



Structured Reporting and Artificial Intelligence

8

Salvatore Claudio Fanni,
Michela Gabelloni,
Angel Alberich-Bayarri,
and Emanuele Neri

Contents

| | | |
|-------|----------------------------------------------------------------------------------------------|-----|
| 8.1 | Introduction..... | 170 |
| 8.2 | Natural Language Processing: How Does It Work? An Overview on the Technical Workflow..... | 171 |
| 8.2.1 | Feature Extraction..... | 171 |
| 8.2.2 | Feature Processing, from Machine Learning to Deep Learning..... | 173 |
| 8.3 | Application of Natural Language Processing in Radiology..... | 175 |
| 8.4 | Structured Reporting as AI Annotation Strategy..... | 177 |
| 8.5 | Quantitative Structured Reporting..... | 178 |
| | References..... | 181 |

S. C. Fanni · M. Gabelloni · E. Neri
Academic Radiology, Department of Translational Research, University
of Pisa, Pisa, Italy

Italian Society of Medical and Interventional Radiology, SIRM
Foundation, Milan, Italy
e-mail: michela.gabelloni@unipi.it; emanuele.neri@unipi.it

A. Alberich-Bayarri (✉)
Quantitative Imaging Biomarkers in Medicine (QUIBIM SL), Valencia, Spain
e-mail: angel@quibim.com

© European Society of Medical Imaging Informatics
(EuSoMII) 2022

169

M. Fatehi, D. Pinto dos Santos (eds.), *Structured Reporting in
Radiology*, Imaging Informatics for Healthcare Professionals,
https://doi.org/10.1007/978-3-030-91349-6_8

8.1 Introduction

Structured radiology reporting has proved to be not only useful but also necessary in order to achieve completeness, comparability, and quantification and to minimize ambiguity [1]. The introduction of electronic medical record (EMR) holds the promise of advancing clinical research by allowing analysis of data contained in the radiology reports; unfortunately, this is extremely difficult in free-form text, while it is quicker and easier in structured reports [2].

Nowadays, structured reporting is still not widely used due to many reasons, such as the fact that technical difficulties and lack of integration make it time consuming; therefore, many radiology reports remain unstructured and use a free-form language [3].

Artificial intelligence (AI) may be the way to overcome these issues.

AI is a large area of study in the field of computer science, which deals with the development of tools able to perform human tasks or processes such as learning, reasoning, and self-correction [4].

A subfield of AI is natural language processing (NLP), also defined as “information extraction” or “text mining.”

NLP is already part of our daily life, although little is known. For example, the system that separates valid e-mails from spam is based on text classification performed by an NLP tool.

NLP is a computer-based method that analyzes free-form text, in our case radiology reports, by combining linguistics, statistical, and AI methods, like machine learning (ML) or deep learning (DL).

The final output of this process is a structured format of specific itemized elements with a predefined organization and standardized terminology for each element [3].

From this analysis, NLP automatically identifies and extracts features, which ML or DL algorithm process, for example, to classify radiology reports [5].

Nevertheless, NLP will be useful in a transition period, passing from unstructured to structured reporting. The appropriateness of

software and templates integration will allow for fast reporting also in a structured way [6–8], shortening the elapsed time in the reporting process [9]. The two biggest radiological scientific societies, the Radiological Society of North America (RSNA) and the European Society of Radiology (ESR), established the template library advisory panel (TLAP) to endorse specific structured reporting templates. The most relevant template database can be accessed through the RadReport portal (www.radreport.org), created by the RSNA.

The use of structured reporting templates is also the way in which images to be used for the creation of AI models can be properly annotated with the radiological findings.

A further step in the structured reporting is the inclusion of automatically generated quantitative imaging biomarkers in the report. The goal is not to create a fully quantitative report, which would resemble the way in which blood tests are reported, but to combine the findings detected by the radiologist with the associated annotations and quantitative metrics derived with a perfect combination between quantitative data and radiologist impressions.

8.2 Natural Language Processing: How Does It Work? An Overview on the Technical Workflow

8.2.1 Feature Extraction

NLP analysis starts off with preprocessing feature extraction, which is articulated in various steps. The different tools used in clinical practice and research implement in various ways the different possible steps that we are going to describe.

