **Step 2:** We find the normal sentences in normal report template which matches with normal descriptions found in step 1.
- **Step 3:** We replace matched normal sentences in the template with the corresponding generated pathological descriptions.

Figure 3 shows the system interface of radiology report generation.

10 Conclusion

We have constructed a parallel dataset of radiologists' notes and corresponding pathological descriptions. KG-BART for radiology text generation produces high-quality sentences by capturing relationships between the concepts in the input. It also considers default properties from the KG if they are missing in the input concept set to generate more logical and natural sentences. Our approach to construct grounded KGs does not add noise since it only considers entities in the hierarchical path from concept to root node and only adds neighbors with default properties. Experimental results show

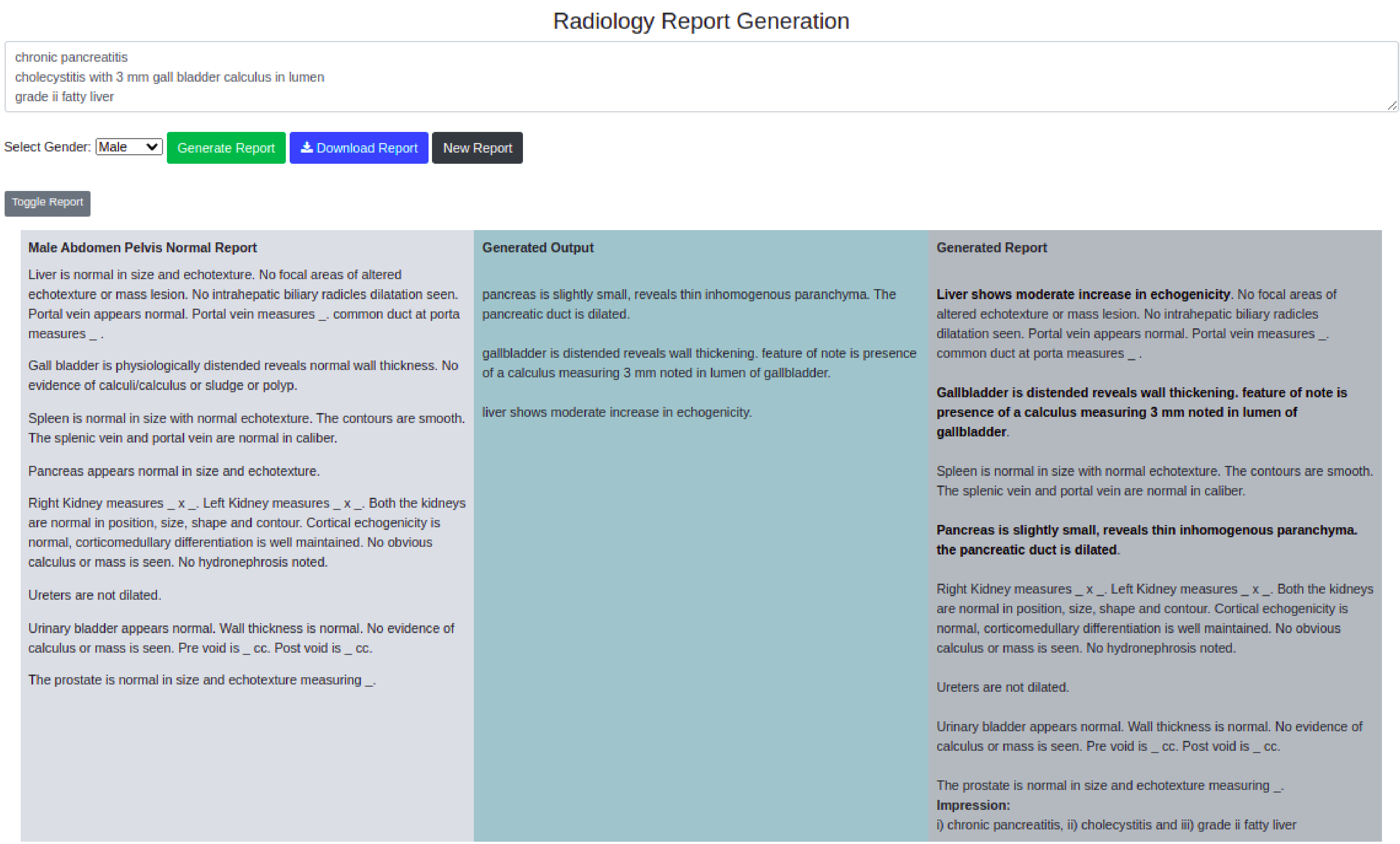

Figure 3: UI interface of our Radiology Report Generation system. The upper text area shows the inputs (radiologist's notes) given to the system. Drop-box gives the option to select the template for male or female. After clicking on **Generate Report** button, the system generates a pathological description using our KG-BART radiology text generator and then generates the whole report. The lower first column shows the normal report template. We display the generated pathological description in the lower mid column. The third column shows the generated report. Once the report is generated, we can download it by clicking on the **Download Report** button.

