CHAPTER

17

Natural Language Processing for Large-Scale Medical Image Analysis Using Deep Learning

Hoo-Chang Shin, Le Lu, Ronald M. Summers
National Institutes of Health Clinical Center, Bethesda, MD, United States

CHAPTER OUTLINE

Deep Learning for Medical Image Analysis. DOI: 10.1016/B978-0-12-810408-8.00023-7

CHAPTER POINTS

- Structured information can be obtained from the radiology reports in Picture Archiving and Communication Systems (PACS) using natural language processing (NLP)
- Recent advances in deep learning enables us to analyze large amount of images effectively, though collecting and annotating many medical images has been always a challenge
- Using NLP, lots of radiology images can be collected, annotated, and analyzed, without much human manual efforts

17.1 INTRODUCTION

Recent advances in deep learning methods enable us to analyze complex data types such as images in a much larger scale more effectively. Computer vision research and application has made leaping improvements with the state-of-the-art deep learning methods, setting new records every year.

Despite the tremendous progress in computer vision, there has not been an attempt for machine learning on very large-scale medical image databases. The foremost challenges in performing large-scale medical image analysis somewhat comparable to the scale of recent computer vision studies are:

- The challenge in collecting the images
- The challenge in annotating such collected images

The ImageNet [9] (now-standard computer vision large-scale image dataset) is constructed by collecting images from the Internet, and annotating them using crowd-sourcing. Such approach is infeasible for medical images, due to:

- Privacy concerns
- Lack of expertise (of crowd-sourced annotators)

On the other hand, large collections of radiology images and reports are stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). The invaluable semantic diagnostic knowledge inhabiting the mapping between hundreds of thousands of clinician-created high-quality text reports and linked image volumes remains largely unexplored.

Natural language processing (NLP) provides methods that can help the conversion of text into a structured representation, thereby enabling computers to derive meaning from human input, such as text or speech. Applied on radiology reports, NLP methods can help us to automatically identify and extract relevant informations from the reports so that (i) image collection and (ii) image annotation can be automated.

A radiology report provides a diagnostic imaging referral, and is used for communication and documentation purposes. Although there exist some guidelines for reporting diagnostic imaging, a report mostly contains free text, often organized in a few standard sections. For that reason, it is challenging to convert such free text into a representation that computer can efficiently manage and interact. NLP is a set of method to enable that, to (i) convert unstructured text into a structured form, so that (ii) computer can efficiently store and interact with the information stored.

In this chapter, we show and demonstrate how we can use NLP to collect large radiology image dataset by analyzing radiology text reports. Furthermore, we introduce some existing NLP tools and ontology databases which can be used to annotate large dataset of radiology images. With NLP, images already stored in most hospital PACS database can be retrieved and analyzed, so that we can leverage upon the vast learning capacity of recent deep learning methods.

17.2 FUNDAMENTALS OF NATURAL LANGUAGE PROCESSING

17.2.1 PATTERN MATCHING

The most fundamental method and an integral part of many other NLP tasks is pattern matching. A pattern is a sequence of characters that can be matched, character for character, to a given text. Pattern matching can be often performed with regular expressions (called REs, or regexes, or regex patterns) [20,4], which are essentially a tiny, highly specialized programming language embedded inside most of major programming languages such as Python or Java. A set of possible strings can be matched by specifying the rules using this little programming language.

There are some metacharacters that can be used to match specific character patterns. For example, `[` and `]` are used for specifying a character class. Characters can be listed individually, or a range of characters can be indicated by giving two characters and separating them by a '`-`'. For example, any lowercase letters can be matched with a regular expression of `[a-z]`.

By including a '`^`' as the first character of the class, any characters not listed within the class by complementing the set can be matched. For example, `[^0]` will match any character except '0'.

One of the most important metacharacters for regex is the backslash \, which can be followed by various characters to signal different special sequences. It is also used to escape all the metacharacters so that they can be matched in patterns. For example, `[` or `\` can be matched by preceding them with a backslash to remove their special meaning.

Some of the special sequences beginning with '\' represent predefined sets of characters that are often useful, such as the set of digits, the set of letters, or the set of anything that is not a whitespace. The following predefined special sequences are a subset of those:

