



## Natural language processing to convert unstructured COVID-19 chest-CT reports into structured reports

Salvatore Claudio Fanni<sup>a</sup>, Chiara Romei<sup>b,\*</sup>, Giovanni Ferrando<sup>c</sup>, Federica Volpi<sup>a</sup>, Caterina Aida D'Amore<sup>a</sup>, Claudio Bedini<sup>d</sup>, Sandro Ubbiali<sup>c</sup>, Salvatore Valentino<sup>c</sup>, Emanuele Neri<sup>a</sup>

<sup>a</sup> Department of Translational Research, Academic Radiology, University of Pisa, Pisa, Italy

<sup>b</sup> Department of Diagnostic Imaging, 2nd Radiology Unit, Pisa University-Hospital, Pisa, Italy

<sup>c</sup> EBIT sr.l. Esaote Group, Via di Caciolle, Florence, Italy

<sup>d</sup> EBIT sr.l. Esaote Group, Via Melen 77, Genoa, Italy

### ARTICLE INFO

#### Keywords:

Natural language processing  
Structured reporting  
COVID-19  
Artificial intelligence  
Deep learning

### ABSTRACT

**Background:** Structured reporting has been demonstrated to increase report completeness and to reduce error rate, also enabling data mining of radiological reports. Still, structured reporting is perceived by radiologists as a fragmented reporting style, limiting their freedom of expression.

**Purpose:** A deep learning-based natural language processing method was developed to automatically convert unstructured COVID-19 chest CT reports into structured reports.

**Methods:** Two hundred-two COVID-19 chest CT were retrospectively reviewed by two experienced radiologists, who wrote for each exam a free-form text radiological report and coherently filled the template provided by the Italian Society of Medical and Interventional Radiology, used as ground-truth. A semi-supervised convolutional neural network was implemented to extract 62 categorical variables from the report. Two iterations were carried-out, the first without fine-tuning, the second one performing a fine-tuning. The performance was measured using the mean accuracy and the F1 mean score. An error analysis was performed to identify errors entirely attributable to incorrect processing of the model.

**Results:** The algorithm achieved a mean accuracy of 93.7% and an F1 score 93.8% in the first iteration. Most of the errors were exclusively attributable to wrong inference (46%). In the second iteration the model achieved for both parameters 95,8% and percentage of errors attributable to wrong inference decreased to 26%.

**Conclusions:** The convolutional neural network achieved an optimal performance in the automated conversion of free-form text into structured radiological reports, overcoming all the limitation attributed to structured reporting and finally paving the way for data mining of radiological report.

### 1. Introduction

Structured report is defined as a radiological report using pre-defined formats, based on templates or checklists, and standardized terms [1]. Structured reporting could integrate other information than radiological ones, such as clinical information, technical parameters, measurements, annotations, and images of key findings [2]. Structured reporting has many advantages compared to free-form text reporting, such as the use

of a systematic checklist helpful in error rate reduction, the increase of report completeness, a reduced reporting time and a higher quality perceived by the referring physicians [3,4].

These advantages also emerged in one of the first surveys concerning structured reporting of COVID-19 pneumonia showing that the referring physicians strongly preferred structured reporting over free-form text reporting, due to the reduction of omissions of relevant imaging findings [5].

**Abbreviations:** SIRM, Italian Society of Medical and Interventional Radiology; CNN, Convolutional neural network; PPV, Positive predictive value; GGO, Ground-glass opacities; ESR, European Society of Radiology; RSNA, Radiological Society of North America; NLP, natural language processing; AI, Artificial Intelligence; ML, Machine-learning; DL, Deep-learning; RIS, Radiological Information System; MTM, Machine Translation Model.

\* Correspondence to: Via Paradisa 2, Pisa, Italy.

E-mail address: [chiara.romei@gmail.com](mailto:chiara.romei@gmail.com) (C. Romei).

<https://doi.org/10.1016/j.ejro.2023.100512>

Received 27 May 2023; Received in revised form 17 July 2023; Accepted 18 July 2023

2352-0477/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Indeed, CT findings in COVID-19 patients are not specific, often resembling those associated with other infections. The lung abnormalities created a complex radiological pattern, and it is often challenging for the radiologist to report all lung abnormalities, even more in urgency clinical setting and to write consistent and clear conclusions in the free-form text radiological report [6].