Input	bulky and retroverted uterus 1 x 5 x 3.4 mm with endometrial thickness 7 mm.
T5-large	uterus is retroverted and bulky in size 1 x 5 x 3.4 mm. myometrial reflectivity is inhomogeneous with an echogenic lesion measuring 7 mm in the mid anterior wall.
BART-large	Uterus is retroverted and bulky in size 1 x 5 & 3.4 mm. Myometrial reflectivity is inhomogeneous and shows normal echotexture. It has uniform epigastrophic echopattern.
KG-BART	uterus is retroverted and bulky in size 1 x 5 x 3.4 mm. myometrial reflectivity is inhomogeneous. Endometrial thickness is 7 mm in size.
Target	Uterus is retroverted and bulky in size measuring 1 x 5 x 3.4 mm. myometrial reflectivity is inhomogeneous and shows bulky echotexture. Endometrial thickness is 7 mm in size.
Input	gross splenomegaly, maximum span of spleen is 8.2 mm with focal calcifications noted in spleen.
T5-large	e/o spleen is grossly enlarged and normal in echotexture, maximum span of stiple is 8.2 mm. multiple calcified granulomas noted in adnexa.
BART-large	spleen is grossly enlarged and normal in echotexture, maximum span of spleen is 8.2 mm. multiple calcified granulomas noted in gb.
KG-BART	Spleen is grossly enlarged and normal in echotexture, maximum span of spleen is 8.2 mm. multiple calcified granulomas noted in spleen.
Target	Spleen is grossly enlarged and normal in echotexture, maximum span of spleen is 8.2 mm. Multiple calcified granulomas noted in spleen.

Table 5: Examples of input (radiologist's notes) and output (pathological description) generated by T5-large, BART-large and KG-BART model.

Radiologists' Notes	**Normal Description**
fatty liver	Liver is normal in size and echotexture.
acute pancreatitis	Pancreas is normal in size and echotexture.
chronic pancreatitis	Pancreas is normal in size and echotexture.
cholecystitis with gallbladder calculus in lumen	Gall bladder is physiologically distended reveals normal wall thickness. No evidence of calculi/calculus or sludge or polyp.

Table 6: Samples from the parallel corpus of radiologists' notes and normal descriptions.

that KG-BART is more capable of producing radiology text than the state-of-the-art T5-base/large and BART-base/large models. In future, we plan to apply the proposed method to generate radiology reports for CT, MRI, *etc.*

Limitations

Available medical reports used to construct the parallel dataset are biased as the abnormal findings for the liver, pancreas, kidney, gallbladder, and uterus are more weighted than organs like ovary, prostate, urinary bladder, etc. Hence, generated pathological descriptions for organs with fewer data are influenced by data with highly weighted organs.

Broader Impact

Automatic radiology report generation can augment radiologists' capabilities to enhance clinical workflows. Our work will help to avoid delays in report turnaround time and human typographical errors in the report. It will speed up the report generation process resulting in faster medical treatment. This will help faster treatment of patients thereby saving many lives.

References

Richa Arora. 2014. The training and practice of radiology in india: current trends. *Quantitative imaging in medicine and surgery*, 4(6):449.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Shivani Choudhary, Tarun Luthra, Ashima Mittal, and Rajat Singh. 2021. A survey of knowledge graph embedding and their applications. *arXiv preprint arXiv:2107.07842*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Janice C Honeyman. 1999. Information systems integration in radiology. *Journal of Digital Imaging*, 12(1):218–222.

Kaveri Kale, Pushpak Bhattachryya, Aditya Shetty, Milind Gune, Kush Shrivastava, Rustom Lawyer, and

Spriha Biswas. 2022. Knowledge graph construction and its application in automatic radiology report generation from radiologist's dictation. *arXiv preprint arXiv:2206.0630808*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2901–2908.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and S Yu Philip. 2021. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6418–6425.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ofir Press and Lior Wolf. 2016. Using the output embedding to improve language models. *arXiv preprint arXiv:1608.05859*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2021. Kepler: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9:176–194.

Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. 2022. A survey of knowledge-enhanced text generation. *ACM Computing Surveys (CSUR)*.

Yixiao Zhang, Xiaosong Wang, Ziyue Xu, Qihang Yu, Alan Yuille, and Daguang Xu. 2020. When radiology report generation meets knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12910–12917.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

A Appendix

A.1 Data Preprocessing

In the corpus, there are a lot of extra spaces and unwanted punctuation marks found. We have removed these unwanted characters from the corpus using regular expressions.

For example, *Liver is enlarged in size(16. 45cm)& normal in shape and shows raised echo reflectivity. No focal or diffuse lesion is seen. The portal and hepatic veins are normal.* In the above example, there is no space between *size, (16.45cm)* and *&*. Also, there is no space between . and *No* and therefore sentence tokenization is challenging. *Liver is enlarged in size (16.45 cm) & normal in shape and shows raised echo reflectivity. No focal or diffuse lesion is seen. The portal and hepatic veins are normal.* The text is then further divided into sentences.

A.1.1 Spelling Correction

In corpus, there are a lot of spelling mistakes also. To correct the spellings we have used the SymSpell library.

Single Word Spelling Correction We have created unigram and bigram dictionaries for corpus text.
Unigram Dictionary: Dictionary of unique correct spelling words, and the frequency count for each word.
Bigram Dictionary: Dictionary of the unique correct spelling of a pair of words, and the frequency count for each pair.

Levenshtein algorithm is used to compute edit distance metric between two strings. Edit distance algorithm finds the correct suggestion for words in input text with words in unigram dictionary.

For example, *enlaregd, billiary, radicals* are the incorrect words found in the corpus. In dictionary enlarged, biliary, radicals these correct words are present. Edit distance algorithm suggests enlarged word for *enlaregd*. Similarly *biliary* for *billiary* and *radicles* for *radicals*.

Multi-word Spelling Correction

- We remove mistakenly inserted spaces within a correct word

Input: *Liver is normal in size and reveals diffuse hypo attenuation*
Output: *Liver is normal in size and reveals diffuse hypoattenuation*

- We add mistakenly omitted spaces between two correct words
  Input: *Liver appears normal in size and reveals mild generalized increasedparenchymal reflectivity.*
  Output: *Liver appears normal in size and reveals mild generalized increased parenchymal reflectivity.*