- \d – matches any decimal digit; equivalent to `[0-9]`
- \D – matches any non-digit character; equivalent to `[^0-9]`
- \s – matches any whitespace character; equivalent to `[ \t\n\r\f\v]`
- \S – matches any non-whitespace character; equivalent to `[^\t\n\r\f\v]`

```
Dictating Radiologist:      Jane Doe        11/11/2011
Transcribed by:
Signed by:
Signed by:      John Doe 11/11/2011
Approved by:    John Doe 11/11/2011 11:11 AM"
123456,"MR1000012340","MRI Results

Exam:      Exam Date:      Accession #:    Ordered By:
MRI Brain -      11/11/2011      MR1000012340      HOMER, SIMPSON
Patient:  LISA SIMPSON


REASON FOR EXAM: Stroke

CLINICAL HISTORY: Glycogen storage disease. Marked leukomalacia bilaterally ...

TECHNIQUE: Sagittal T1, axial T1 pre- and postcontrast axial FLAIR, axial T2, axial
gradient echo, coronal T1 postcontrast, sagittal T1 3-D postcontrast …

FINDINGS: There is evidence of left parietal encephalomalacia  consistent with known
history of prior stroke. Small focal area of hemosiderin deposition along the lateral
margins of the left lateral ventricle  series 601  image 18.  The orbits and globes are
normal …

IMPRESSION: 1. Diffuse white matter encephalomalacia that is stable when compared
to XX/XX/XXXX examination but has minimally increased ...
```

FIGURE 17.1

An example of a radiology report stored as a free-form text. Pattern matching can be applied to the free-form text of radiology report to (i) convert to a structured text and to (ii) extract image mentioned in the report (highlighted blue).

- \w – matches any alphanumeric character; equivalent to `[a-zA-Z0-9_]`
- \W – matches any non-alphanumeric character; equivalent to `[^a-zA-Z0-9_]`

17.2.1.1 Structuring a Free Form Text

With pattern matching, a free-form text downloaded from a PACS containing many radiology reports can be automatically structured. An example of a radiology report stored in a free-form text is shown in Fig. 17.1. If many reports are stored in a free-form text, pattern matching can be applied to divide them into the reports of separate incidents. For example, a regex like the following example can be built to match the start of a radiology report:

```
re1  =  '(Dictating)(\\s+)(Radiologist)(:)'
re2  =  '(\\s+)'
re3  =  '((?:[a-z][a-z]+))(\\s+)((?:[a-z][a-z]+))'
re4  =  '(\\s+)'
re5  =  '(?:[0]?[1-9]|[1][012])[-:\\/.]'
re6  =  '(?:(?:[0-2]?\\d{1})|(?:[3][01]{1}))[-:\\/.]'
re7 = '(?:(?:[1]{1}\\d{1}\\d{1}\\d{1})|(?:[2]{1}\\d{3}))'
re8  =  '(?![\\d])'
re  =  re1+re2+re3+re4+'('+re5+re6+re7+')'+re8
```

In the above regex, `re1` matches the exact pattern ‘Dictating Radiologist:’, `re2` a white space afterward, `re3` a pattern for a name, `re4` a following white space, and `re5` to `re8` are for matching a calendar date pattern.

17.2.1.2 Extracting the Image Information

An important pattern to extract image information is the *accession number*, *series number*, and *image (slice) number*. The information about the image modality (MR, CT, etc.) can be relevant too, and that can be extracted from the accession number: accession numbers start with `MR` for studies with magnetic resonance imaging (MRI), `CT` for studies with computed tomography (CT), and `NM` for those with positron emission tomography (PET) imaging. Patient identification number (ID) can be relevant if patient history is of an interest.

Patient ID, accession number are listed in the sixth line of the radiology report shown in Fig. 17.1, which can be matched by regex like:

```
re = '(\\d+)(\\s+)(MR)(\\d+)(\\s+)((?:[a-z][a-z]+))'
```

Once the accession number (and possibly the patient ID too) is extracted with regex like shown above, the series number and the image number can be extracted with regex like the following example to get the images mentioned in the report:

```
re1 = ((?:[a-z][a-z]+))(\\s+)
re2 = (series)(\\s+)(\\d+)(\\s+)
re3 = (image)(\\s+)(\\d+)(\\.)
re = re1+re2+re3+re4
```

An example of a pulled image from the radiology report shown in Fig. 17.1, using the pattern matching methods listed above, is shown in Fig. 17.2. We can indeed see that the pulled image is a magnetic resonance image indicating a stroke.

17.2.1.3 Anonymization

Before conducting a research with patient data, the obtained data should be properly anonymized. Anonymization can be also performed using pattern matching, for example, matching the likely patterns of names, patient IDs, dates, telephone numbers, or any other unique identifying number. Moreover, machine learning methods can be applied to complement the pattern matching methods, in order to achieve better performance. More complete studies about anonymization can be found in [2,26,12,39].

17.2.1.4 Negation Detection

Pattern matching is also used for negation detection algorithms such as the popular `NegEx` [6] algorithm. Some common negation phrases like *no signs of*, *unlikely*,

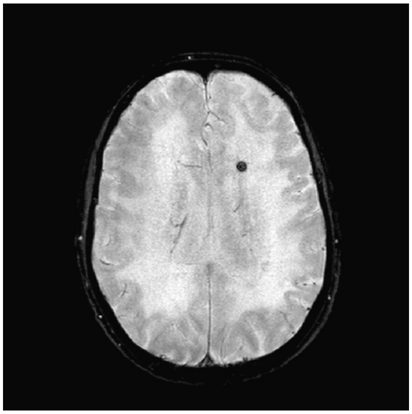

FIGURE 17.2

An example of pulled image from the radiology report shown in Fig. 17.1 with pattern matching.

absence of, or *no evidence of* are found with the pattern matching. These patterns are associated with the following disease mentioning to form negated mentioning of the disease. Nonetheless, machine learning methods can be used for negation detection as well to complement the pattern matching. More details on negation detection can be found in [6,8,21,37].