The most typical findings are ground-glass opacities (GGO), interlobular septal thickening, the association between the latter two, the so-called “crazy paving” pattern, and consolidations. These patterns may show a focal, multifocal, or diffuse distribution, typically bilateral, with a peripheral or subpleural prevalence [7]. Other manifestations less frequently observed on CT are bronchiectasis, reverse halo-sign, pleuro-pericardial effusion, lymphadenopathy, and pneumothorax [8].

Based on these reasons, structured reporting together with a standardized language, should be implemented in imaging of COVID-19 pneumonia. This need prompted international radiological societies to develop and publish templates for structured reporting of chest CT in COVID-19 pneumonia [6,9].

Despite efforts in favor of structured reporting promoted by the European Society of Radiology (ESR) and the Radiological Society of North America (RSNA) the majority of radiologists prefer free-form text radiological reports [10]. Indeed, structured reporting also has some limitations, such as the perception by reluctant radiologists of a fragmented approach to reporting, a limitation to their freedom of expression, and the learning curve associated with the new reporting style or the time necessary to fill the structured report [11]. A further issue to address is the “inhabiting eye”, that happen when radiologists focus too much attention on the reporting system than on the images. Each moment the radiologist looks away from the images, time is lost to visually and cognitively refocus and there is an increased risk of errors [12,13].

Therefore, a way to shorten the distance between structured reporting and the radiologists and to tackle these issues, must be identified. Natural Language Processing (NLP) might be a solution: NLP is a sub-field of computer science including any computer-based methods able to convert free-form text into mineable structured data through the combination of linguistic, statistical and Artificial Intelligence (AI) methods, such as machine-learning (ML) and deep-learning (DL) [14]. NLP could be used to automatically convert free-form text radiological report into structured report, leaving the radiologist free to dictate the report as they are used to, overcoming most of the issues above-mentioned, such as the learning curve required or the “inhabiting eye”.

This is a feasibility study aimed at implementing a DL-based NLP method to automatically convert unstructured COVID-19 chest CT radiological reports into structured reports.

## 2. Material and methods

### 2.1. Case selection

Patients affected by COVID 19 and hospitalized in Pisa University Hospital between March 1st and April 9th, 2020, were retrospectively enrolled in this study.

Inclusion criteria were a non-enhanced chest CT and a positive RT-PCR for Sars-Cov 2 tested of nasopharyngeal swab no longer than a week after the CT acquisition. Patients with Severe motion artifacts in the CT scan and unavailable RT-PCR outcome or date were excluded. The final cohort included 202 patients. All the procedures, from the collection to the amplification of mRNA, were performed in full compliance with the WHO Guidelines [15]. The internal review board approved the research with protocol number 51834.

### 2.2. Image acquisition

Non-enhanced Chest CT were performed during a single full

inspiratory breath-hold in a supine position. Two different CT scanner were used to perform all the scans, respectively a 64-slice General Electric Light Speed scanner (General Electric co) and a 40-slice Siemens Somatom Sensation scanner (Siemens Healthineers). The acquisition parameters are summarized in Table 1.

### 2.3. Structured report template

The structured report template adopted in the study was developed by members of the college of thoracic radiology and imaging informatics of the Italian Society of Medical and Interventional Radiology (SIRM) [6] and it is freely accessible through the RadReport portal (www.radreport.org), created by the RSNA. The structured report template is divided into the following sections headings: *procedure information, clinical information, previous exams, parenchyma: ground glass opacities, consolidation and nodules, mediastinum, vascular findings, barotrauma, impression and classification.*

### 2.4. Dataset

Once developed, the template has been integrated into our department’s radiological information system (RIS), which allowed the radiologist to easily fill the structured report and to dictate a free-text report. All the 202 CTs have been revised in consensus by two experienced radiologists (C.R. and E.N.). The authors dictated a new free-form text radiological report and coherently filled the structured report. The free-form text radiological reports have been used as input, whereas the structured reports have been used as ground truth.

### 2.5. Model architecture

The extraction of data from a text written in natural language can be framed as a text classification problem, that is the determination of whether a given characteristic (e.g., the presence of fever) is contained in the text. A traditional approach would have required many manually classified examples to achieve acceptable performance. It would have been necessary to train the specific classification models for each information that should be extracted from the text, using a long short-term memory architecture [13].

Recently, together with more efficient architectures based on text comprehension, a system for the development of natural language models based on pre-training has emerged [16,17]. This pre-training step is particularly powerful because it allows the model to learn on a non-manually classified dataset.