The first preprocessing steps are segmentation, sentence splitting, and tokenization.

Segmentation is defined as the identification of radiology reports sections, and the successive processing steps may be performed on every section or just a subset.

Further processing steps are divided into sentences, defined as sentence splitting, and into words, that is, tokenization [5].

Words, when separated, are characterized by considering the respective lexical root (stemming). Eventual spelling mistakes are fixed, and eventual abbreviations are expanded.

After normalization of the words, the syntactic analysis assigns part of speech of the words (noun, adjective, verb), their grammatical structure, and dependency relations [10].

The next stage is the semantic analysis, in order to identify for each word an individual concept and their modification by other contiguous terms. A concept is defined as a unique entity with a definite and unambiguous meaning. To standardize the medical language processing, the different software adopted medical lexicons. Lexicons are collections of precise definitions of concepts, each one with a preferred term and a list of possible synonymous or specific semantic [3]. Such lexicons are manually created by experts but may also be combined with existing lexicons; one of the most used is the Unified Medical Language System (UMLS) Metathesaurus [11].

When semantic analysis is completed, each individual concept is ideally output as a separate item in a structured format, which includes other contiguous concepts that modify it (e.g., for the concept of pneumonia, the anatomic location, or chronicity).

The primary NLP technologies used for these purposes are pattern matching and linguistic analysis.

Pattern matching is the simplest technique for searching text, and it is frequently integrated into more complex NLP tasks: it is based on matching of pattern, that is, a sequence of characters, to a given text.

Pattern matching, for example, is used in the above-mentioned process of stemming, in order to reduce a given word to its root and facilitate the connection to the relative lexicon concept.

Pattern matching could be used even to determine whether a concept is present or absent. NegEx is a pattern matching based on an algorithm, used to detect negation lexical words, such as “no” or “absent,” within a small number of words before and after a specific concept [12].

Linguistic analysis is a more complex computer algorithm that uses syntactic and semantic knowledge to infer what concepts are cited in the text and how each concept is related to other contiguous concepts.

Limitations of this approach are ambiguity, incorrect grammar use, and misspellings.

An example of NLP resource based exclusively on linguistic analysis is Medical Extraction and Encoding (MedLEE), developed at New York Presbyterian Hospital [13]. MedLEE processes chest X-ray reports using semantic knowledge, and the final output is a structured format with a list of findings and associated modifiers for each finding [14].

8.2.2 Feature Processing, from Machine Learning to Deep Learning

The combined steps mentioned above produce the NLP features. Features are individual properties or characteristics of the subject of analysis. One of the simplest features in NLP is the n -grams, i.e., the consecutive number of words in a text.

However, concepts identified by semantic analysis have been shown to be more predictive features compared to n -grams [15]. Unfortunately, not all the words contained in the text can be reduced to a concept, such as conjunction or adverbs, however relevant and significant to achieve a complete comprehension of the radiological report.

The extracted features could be used to achieve text classification or information extraction. To solve this task, textual features can be processed by statistical, machine learning (ML), deep learning (DL) approach, or even hybrid approach.

ML is the branch of AI that studies the development of computer algorithms able to learn from data [16]. While the statistical approach utilizes hand-crafted statistics rules, the machine learning approach automatically generates the classification rules.

ML can be used even to achieve linguistic tasks.

The Statistical Assertion Classifier (StAC) performs the same function of the previously mentioned pattern-matching-based tool NegEx. However, StAC works with a completely different and more complex technique. In fact, StAC is an ML algorithm that learns what negations are by analyzing radiology reports previously labeled by humans for the presence/absence of negations [17].

The ML algorithm is mostly integrated in NLP processing with the purpose of classification of radiology reports analyzing the extracted features.

The simplest way is to classify reports by analyzing the presence/absence of findings and their possible combination.

For example, if findings such as pneumonia or infiltrates are described in a chest x-ray report by an NLP tool, then the report is likely classified as positive for pneumonia [18].

Machine learning algorithms perform report classification tasks by analyzing data and automatically determining which features correlate with a positive or negative result.