17.2.2 TOPIC MODELING

Topic modeling is a useful method to summarize a dataset with large text corpus and to obtain gross insight over the dataset. Many popular topic modeling methods characterize document content based on key terms and estimate topics contained within documents. For example, documents associated with the topic *MRI of brain tumor* would comprise one cluster with key terms such as *axial, contrast, mri, sagittal, enhancement, etc.* Documents associated with *arthritis imaging* would comprise another cluster with key terms like *joint, views, hands, lateral, feet*, etc.

One of the most popular methods for topic modeling is Latent Dirichlet Allocation (LDA), originally proposed in [5] to find latent topic models for a collection of text documents such as newspaper articles. There are some other popular methods for document topic modeling, such as Probabilistic Latent Semantic Analysis (pLSA) by [13] and Non-negative Matrix Factorization (NMF) by [18]. Simpler methods such as *k-means clustering* or *k-nearest neighbor* are also often used for topic modeling. More details of different topic modeling methods and comparisons of them can be found in [33,36,38].

For the application of topic modeling in radiology text, document topic modeling using LDA was applied to large corpus of radiology report in [30]. Combined with the images extracted and identified to be associated with each radiology report using pattern matching, convolutional neural networks (CNNs) are trained to classify images into document topics. Given the document topic is useful, it can provide a first-level understanding of a new image classified into a document topic of radiology reports.

17.3 NEURAL LANGUAGE MODELS

Statistical language modeling aims to predict the next word in textual data given context. Therefore, we deal with sequential data prediction problem when constructing language models. Traditionally, such statistical models were the approaches that are very specific for language domain. For instance, assuming that natural language sentences can be described by parse trees, or that they can be considered as morphology of words, syntax, and semantics. Most widely used language models are n-gram models, which assume that language consists of sequences of words that form sentences.

Likewise in computer vision, the recent advances in deep learning has changed NLP research. Some of the previous state-of-the-art NLP methods are replaced by deep learning based approaches, and the trend is continuing. In this section we will briefly introduce neural language modeling – a more recent approach to NLP using neural networks, and discuss how they are used in radiology text mining applications.

17.3.1 WORD EMBEDDINGS

Most of the machine learning algorithms benefits from rich feature representations – the better the feature representations are (usually), the better a machine learning algorithm performs. On the contrary, even very sophisticated algorithms are basically limited by the dimension of the features representing the data. Traditionally in NLP, words in text were given a unique ID (e.g., an integer number). Such encodings are not only arbitrary, but also provide no useful information regarding the relationships or similarities between the individual symbols.

Word embedding is a technique to convert words in to vectors, where words with similar vectors would have close cosine distance, and can improve many NLP tasks. By representing words as vectors, the data representation is less sparse, thereby statistical models can be more successfully trained with less data.

Words can be converted to vectors in an unsupervised manner using feed-forward neural networks. A feed-forward neural network such as shown in Fig. 17.3 can be used to learn word embeddings from a large corpus of text. The feed-forward neural network shown in Fig. 17.3 learns the neural word embeddings by learning to predict the next word after seeing three consecutive words in the text corpus.

There are many options and advanced techniques to learn better word embeddings, which can be reviewed in [23,22,25,24]. Fig. 17.4 shows some example of the words trained with the word-to-vector model of [23,22,25,24], applied to large cor-

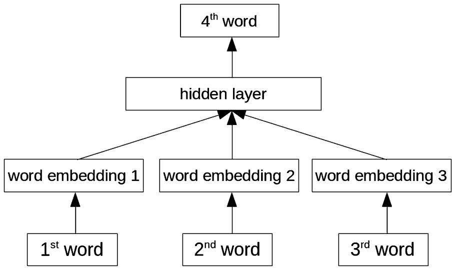

FIGURE 17.3

An example of a feed-forward neural network architecture to learn word embeddings.