It has been shown that more advanced models, such as convolutional neural network, are able to perform extraction tasks without ever having performed a specific training (zero-shot learning) or with very few classification data (few-shot learning) [18].

Question Answering (QA) neural language model was adopted in this study, as it gives the capability to work in zero shot or few shots mode and allows the user to easily configure the model for the extraction of a new field [18]. Given the limited number of reports available, a pre-trained model based on Text-to-Text Transfer Transformer architecture was chosen [19].

In particular, as a trade-off between effectiveness and generalizability, the chosen model was based on UnifiedQA, developed in 2020 by Allen Institute for AI [18], that is a parameterized sequence-to-sequence model for QA [20]. The training scripts and trained weights are open sourced with commercial use licenses. The model casts multiple QA problems (Yes/No, Multiple choice, extractive and abstractive) into a text-to-text paradigm with improved generalization capabilities.

The processing input consists of a medical report containing information attributable to a structure consisting of 62 categorical or Boolean values, which are summarized in Table 2.

The architecture of the solution foresees that the textual data is pre-

**Table 1**

Acquisition parameters of non-enhanced chest CT.

Scanner	Number of CT scans	Kv	mAs	Spiral pitch factor	Collimation width	Matrix	Kernel
General Electric Light Speed scanner	112	120	169	0.98	0.625	512 × 512	Standard
Siemens Somatom Sensation scanner	90	120	284	1.84	0.6	512 × 512	B31

**Table 2**

List of the 62 categorical or Boolean values.

Procedure Information	
1. Thickness (mm)	4. Patient decubitus
2. Reconstruction algorithm	5. Contrast medium
3. Kernel	
Clinical information and anamnesis	
6. Positive RT-PCR test	17. Symptoms (nausea)
7. Severity of symptoms	18. Symptoms (vomit)
8. Time onset of symptoms	19. Symptoms (diarrhea)
9. Symptom (fever)	20. Symptoms (pharyngodynia)
10. Symptoms (cough)	21. Symptoms (anosmia and dysgeusia)
11. Symptoms (rhinitis)	22. Comorbidities/risk factor (hypertension)
12. Symptoms (dyspnea)	23. Comorbidities/risk factor (cardiopathy)
13. Symptoms (myalgia)	24. Comorbidities/risk factor (interstitial lung disease)
14. Symptoms (asthenia)	25. Comorbidities/risk factor (neoplastic lung disease)
15. Symptoms (conjunctivitis)	26. Comorbidities/risk factor (liver function test)
16. Symptoms (headache)	
Parenchyma – Ground glass	
27. Location	29. Craniocaudal distribution
28. Axial distribution	
Parenchyma - Consolidation	
30. Location	33. Solid
31. Axial distribution	34. Subsolid
32. Craniocaudal distribution	35. Cavitated
Parenchyma – Other findings	
36. Septal thickness	40. Fibrotic distortion
37. Crazy paving	41. Pulmonary emphysema
38. Reversed halo sign	42. Perilobular sign
39. Subpleural sparing	
Parenchyma - Nodules	
43. Nodules > 3 mm diameter ≤ 30 mm	48. Tree in bud
44. Micronodules diameter ≤ 3 mm	49. Centrilobular
	50. Perilymphatic
	51. Random
45. Solid	
46. Subsolid	
47. Cavitated	
Mediastinum – Lymph nodes	
52. Right hilar	54. Subcarinal
53. Left hilar	55. Antero-superior mediastinum
Vascular findings	
56. Pulmonary artery trunk diameter	57. Subsegmental peripheral vessel dilation
Barotrauma	
58. Pneumothorax	60. Pneumomediastinum
59. Subcutaneous emphysema	
Impressions	
61. Classification	62. CO-RADS category

processed before it is submitted to the UnifiedQA neural network. Indeed, the UnifiedQA can perform QA only on English paragraphs of moderate size (< 500 characters). Thus, a top-scoring Machine Translation Model (MTM) was necessary to convert reports from Italian to English and, considering that reports have fields that are highly reliant on the context, a context filtering step has been performed to ensure that UnifiedQA receives only the relevant info.

For each structured field to be predicted, a filtering of the relevant phrases is done. This filtering is the only supervised operation in the pipeline, and it is based on manually created regular expressions.

Below, an example of this step showing in grey the recognized *procedure information* section of the radiological report and in red the identified categorical variables (Patient Position, Reconstruction algorithm, Thickness).

Non-enhanced chest CT was obtained with high resolution technique and layer thickness of 1.5 mm for patient in supine position with standard reconstruction algorithm. The patient has had fever for 7–14 days.