In order to achieve these results, machine learning methods previously require training labeled data to establish a connection between the extracted features and predefined class. Care must be taken in the choice of the number and type of data because the performance of the classifier strongly depends on the training set [19].

A subfield of machine learning is deep learning (DL). In DL, the algorithm learns without any prior human feature selection [20].

DL models are based on artificial neural networks (ANNs), inspired by the neural cortex, where each neuron is connected with other neurons [20].

ANNs are a collection of artificial neurons organized in multiple layers, structured as input, hidden computation, and output layers [21].

The information is fed through the input layer, processed through the hidden layers, and the result is produced from the output layer.

The most used ANNs are convolutional neural networks (CNNs) and recurrent neural networks (RNNs). A CNN model is usually composed of numerous convolutional layers followed by a few fully connected layers [22]. CNN uses a repeating pattern in the dataset [20].

As in images, repeating patterns also appear in the free-form text [23].

Conversely, RNNs process sequential information, which is ideal in NLP, because sentences are sequences of words. The neurons in RNNs are connected sequentially, like a long chain, each passing the respective output to the next neuron. The sequential passing of information creates a memory; unfortunately, in

long-distance sentences, the “memory effect” loses effectiveness, while the memory diminishes passing through numerous layers.

To overcome this issue, a subtype of RNNs has been developed, the long short-term memory network, which is more effective for analyzing long and complex radiology reports [24].

DL algorithms have outperformed traditional NLP methods in various tasks, leading to a significant increase in research in this field [25].

For these reasons, it is expected that DL applications in NLP will play a largest and important role in clinical practice in coming years.

8.3 Application of Natural Language Processing in Radiology

NLP in radiology is already used for many purposes, and the largest application categories are the following:

- Identifying/classifying findings
- Identifying cases/cohort for research studies
- Identifying follow-up recommendations
- Imaging protocol determinations
- Diagnostic surveillance
- Assessing the quality of radiologic practice

The major benefit is automation and evaluation of large amounts of data in a reasonable time, while performing these tasks without using NLP and AI is at least unthinkable.

One of the first applications of NLP was **identifying/classifying findings**. In 1998, Knirsch et al. compared MedLEE, a traditional NLP tool based on linguistic analysis, with experts review in order to identify chest x-ray reports suspicious for tuberculosis. The purpose was to identify automatically from the radiological report of the patient who needs respiratory isolation protocol. The agreement was 89–92% focusing on the presence/absence of six pre-selected keywords in the report [26].

MedLEE is also one of the first NLP tools used for **identifying cases/cohort for research studies**.

Hripcsak et al. used MedLEE for large-scale research on radiological reports, in order to test four different hypotheses. The automated analysis has made it possible to analyze a huge number of reports: 889.921! [27].

AI represented one of the most important innovations in the NLP field; in fact, ML and DL methods outperformed different times the traditional tools. In order to compare different NLP tools, different quantitative parameters have been used. F1 score is one of these parameters; it is a harmonized average of sensitivity and positive predictive value (PPV) and is frequently used as an overall measure of NLP tools' performance.

An application of NLP has been to classify radiology reports of contrast material-enhanced CT of the chest performed to evaluate pulmonary embolism. In 2012, Chapman et al. developed an NLP tool named PeFinder (i.e., pulmonary embolism finder) for this purpose. PeFinder classified reports based on the presence/absence and location of pulmonary embolism, chronicity, and certainty. PeFinder applied an extension of NegEx to identify lexical clues and define concepts (i.e., pulmonary embolism). This simple technology achieved good results, such as high sensitivity and specificity [28].

Cheng et al. in 2018 compared a CNN model with peFinder, which was considered the best available software for this specific purpose.

However, the CNN model outperformed PeFinder based on F1 score (0.938 vs. 0.867) [29].

Miao et al. evaluated the extraction of BI-RADS findings from breast ultrasound reports. They compared three different types of NLP approach: a traditional role-based approach, a machine-learning approach, and an RNN model. The RNN model performed better than the other methods [30].

Another important application for NLP is the **automatic identification of follow-up recommendations** from radiology reports. Nowadays, this task remains challenging due to a lack of standardized/structured reporting.