"heart"		**"brain"**		**"liver"**	
lungs	0.526600	t1	0.615066	spleen	0.759884
mediastinum	0.517008	mri	0.595027	gallbladder	0.648075
consolidating	0.486605	sagittal	0.580841	hepatomegaly	0.642022
pa	0.449816	flair	0.565445	gallstones	0.611837
chest	0.433362	t2	0.555053	pancreas	0.608356
infiltrates	0.428404	axial	0.554040	gallstone	0.606063
hyperinflated	0.413326	spgr	0.520954	steatosis	0.601081
cardiomegaly	0.410785	weighted	0.502047	dome	0.594812
hyperlucent	0.400836	technique	0.487768	portal	0.570008
pectus	0.396142	astrocytoma	0.480527	ascites	0.551869
great	0.395712	gbm	0.476956	hepatosplenomegaly	0.540501
ectatic	0.394560	gradient	0.476593	hepatic	0.537453
shifted	0.389205	oligodendroglioma	0.465892	cirrhosis	0.530389
ray	0.389091	postcontrast	0.463686	fatty	0.522134

FIGURE 17.4

Word-to-vector models trained on a collection of radiology reports. Search words (with quotes) and their closest words in vector-space cosine similarity (the higher the better) are listed in a descending order. Image is adapted from [29].

pus of radiology text [30]. Some search words (with quotes) and their closest words in vector-space cosine similarity (the higher the better).

We can see that words with similar meanings are reasonably well encoded with the word embeddings. Word embedding was applied to radiology text reports in [30] to overcome ambiguities of radiology words with similar meanings, and to generate word vectors of keywords describing radiology images.

17.3.2 RECURRENT LANGUAGE MODEL

Neural language models based on feed-forward neural networks have a limitation that the input dimension is fixed. Still, many attempts to obtain such statistical mod-

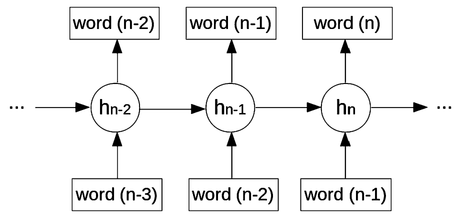

FIGURE 17.5

An example of a recurrent neural network language model.

els involve approaches that are very specific for language domain – for example, assumption that natural language sentences can be described by parse trees, or that we need to consider morphology of words, syntax and semantics.

Even the most widely used and general models, based on n-gram statistics, assume that language consists of sequences of atomic symbols – words – that form sentences, and where the end of sentence symbol plays important and very special role. For example, for the word embedding model shown in Fig. 17.3, once we fix the window size (3: *1st word* . . . *3rd word*), the entire model is limited to the fixed size.

Recurrent language models avoid this limitation, so that a predicted output word can be conditioned on a sequence of arbitrary number of words appeared in the past, not limited by fixed window size. This is an important improvement to language modeling, because a dependency of a word can span to ten or more occurrences of the word in the past. For example, more than ten words in the past needs to be linked to the image location in order to successfully relate the disease mentioned with the image:

```
``Comparison is made with a study of XX/XX/XXXX.
  No change not attributable to a difference in
  technique and bowel content is seen. Findings
  include 1 liver lesions series 601 image 31.''
```

An example of a recurrent neural language model is shown in Fig. 17.5. As shown in the figure, the dependency for an *n*th word prediction can be infinitely long to the occurrences of the previous words. Using recurrent neural networks for NLP, state-of-the-art results were obtained in machine translation [34].

A combination of convolutional neural networks and recurrent neural networks are often used for image caption generation [15,11,35], and this setting was used in [31] to generate image descriptions of chest X-rays. More details of recurrent neural language model can be found in [34,23,3].

17.4 MEDICAL LEXICONS

Lexicons are collections of unique concepts accompanied by a preferred "term" (name) and a list of synonyms and derivational forms. Some of the popular medical lexicons used in radiology are UMLS Metathesaurus [28] and RadLex [16].

17.4.1 UMLS METATHESAURUS

The Unified Medical Language System (UMLS) of [19,14] integrates and distributes key terminology, classification and coding standards, and associated resources to promote the creation of more effective and inter-operable biomedical information systems and services, including electronic health records. It is a compendium of many controlled vocabularies in the biomedical sciences, created in 1986 and maintained by the National Library of Medicine.

The Metathesaurus [28] forms the base of the UMLS and comprises over 1 million biomedical concepts and 5 million concept names, where all of them are collected from the over 100 incorporated controlled vocabularies and classification systems. Arthritis is defined in the Metathesaurus by the unique alphanumeric code (or Concept Unique Identifier (CUI)) C0003864. It includes synonymous terms for each concept (for example, *stroke* and *cerebrovascular accident*) for CUI C0038454.

Moreover, specific semantic types for each concept and their types and relationships are organized as an ontology. There are 133 semantic types in the UMLS Metathesaurus ontology, such as "T017: anatomical structure", "T074: medical device", "T184: sign or symptom", "T033: finding", and "T047: disease or syndrome."

17.4.2 RADLEX

RadLex [16] is a unified language to organize and retrieve radiology imaging reports and medical records. While the Metathesaurus has a vast resource of biomedical concepts, RadLex offers more radiology-specific terms, including imaging techniques and radiology devices. It is designed as a comprehensive lexicon for standardizing indexing and retrieval of radiology information resources.