After the translation and the context identification, the UnifiedQA system answers a predefined question associated with the category. The output is finally mapped to a dictionary of possible values for the category itself. The final architecture of the solution is resumed in Fig. 1.

## 2.6. Algorithm iterations

Two iterations were carried out. In the first one the UnifiedQA neural network was implemented without carrying out any specific fine-tuning operation on it. The validation data consisted of 202 unstructured free-form text radiological reports. After performing the translation into English and filtering the relevant passages, the UnifiedQA system uses the context to answer a predefined question associated with the category. The UnifiedQA output is finally mapped to a dictionary of possible values for the category.

The same dataset was used to carry out a second iteration performing a fine-tuning of the QA model.

The context filtering step has been removed and the QA model was trained on the entire paragraph. The 202 cases were randomly partitioned in training plus validation (90%) and test set (10%) aiming to ensure the widest variability between these datasets.

Within the 202 available chest CT Covid radiological reports:

- 167 were used as training data.
- 15 were used in validation to evaluate the quality of the network output in the training phase.
- 20 were used for testing purposes.

For both the iterations, the system has been evaluated using two metrics:

- the average accuracy on the set of fields evaluated, for the set of reports available (tot. 14,000 values).
- the average F1 score, corresponding to the harmonic average of sensitivity and positive predictive value.

To overcome the problem of information not explicitly mentioned in the medical report, the system has produced a default negative value if the information sought is not included in the medical report itself. Additionally, an analysis of the causes of errors was carried out for both the iterations.

## 3. Results

In the first iteration, the performance obtained was about 93.7% for average accuracy and 93.8% for average F1 score. Detailed results as function of the accuracy are summarized in Table 3.

The performance was lower than-average for some variables (e.g., layer thickness, axial distribution for ground glass and consolidations).

From the analysis of the cause of errors, it turns out that almost half (46%) were related to an incorrect processing of the inference engine, e.g.:

For: Procedure Information - Thickness (mm) (Reported value “1.25”)

“The findings are associated with minimal septal thickening”