In 2019, Carrodegua et al. assessed about 1000 radiology reports for this purpose, evaluating traditional NLP tools (iSCOUT) and ML (Support Vector Machine) and DL models (RNN network). The highest F1 score was achieved by ML models (0.85), while iSCOUT and DL models performed at 0.71.

Imaging protocol determination is a helpful application for NLP in radiology in order to save time and potentially standardize and decrease errors of contrast material injection.

In 2017, Trivedi et al. used the Watson DL protocol to evaluate the need for intravenous contrast injection in musculoskeletal MRI based on the free-text of clinical indication provided for the study. The DL protocol achieved an accepting accuracy (80–90%), resulting in a good clinical decision support tool [31].

Another important NLP task that needs to be mentioned is **diagnostic surveillance** in order to safeguard clinical practice and potentially reduce the chance of errors in communication between radiologists and clinicians. NLP tools developed for this specific task raise alerts for the presence of predetermined findings/conditions contained in the radiology report.

Rink et al. developed a hybrid approach involving a customized lexicon, manually defined patterns and an ML model (support vector machine) able to identify appendicitis based on individual statements of radiological reports. The model achieves a sensitivity of 91% and PPV of 83% [32].

Last but not least, NLP is a helpful tool for **quality assessment of radiologic practice**. NLP tools covering this task identify specific quality indicators used for internal quality assurance, comparison to guidelines, and legal purpose.

For example, Lacson et al. used iSCOUT to select reports with pulmonary nodules and verify the concordance between node management and recommendations from the Fleischner Society Guidelines [33].

8.4 Structured Reporting as AI Annotation Strategy

Appropriate implementation of structured reporting is based on templates. Integrating the healthcare enterprise (IHE) initiative developed a standard for the presentation of structured reports through the working group on Management of Radiology Reporting Templates (MRRT). It specifies which technology should be used for template development and describes how these templates should be managed and integrated into radiology infor-

mation systems or PACS reporting orchestrators and their migration to these environments. In contrast, MRRT does not define how template-based reports are transmitted from a radiology information system or PACS to an electronic health record system.

Structured reports can also be stored in DICOM format since the current standard definition considers the “DICOM-SR” modality. In the standard, the guidelines to be followed in the DICOM-SR object creation and the encoding of the information contained are specified. Furthermore, DICOM-SR objects can also include the annotations (i.e., measurements, regions of interest, among others) performed by the radiologist using the tools available in a PACS workstation. Measurements and annotations provide meaningful information to complement the qualitative findings included in the report.

The combination of the HL7 standard with DICOM-SR enriches the report with clinical information relevant to patient diagnosis through the images obtained.

Structured reporting enables the development of deep learning algorithms thanks to the seamless annotation performed while reporting. Annotation is mainly performed today from retrospective data by NLP techniques, as seen in previous sections. Nevertheless, a risk to generate inaccuracies and uncertainties not only in annotation but also in the creation of deep learning models has already been in the case of the CheXNet paper [34, 35].

As an improved and scalable annotating methodology, research experiences have already demonstrated the feasibility of using the data from structured reports completed in clinical routine for training deep learning algorithms, highlighting the potential of structured reporting for the future of radiology in the context of AI and deep learning as the main technique applied [36–38].

8.5 Quantitative Structured Reporting

Structured reports can also be the way in which AI algorithms and image analysis results are communicated and integrated into hospital information systems such as the PACS, RIS, or EHR.

Quantitative features are today being generated in the form of imaging biomarkers by applying computational algorithms to the analysis of medical images. Computational imaging algorithms can either be based on AI (driven by data) or on conventional computer vision algorithms (driven by model). The main aim of quantitative imaging biomarkers is to early detect disease before symptoms, to establish a diagnosis and staging if the disease and symptoms are already present, to predict patient outcomes, and to evaluate treatment response during follow-up.

The extracted imaging biomarkers provide quantitative information on their spatial distribution (parametric images) and their magnitude (intensity). The textural analysis of signal intensity properties from different voxels in a region of interest, through the extraction of quantitative features, allows for the evaluation of first-order histogram characteristics (intensity, skewness, kurtosis) and second-order parameters (energy, information, correlation, among many others). The process of extracting hundreds or thousands of these features and using AI-based classifiers whose output is a clinical endpoint is called radiomics.