17.5 PREDICTING PRESENCE OR ABSENCE OF FREQUENT DISEASE TYPES

17.5.1 MINING PRESENCE/ABSENCE OF FREQUENT DISEASE TERMS

A large collection of radiology images can be retrieved and labeled with associated disease words, using the lexicons and the pattern matching techniques discussed above. We apply these to the PACS radiology database used in [30] containing about 780,000 imaging examinations of about 60,000 unique patients. We extract images mentioned with (i) radiology words appearing in RadLex; (ii) having "T047: disease or syndrome" semantics in the UMLS Metathesaurus that also appears in the RadLex.

Table 17.1 Some statistics of images-to-disease assertion/negation label matching

# images		Per image mean/std		# assertions per image		# negations per image	
Total matching	18,291	# assertions mean	1.05	1/image	16,133	1/image	1581
With assertions	16,827	# assertions std	0.23	3/image	81	3/image	0
With negations	1665	# negations std	0.22	4/image	0	4/image	0

Furthermore, we use negation detection algorithm [6,7] to detect presence and absence of the disease terms in the associated image. Some frequently mentioned disease terms detected with assertion are: *cyst, disease, infection, pneumonia, bronchiectasis, abscess*, with each of them being detected more than 300 times. These terms are similarly mentioned more than 300 times with negation. We ignore the disease terms mentioned less then 10 times in the dataset.

The total number of disease terms with assertion and negation are 77 (59 with assertion and 18 with negation). If more than one disease term is mentioned for an image, we simply assign the terms multiple times for an image. Some statistics on the number of assertion/negation occurrences per image are shown in Table 17.1.

17.5.2 PREDICTION RESULT AND DISCUSSION

When we train a deep convolutional neural network model of [32] on the image-to-disease-label dataset we mined from the PACS dataset [30], and achieve the top-1 test accuracy achieved of 0.71, and top-5 accuracy of 0.88. Some examples of test cases where top-1 probability output matches the originally assigned disease labels are shown in Fig. 17.6. It is noticeable that disease words are detected with high probability when there is one disease word per image, but with relatively lower top-1 probability for one disease word and other words within the top-5 probabilities (Fig. 17.6B – "...infection abscess").

It can be also observed that automatic label assignment to images can sometimes be challenging. In Fig. 17.6D "cyst" is assigned as the correct label based on the original statement "...possibly due to cyst ...", but it would be unclear whether cyst will be present in the image (and the cyst is not visibly apparent). It applies similarly to Fig. 17.6E where the presence of "osteophyte" is not clear from the referring sentence but is assigned as the correct label (and osteophyte is not visibly apparent on the image). In Fig. 17.6F "no cyst" is labeled and predicted correctly, but it is not obvious what to derive from this prediction that indicates an absence of a disease versus a presence.

Some examples of test cases where top-1 probability does not match the originally assigned labels are shown in Fig. 17.7. Four (A, C, E, and F) of the six examples, however, contain the originally assigned label in the top-5 probability predictions, which is coherent with the relatively high (88%) top-5 prediction accuracy.