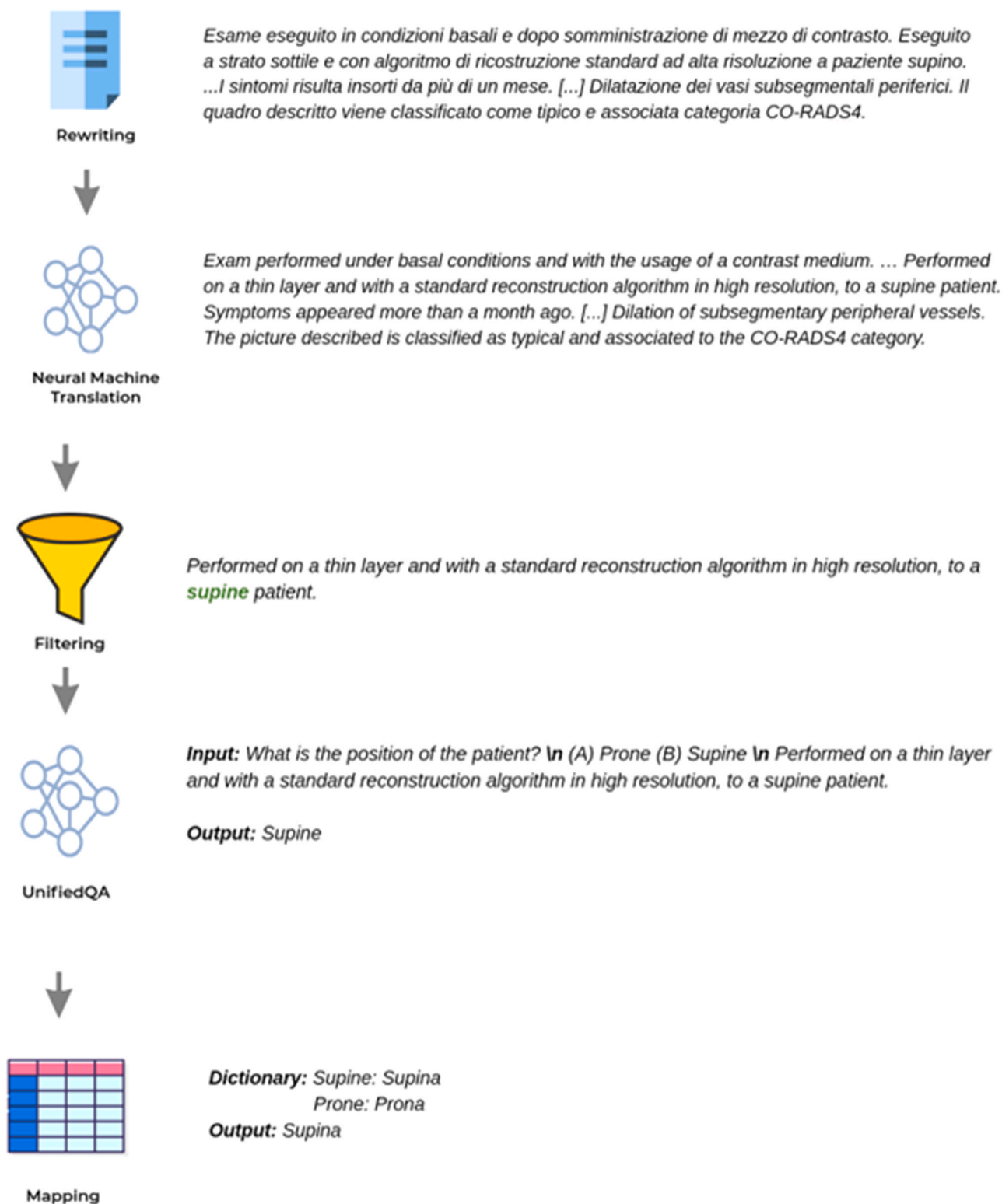


Fig. 1. Final architecture of the solution.

The septal thickening was misclassified as CT thickness.

Another source of error (14%) was the ambiguity in the formulation of the contents. These errors were related to the omission by the radiologists of pivotal information for the machine to properly contextualize the sentences and the terms.

For: Parenchyma – Nodules - Solid (Reported values “True”)

“A few sporadic, nonspecific, parenchymal pulmonary nodules and micronodules”

The information was not explicitly and clearly stated in the radiological report.

No errors were attributed to the pre-processing translation step performed by MTM.

**Table 3**  
Distribution of each variable and correspondent accuracy in the first and the second iteration.

Subheading	Variable	N (%)	Accuracy (%)	
			First Iteration	Second Iteration
Procedure information	Thickness (mm)	201 (99)	66	95
	Reconstruction algorithm	196 (97)	100	100
	Kernel	195 (97)	89	92
	Patient Position	194 (96)	100	100
Clinical information and anamnesis	Contrast medium	105 (52)	94	100
	Positive RT-PCR test	50 (25)	88	100
	Severity of symptoms	17 (8)	96	92
	Time onset of symptoms	107 (53)	84	100
	Symptoms (fever)	124 (61)	97	100
	Symptoms (cough)	71 (35)	97	100
	Symptoms (rhinitis)	2 (1)	100	100
	Symptoms (dyspnea)	72 (36)	97	100
	Symptoms (myalgia)	12 (6)	99	100
	Symptoms (asthenia)	14 (7)	99	100
	Symptoms (conjunctivitis)	2 (1)	100	100
	Symptoms (headache)	4 (2)	100	100
	Symptoms (nausea)	6 (3)	100	100
	Symptoms (vomit)	6 (3)	100	95
	Symptoms (diarrhea)	10 (5)	100	100
	Symptoms (pharyngodynia)	2 (1)	100	100
	Symptoms (anosmia e dysgeusia)	5 (2)	99	100
	Comorbidities/RF (hypertension)	6 (3)	98	100
	Comorbidities/RF (cardiopathy)	7 (3)	99	100
	Comorbidities/RF (interstitial lung disease)	2 (1)	99	100
Comorbidities/RF (neoplastic lung disease)	4 (2)	98	100	
Comorbidities/FR (liver function test)	1 (1)	100	100	
Parenchyma - Ground Glass	Location	187 (93)	86	100
	Axial distribution	184 (91)	63	70
Parenchyma - Consolidations	Craniocaudal distribution	184 (91)	98	100
	Location	158 (78)	79	85
	Axial distribution	151 (75)	68	65
	Craniocaudal distribution	154 (76)	100	100
	Solid	84 (42)	73	84
	Subsolid	58 (29)	85	83
Parenchyma - Other findings	Cavitated	6 (3)	100	100
	Septal thickness	165 (82)	80	89
	Crazy paving	121 (60)	98	94
	Reversed Halo sign	26 (13)	100	100