With regard to radiological workflow integration, even if these imaging biomarkers and radiomics capabilities may be available in a research or academic domain, their integration within radiology information systems such as the RIS and PACS is still not straightforward. As an example, we can obtain the percentage of the affected lung in the computed tomography images of a COVID-19 patient, but current systems will not allow integrating this value seamlessly in the radiology report (without manually typing it) or performing population-based queries such as “show me all cases analyzed during the last year with a % of affected lung higher than 20%.”

The final results of AI and imaging biomarker extraction algorithms must be available in the radiology structured reporting environment in order for the radiologists to be able to accept, amend, or reject this information.

As of now, quantitative structured reports can be generated in a parallel streamline that allows integrating final reports as an annex to the conventional radiology reporting. These quantitative reports can be generated by the use of HTML or Jade templates that are

installed in an environment or ecosystem hosting different applications and orchestrating AI analysis in the radiology routine. An example of the quantitative structured report obtained from the application of convolutional neural networks (CNNs) for the detection of ground glass opacities and the quantification of lung damage can be appreciated in Fig. 8.1.



Fig. 8.1 Quantitative structured report generated from an AI pipeline that calculates the percentage of the affected lung by COVID-19 opacities and the probability of being a positive case

References

1. European Society of Radiology (ESR). ESR paper on structured reporting in radiology. *Insights. Imaging.* 2018;9(1):1–7. <https://doi.org/10.1007/s13244-017-0588-8>. Epub 2018 Feb 19.
2. Institute of Medicine (US) Committee on Data Standards for Patient Safety. *Key Capabilities of an Electronic Health Record System: Letter Report.* Washington, DC: National Academies Press (US); 2003. <https://www.ncbi.nlm.nih.gov/books/NBK221802/>.
3. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, Rybicki FJ, Mitsouras D. Natural language processing technologies in radiology research and clinical applications. *Radiographics.* 2016;36(1):176–91. <https://doi.org/10.1148/rg.2016150080>.
4. Kok JN, Boers EJW, Kusters WA, van der Putten P, Poel M. Artificial intelligence: definition, trends, techniques and cases. *Encyclopedia of Life Support Systems (EOLSS).* UNESCO-EOLSS.
5. Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. *Radiology.* 2016;279(2):329–43. <https://doi.org/10.1148/radiol.16142770>.
6. Martí-Bonmatí L, Ruiz-Martínez E, Ten A, Alberich-Bayarri A. How to integrate quantitative information into imaging reports for oncologic patients. *Radiologia.* 2018;60(Suppl 1):43–52. <https://doi.org/10.1016/j.rx.2018.02.005>. Epub 2018 Mar 28.
7. Montoliu-Fornas G, Martí-Bonmatí L. Magnetic resonance imaging structured reporting in infertility. *Fertil Steril.* 2016;105(6):1421–31. <https://doi.org/10.1016/j.fertnstert.2016.04.005>. Epub 2016 Apr 19.
8. Medina García R, Torres Serrano E, Segrelles Quilis JD, Blanquer Espert I, Martí Bonmatí L, Almenar Cubells D. A systematic approach for using DICOM structured reports in clinical processes: focus on breast cancer. *J Digit Imaging.* 2015;28(2):132–45. <https://doi.org/10.1007/s10278-014-9728-6>.
9. Segrelles JD, Medina R, Blanquer I, Martí-Bonmatí L. Increasing the efficiency on producing radiology reports for breast cancer diagnosis by means of structured reports. A comparative study. *Methods Inf Med.* 2017;56(3):248–60. <https://doi.org/10.3414/ME16-01-0091>. Epub 2017 Feb 21.
10. Kao A, Poteet S. Overview. In: Kao A, Poteet S, editors. *Natural language processing and text mining.* New York, NY: Springer; 2007. p. 1–7.
11. National Library of Medicine (U.S.). *UMLS reference Manual, Chapter 1, Introduction to the UMLS.* 2009. <http://www.ncbi.nlm.nih.gov/books/NBK9675/>. Accessed 20 Mar 2021.
12. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* 2001;34(5):301–10.

13. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc.* 1994;1(2):161–74.
14. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004;11(5):392–402.
15. Yu S, Kumamaru KK, George E, et al. Classification of CT pulmonary angiography reports by presence, chronicity, and location of pulmonary embolism with natural language processing. *J Biomed Inform.* 2014;52:386–93.
16. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell.* 2020;181(1):92–101. <https://doi.org/10.1016/j.cell.2020.03.022>.
17. Uzuner O, Zhang X, Sibanda T. Machine learning and rule-based approaches to assertion classification. *J Am Med Inform Assoc.* 2009;16(1):109–15.
18. Elkin PL, Froehling D, Wahner-Roedler D, et al. NLP-based identification of pneumonia cases from free-text radiological reports. *AMIA Annu Symp Proc.* 2008;2008:172–6.
19. Cheng LT, Zheng J, Savova GK, Erickson BJ. Discerning tumor status from unstructured MRI reports—completeness of information in existing reports and utility of automated natural language processing. *J Digit Imaging.* 2010;23(2):119–32. <https://doi.org/10.1007/s10278-009-9215-7>. Epub 2009 May 30.
20. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology.* 2019;290:590–606.
21. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *Radiographics.* 2017;37:2113–31.
22. Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. *Front Plant Sci.* 2020;10:1750. <https://doi.org/10.3389/fpls.2019.01750>.
23. Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:14085882. 2014. <https://arxiv.org/abs/1408.5882>. Accessed Mar 2021.
24. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9:1735–80.
25. Ruder S. NLP's ImageNet moment has arrived. 2018. <https://thegradient.pub/nlp-imagenet/>. Accessed Mar 2021.
26. Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. *Infect Control Hosp Epidemiol.* 1998;19(2):94–100.
27. Hripcsak G, Austin JH, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology.* 2002;224(1):157–63.

28. Chapman BE, Lee S, Kang HP, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. *J Biomed Inform.* 2011;44(5):728–37. <https://doi.org/10.1016/j.jbi.2011.03.011>. Epub 2011 Apr 1.
29. Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports. *Radiology.* 2018;286:845–52.
30. Miao S, Xu T, Wu Y, Xie H, Wang J, Jing S, Zhang Y, Zhang X, Yang Y, Zhang X, Shan T, Wang L, Xu H, Wang S, Liu Y. Extraction of BI-RADS findings from breast ultrasound reports in Chinese using deep learning approaches. *Int J Med Inform.* 2018;119:17–21. <https://doi.org/10.1016/j.ijmedinf.2018.08.009>. Epub 2018 Aug 18.
31. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH. Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson’s natural language processing algorithm. *J Digit Imaging.* 2018;31:245–51.
32. Rink B, Roberts K, Harabagiu S, et al. Extracting actionable findings of appendicitis from radiology reports using natural language processing. *AMIA Jt Summits Transl Sci Proc.* 2013;2013:221.
33. Lacson R, Prevedello LM, Andriole KP, et al. Factors associated with radiologists’ adherence to Fleischner Society guidelines for management of pulmonary nodules. *J Am Coll Radiol.* 2012;9(7):468–73.
34. Rajpurkar P, Irvin J, Zhu K et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. 2017. <http://arxiv.org/abs/1711.05225v3>. Accessed 30 Mar 2021.
35. Oakden-Rayner L. CheXNet: an in-depth review. 2018. <https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/>. Accessed 30 Mar 2021.
36. Pinto dos Santos D. The value of structured reporting for AI: opportunities, applications and risks. 2019. https://doi.org/10.1007/978-3-319-94878-2_7.
37. Bizzo BC, Almeida RR, Alkasab TK. Computer-assisted reporting and decision support in standardized radiology reporting for cancer imaging. *JCO Clin Cancer Inform.* 2021;5:426–34. <https://doi.org/10.1200/CCI.20.00129>.
38. Goldberg-Stein S, Chernyak V. Adding value in radiology reporting. *J Am Coll Radiol.* 2019;16(9 Pt B):1292–8. <https://doi.org/10.1016/j.jacr.2019.05.042>.