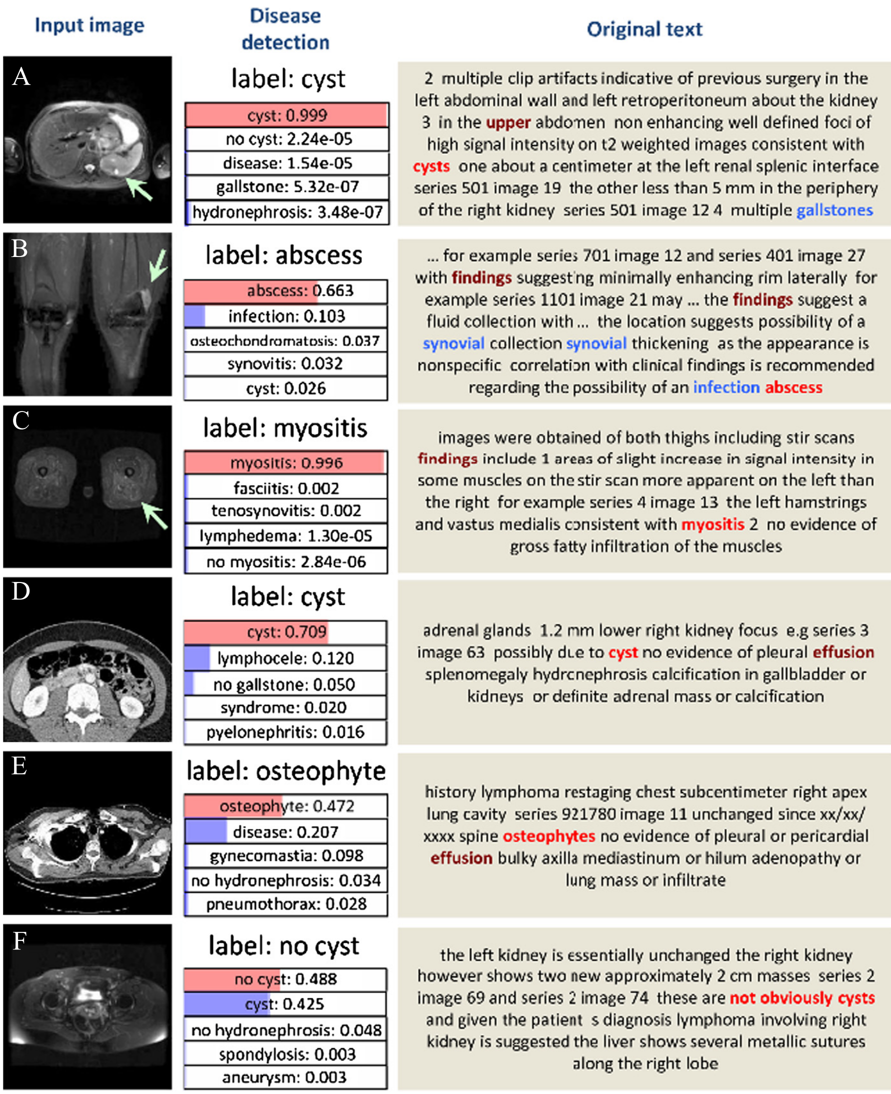

FIGURE 17.6

Some examples of final outputs for automated image interpretation, where top-1 probability matches the originally assigned label. Specific disease words assigned as label mentioned in the reports are shown in bold red, and disease words predicted with top-5 probability in the reports are shown in bold blue. The probability assigned to the originally assigned label is shown with a red bar, and the other top-5 probabilities are shown with blue bars. Disease region identified in an image is pointed by arrow. Image is adapted from [29].

Here again, Fig. 17.7A is automatically labeled as "cyst", but the cyst is not clearly visible on the image where the original statement "...too small to definitely characterize cyst ..." supports this. The example of Fig. 17.7B shows a failed case of assertion/negation algorithm, where "cyst" is detected as negated based on the statement "...small cyst". Nonetheless, true label ("cyst") is detected as its top-1 probability. For Fig. 17.7C "cyst" is predicted where the true label assigned was "abscess"; however, cyst and abscess are sometimes visibly similar.

It is unclear whether we should expect to find emphysema in the image from the statement such as "...possibly due to emphysema" (and emphysema is not visibly present), similarly to Fig. 17.6D. Therefore, it would be challenging to correctly interpret such statement for label assignment. Fig. 17.7E shows a disease which can be bronchiectasis, though it is not clear from the image. Nonetheless, bronchiectasis is predicted with the second highest probability. Bronchiectasis is visible in Fig. 17.7F, and it is predicted with second highest probability, too.

17.5.2.1 Discussion

Automated mining of disease terms enables us to predict a disease from a large amount of images with promising result. However, matching the images to more specific disease words and assigning labels is not always straightforward as was shown for the examples in Fig. 17.7. A big part of challenge is on developing a better natural language processing algorithms to (i) infer better image labels, to (ii) deal with uncertainties in radiology text, and to (iii) better relate words of different sentences for more composite labeling.

Some NLP challenges in mining more images and assigning better image labels are:

- Stemming (e.g., kidneys → kidney, vertebrae → vertebra)
- Anaphora resolution [17] (e.g., findings are mentioned in a different sentence than the sentence referring to the image)
- Part-of-speech (POS) tagging [10] (e.g., whether "spleen" is a noun, or "unremarkable" is an adjective, etc.)
- Quantification of hedging statements [27,37]
  - "Too small to characterize" → present with 40% confidence
  - "possibly degenerative" → present with 50% confidence
- Coreference resolution [1] (identify words for size, location, etc., and relate these with the words they are describing)

Some of the above mentioned challenges are specific to NLP in radiology text (e.g., stemming, POS tagging are regarded not challenging in general NLP), though the others are more generic NLP challenges. Also, comprehensive analysis of hospital discharge summaries, progress notes, and patient histories might address the need to obtain more specific information relating to an image even when the original image descriptions are not very specific.