**Table 3 (continued)**

Subheading	Variable	N (%)	Accuracy (%)		
			First Iteration	Second Iteration	
Parenchyma - Nodules	Subpleural sparing	58 (29)	85	100	
	Fibrotic distortion	59 (29)	96	100	
	Pulmonary emphysema	39 (19)	96	95	
	Perilobular sign	128 (63)	91	90	
	Nodules < 3 mm diameter < = 30 mm	38 (19)	95	95	
	Nodules diameter < = 3 mm	34 (17)	94	95	
	Solid	84 (42)	96	92	
	Subsolid	58 (29)	97	100	
	Cavitated	6 (3)	100	100	
	Tree in bud	9 (4)	100	100	
Mediastinum - Lymph nodes	Centrilobular	7 (3)	100	100	
	Perilymphatic	6 (3)	100	100	
	Random	7 (3)	99	100	
	Right hilar	98 (49)	87	85	
	Left hilar	77 (38)	93	94	
	Subcarinal	125 (62)	84	90	
	Antero-superior mediastinum	134 (66)	84	73	
	Vascular findings	Pulmonary artery trunk diameter	104 (51)	86	90
		Subsegmental peripheral vessel dilation	139 (69)	89	90
	Barotrauma	Pneumothorax	37 (18)	100	100
Pneumomediastinum		36 (18)	100	100	
Impressions	Subcutaneous emphysema	6 (3)	98	100	
	Classification	200 (99)	100	100	
CO-RADS Category		199 (99)	100	100	

The detailed errors distribution of the first iteration is showed in [Table 4](#).

When considering the second iteration, the performance improved obtaining about 95.8% for average accuracy and 95.8% for average F1 score with a percental gain of about the 2%. Detailed results as function of the accuracy are summarized in [Table 3](#). A detailed analysis of the errors was carried out and demonstrated a percentage gain of about 20% on errors due to the incorrect processing of the inference engine ([Table 5](#)).

**Table 4**  
Errors distribution in the first iteration.

Problem	Description	%
Wrong inference	Incorrect processing of the inference engine	~ 46%
Inconsistent report	The dictated data and the manually entered data did not match	~ 28%
Ambiguous	Ambiguous text in the report	~ 14%
Wrong context inference	Incorrect identification of the context based on regular expression rules	~12%



**Table 5**  
Errors distribution in the second iteration.

Problem	Description	%
Ambiguous	Ambiguous text in the report	~ 45%
Inconsistent report	The dictated data and the manually entered data did not match	~ 29%
Wrong inference	Incorrect processing of the inference engine	~ 26%

#### 4. Discussion

In our study, two iterations were carried out to automatically convert unstructured into structured radiological reports and both achieved satisfactory performance in terms of average accuracy and average F1 score. Even though inferior compared to the second iteration, the first one, carried out without fine-tuning, achieved an optimal result. This result paves the way for potential applications also in other clinical scenarios, as other viral pneumonias presenting a significant overlap of radiological findings with COVID-19 pneumonia [21].

In addition, the second iteration improved both accuracy and F1 score, and, most importantly, achieved this result through a low number of reports, considering that usually fine tuning requires large amount of data to obtain acceptable performance. It might be expected an improved performance as the number of reports used for training increases. We further performed an error analysis demonstrating a significant reduction in errors related to the model misinterpretation between the first and the second iterations. However, the majority of the errors, in both iterations, was related to ambiguity or inconsistency of free-form text radiological reporting, demonstrating once again the limits of this reporting style.

In literature various applications of NLP in radiology have been already described: the identification/classification of findings, of cases/cohort for research studies, of follow-up recommendation, the determination of imaging protocol, the quality assessment of radiologic practice, the diagnostic surveillance [14].

In 1998, Knirsch et al. compared an NLP tool based on linguistic analysis with experts review for the identification of chest x-ray reports suspect of tuberculosis, focusing on the presence/absence of six key elements. As result, the agreement was 89–92% [22].

Similarly, in 2012, Chapman et al. developed an NLP tool to classify radiological reports of CT pulmonary angiography based on the presence/absence, location, chronicity, and certainty of pulmonary embolism, and achieved good results in terms of both sensitivity and specificity [23].