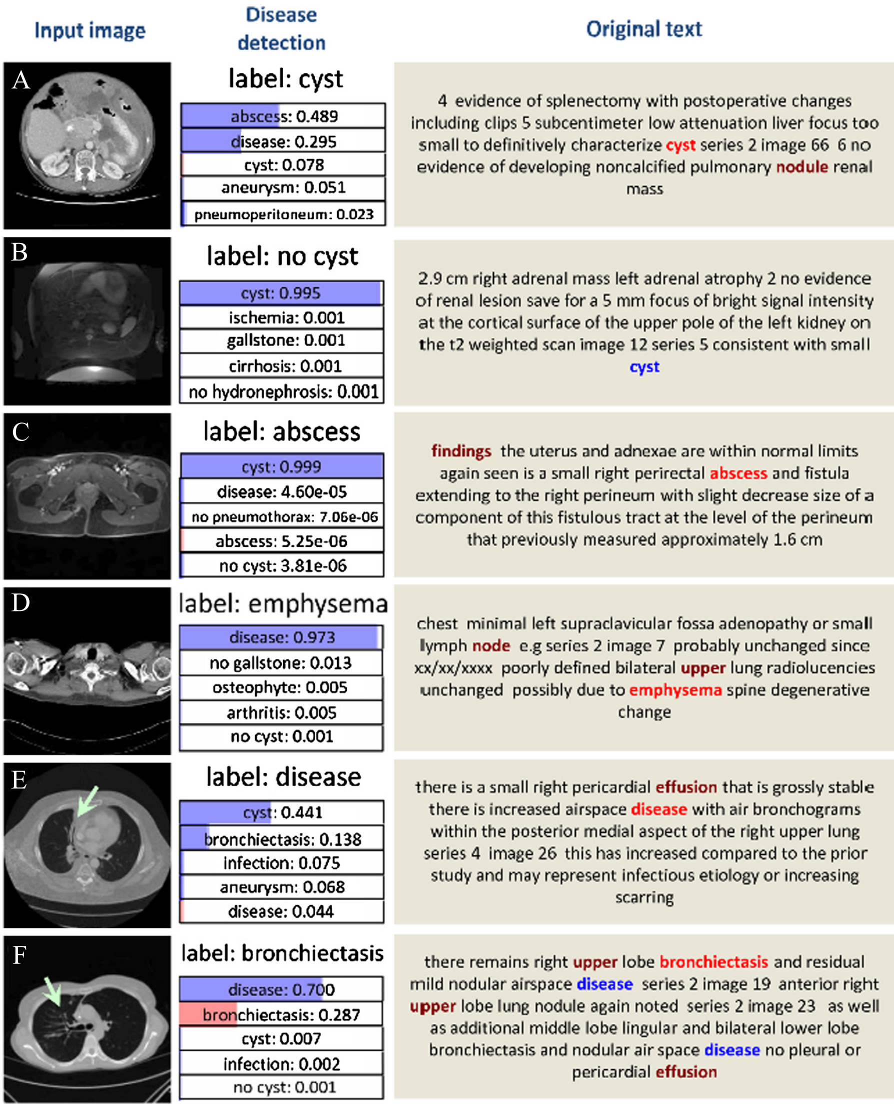

FIGURE 17.7

Some examples of final outputs for automated image interpretation where top-1 probability does not match the originally assigned label. One of the top-5 probabilities match the originally assigned labels in the examples of images A, C, D, and F. None of the top-5 probabilities match the originally assigned labels in the examples of image B and D. However, label assignment for image B is incorrect, as a failed case of assertion/negation detection algorithm used. Nonetheless, the CNN predicted “true” label correctly (“cyst”). Image is adapted from [29].

17.6 CONCLUSION

It has been unclear how to extend the significant success in image classification using deep convolutional neural networks from computer vision to medical imaging. What are the clinically relevant image labels to be defined, how to annotate the huge amount of medical images required by deep learning models, and to what extent and scale the deep CNN architecture is generalizable in the medical image analysis are open questions.

In this chapter, we present an approach to mine and label images from a hospital PACS database using natural language processing. Natural language processing enables us to conduct large-scale medical image analysis, which has been very challenging with manual data collection and annotation approaches. Basic natural language processing methods as well as more recent deep-learning-based approaches for mining large radiology reports are introduced.

We also demonstrated to mine and match frequent disease types using disease ontology and semantics from a large PACS database, and demonstrate prediction of the presence/absence of disease from a radiology image with probability outputs. Exploring effective deep learning models on this database opens new ways to parse and understand large-scale radiology image informatics.

ACKNOWLEDGEMENTS

This work was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center, and in part by a grant from the KRIBB Research Initiative Program (Korean Biomedical Scientist Fellowship Program), Korea Research Institute of Bioscience and Biotechnology, Republic of Korea. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). We thank NVIDIA for the K40 GPU donation.

REFERENCES

1. E. Apostolova, D. Demner-Fushman, Towards automatic image region annotation: image region textual coreference resolution, in: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, Association for Computational Linguistics, 2009, pp. 41–44.
2. A. Aronson, O. Bodenreider, H. Chang, S. Humphrey, J. Mork, S. Nelson, T. Rindflesch, W. Wilbur, I. Initiative, et al., A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications, 1999.
3. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv:1409.0473, 2014.
4. S. Bird, NLTK: the natural language toolkit, in: Proceedings of the COLING/ACL on Interactive Presentation Sessions, Association for Computational Linguistics, 2006, pp. 69–72.