Indeed, the automated conversion in a structured form of free-form text radiological report can be considered an extension of the task of identification of findings, multiplied by the number of items of the structured report. In 2019, Spandorfer et al. developed a deep-learning-based algorithm for the automated conversion of CT pulmonary angiography free-form text reports into structured reports. The authors trained a convolutional neural network using 475 manually structured report. The accuracy was evaluated on a test set of 400 CT, resulting in an amount of 4.157 statements. The per-statement accuracy was 91.6% and 95.9% depending on whether strict or modified criteria were used [24].

Similarly to this study, we adopted a deep-learning based algorithm and we reached a result of 95.8% both for mean accuracy and F1 score. Some differences have to be highlighted beyond the different clinical scenario, as the absence of a specific training in the first iteration, and the lower number of radiological reports also used in our study.

Another study by Jorg et al. investigated the potential role of NLP in facilitating the adoption of structured reporting in the radiological workflow [25]. The authors developed an algorithm able to automatically convert dictated into structured reports of Urolithiasis CTs and

achieved an F1 score of 0.90 in the test set.

Coherently to our results, the majority of the errors were related to ambiguous description of the radiological findings. These results further demonstrate the feasibility of the automated conversion unstructured free-form text into the structured report template using AI algorithm based on NLP techniques. This strategy may enable to combine the advantages of both free-form reporting (namely, increased radiologist productivity) and structured reporting (namely, improved communication with providers). Additionally, increased structuring allows for data mining applications and the algorithm could be retrospectively applied to other reports of COVID-19 pneumonia chest CT.

NLP was already determined by Cury et al. capable of performing syntactic analysis of a large number of chest CT free-form text radiological report to generate predictions of positivity for COVID-19 pneumonia [26]. The authors concluded that a so-developed surveillance algorithm has the potential to assist healthcare entities in developing strategies against COVID-19 or similar pandemics in the future. A similar forecasting could apply also to our strategy, empowered by the ability of providing structured information about 62 different findings of COVID-19.

Although the promising results, this study is not to be considered without limitations. The overall number of reports used for both training and testing is to be considered limited, especially the one used for testing, and these reports were generated by the consensus of only two different radiologists from the same department, therefore with low heterogeneity in individual reporting style and the lack of an external validation. Finally, as COVID-19 pneumonia CT patterns of the first wave are now rarely encountered, the inclusion of only patients from the first COVID-19 wave should be considered a limitation, and further studies are needed to confirm the model applicability to the subsequent waves.

In conclusion, we demonstrated the feasibility of a highly accurate AI-based conversion of radiological unstructured into structured reports. In addition to the advantage of providing more structured data for research/data mining applications, the implementation of NLP in the radiological workflow could also affect the radiological workflow. However, the entity of this effect will be explored with following studies addressing the acceptance of automated compared to conventional structured reporting, the reporting time or the reporting quality perceived by the referring physician.

#### 5. Conclusions

The UnifiedQA neural network achieved an optimal performance in the automated conversion of free-text COVID-19 chest CT reports into structured reports. This strategy may effectively enable to overcome the limitations of structured reporting and to carry out extensive data mining from radiological reports. Further studies are necessary to fully investigate the potential clinical impact of automated structured reporting and to externally validate the model.

#### Ethics approval

The review board of Azienda Ospedaliera Universitaria Pisana approved the research with protocol number 51834.

#### Funding

This research received no external funding

#### CRediT authorship contribution statement

**Bedini Claudio:** Software, Resources. **Ubbiali Sandro:** Software, Resources. **Valentino Salvatore:** Software, Resources. **Fanni Salvatore Claudio:** Writing – original draft, Resources, Methodology, Data curation. **Neri Emanuele:** Writing – review & editing, Visualization,

Validation, Supervision, Methodology, Data curation, Conceptualization. **Volpi Federica:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **D’Amore Caterina Aida:** Formal analysis, Data curation. **Romei Chiara:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Data curation, Conceptualization. **Ferrando Giovanni:** Writing – original draft, Software, Methodology, Formal analysis, Data curation.

### Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Giovanni Ferrando, Claudio Bedini, Sandro Ubbiali and Salvatore Valentino declare personal fees from Ebit s.r.l. Esaote group. The other authors of this manuscript declare no conflict of interest.