5. D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
6. W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, J. Biomed. Inform. 34 (5) (2001) 301–310.
7. W.W. Chapman, D. Hilert, S. Velupillai, M. Kvist, M. Skeppstedt, B.E. Chapman, M. Conway, M. Tharp, D.L. Mowery, L. Deleger, Extending the NegEx lexicon for multiple languages, Stud. Health Technol. Inform. 192 (2013) 677.
8. M.-C. De Marneffe, A.N. Rafferty, C.D. Manning, Finding contradictions in text, in: ACL, vol. 8, 2008, pp. 1039–1047.
9. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, IEEE, 2009, pp. 248–255.
10. S.J. DeRose, Grammatical category disambiguation by statistical optimization, Comput. Linguist. 14 (1) (1988) 31–39.
11. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625–2634.
12. O. Ferrández, B.R. South, S. Shen, F.J. Friedlin, M.H. Samore, S.M. Meystre, Bob, a best-of-breed automated text de-identification system for VHA clinical documents, J. Am. Med. Inform. Assoc. 20 (1) (2013) 77–83.
13. T. Hofmann, Probabilistic latent semantic indexing, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 1999, pp. 50–57.
14. B.L. Humphreys, D.A. Lindberg, H.M. Schoolman, G.O. Barnett, The unified medical language system an informatics research collaboration, J. Am. Med. Inform. Assoc. 5 (1) (1998) 1–11.
15. A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, June 2015.
16. C.P. Langlotz, RadLex: a new method for indexing online educational materials 1, Radiographics 26 (6) (2006) 1595–1597.
17. S. Lappin, H.J. Leass, An algorithm for pronominal anaphora resolution, Comput. Linguist. 20 (4) (1994) 535–561.
18. D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
19. D.A. Lindberg, B.L. Humphreys, A.T. McCray, The unified medical language system, Methods Inf. Med. 32 (4) (1993) 281–291.
20. J.H. Martin, D. Jurafsky, Speech and Language Processing, International Edition, 2000, p. 710.
21. S. Mehrabi, A. Krishnan, S. Sohn, A.M. Roch, H. Schmidt, J. Kesterson, C. Beesley, P. Dexter, C.M. Schmidt, H. Liu, et al., DEEPEN: a negation detection system for clinical text incorporating dependency relation into NegEx, J. Biomed. Inform. 54 (2015) 213–219.
22. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv:1301.3781, 2013.
23. T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, S. Khudanpur, Recurrent neural network based language model, in: INTERSPEECH, vol. 2, 2010, p. 3.

24. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.
25. T. Mikolov, W.-t. Yih, G. Zweig, Linguistic regularities in continuous space word representations, in: HLT-NAACL, 2013, pp. 746–751.
26. I. Neamatullah, M.M. Douglass, H.L. Li-wei, A. Reisner, M. Villarroel, W.J. Long, P. Szolovits, G.B. Moody, R.G. Mark, G.D. Clifford, Automated de-identification of free-text medical records, BMC Med. Inform. Decis. Mak. 8 (1) (2008) 1.
27. F. Salager-Meyer, Hedges and textual communicative function in medical English written discourse, Engl. Specif. Purp. 13 (2) (1994) 149–170.
28. P.L. Schuyler, W.T. Hole, M.S. Tuttle, D.D. Sherertz, The UMLS metathesaurus: representing different views of biomedical concepts, Bull. Med. Libr. Assoc. 81 (2) (1993) 217.
29. H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, R.M. Summers, Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation, J. Mach. Learn. Res. 17 (107) (2016) 1–31.
30. H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, R.M. Summers, Interleaved text/image deep mining on a very large-scale radiology database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1090–1099.
31. H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, R.M. Summers, Learning to read chest X-rays: recurrent neural cascade model for automated image annotation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
32. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556, 2014.
33. K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring topic coherence over many models and many topics, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, 2012, pp. 952–961.
34. I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, in: Advances in Neural Information Processing Systems, 2014, pp. 3104–3112.
35. O. Vinyals, A. Toshev, S. Bengio, D. Erhan, Show and tell: a neural image caption generator, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3156–3164.
36. H.M. Wallach, Topic modeling: beyond bag-of-words, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, 2006, pp. 977–984.
37. A.S. Wu, B.H. Do, J. Kim, D.L. Rubin, Evaluation of negation and uncertainty detection and its impact on precision and recall in search, J. Digit. Imaging 24 (2) (2011) 234–242.
38. P. Xie, E.P. Xing, Integrating document clustering and topic modeling, arXiv:1309.6874, 2013.
39. R. Yeniterzi, J. Aberdeen, S. Bayer, B. Wellner, L. Hirschman, B. Malin, Effects of personal identifier resynthesis on clinical text de-identification, J. Am. Med. Inform. Assoc. 17 (2) (2010) 159–168.