### References

- J.M. Nobel, E.M. Kok, S.G.F. Robben, Redefining the structure of structured reporting in radiology, *Insights Imaging* 11 (2020), <https://doi.org/10.1186/s13244-019-0831-6>.
- Reiner, B.I., Knight, N., Siegel, E.L., 2007, Radiology Reporting, Past, Present, and Future: The Radiologist’s Perspective. *Journal of the American College of Radiology* 4:313–319. <https://doi.org/10.1016/j.jacr.2007.01.015>.
- C.C. Barbisan, M.P. Andres, L.R. Torres, et al., Structured MRI reporting increases completeness of radiological reports and requesting physicians’ satisfaction in the diagnostic workup for pelvic endometriosis, *Abdom. Radiol.* 46 (2021) 3342–3353, <https://doi.org/10.1007/s00261-021-02966-4>.
- B.P. Ernst, F. Katzer, J. Künzel, et al., Impact of structured reporting on developing head and neck ultrasound skills, *BMC Med Educ.* (2019) 19, <https://doi.org/10.1186/s12909-019-1538-6>.
- A. Stanzione, A. Ponsiglione, R. Cuocolo, et al., Chest CT in COVID-19 patients: Structured vs conventional reporting, *Eur. J. Radio.* (2021) 138, <https://doi.org/10.1016/j.ejrad.2021.109621>.
- E. Neri, F. Coppola, A.R. Larici, et al., Structured reporting of chest CT in COVID-19 pneumonia: a consensus proposal, *Insights Imaging* (2020) 11, <https://doi.org/10.1186/s13244-020-00901-7>.
- C. Romei, Z. Falaschi, P.S.C. Danna, et al., Lung vessel volume evaluated with CALIPER software is an independent predictor of mortality in COVID-19 patients: a multicentric retrospective analysis, *Eur. Radio.* 32 (2022) 4314–4323, <https://doi.org/10.1007/s00330-021-08485-6>.
- V. Ojha, A. Mani, N.N. Pandey, et al., CT in coronavirus disease 2019 (COVID-19): a systematic review of chest CT findings in 4410 adult patients, *Eur. Radio.* 30 (2020) 6129–6138, <https://doi.org/10.1007/s00330-020-06975-7>.
- S. Simpson, F.U. Kay, S. Abbara, et al., Radiological society of North America expert consensus document on reporting chest CT findings related to COVID-19: Endorsed by the society of thoracic radiology, the American college of radiology, and RSNA, *Radio. Cardiothorac. Imaging* 2 (2020), <https://doi.org/10.1148/ryct.2020200152>.
- European Society of Radiology (ESR), ESR paper on structured reporting in radiology, *Insights Imaging* 9 (2018), <https://doi.org/10.1007/s13244-017-0588-8>.
- D. Buccicardi, A. Panunzio, L. Faggioni, F. Coppola, Structured reporting in radiology: an overview, *J. Radiol. Rev.* 7 (2021), <https://doi.org/10.23736/s2723-9284.20.00026-4>.
- R.B. Gunderman, L.R. McNeive, Is structured reporting the answer? *Radiology* 273 (2014) 7–9.
- S.C. Fanni, L. Colligiani, N. Spina, et al., Current knowledge of radiological structured reporting, *J. Radiol. Rev.* 9 (2022), <https://doi.org/10.23736/s2723-9284.22.00189-1>.
- Fanni S.C., Gabelloni M., Alberich-Bayarri A., Neri E. (2022) Structured Reporting and Artificial Intelligence. In: *Structured Reporting in Radiology*.
- W.H.O. Coronavirus disease (COVID-19) technical guidance: Laboratory testing for 2019-nCoV in humans, 2020.
- Vaswani A., Shazeer N., Parmar N., et al. (2017) Attention Is All You Need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. pp 6000–6010.
- Devlin J., Chang M.-W., Lee K., et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, pp 4171–4186.
- D. Khashabi, S. Min, T. Khot, et al., UnifiedQA: Crossing Format Boundaries With a Single QA System. *Findings of the Association for Computational Linguistics, EMNLP 2020*, 2020, pp. 1896–1907.
- R. Raffel, N. Shazeer, A. Roberts, et al., Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, *J. Mach. Learn. Res.* 21 (2020) 1–67.
- Sutskever Google I., Vinyals Google O., le Google Q. v (2014) Sequence to Sequence Learning with Neural Networks. In: *Advances in neural information processing systems*.
- M. Churrua, E. Martínez-Besteiro, F. Couñago, P. Landete, COVID-19 pneumonia: A review of typical radiological characteristics, *World J. Radio.* 13 (2021) 327–343, <https://doi.org/10.4329/wjr.v13.i10.327>.
- C.A. Knirsch, N.L. Jain, A. Pablos-Mendez, et al., Respiratory Isolation of Tuberculosis Patients Using Clinical Guidelines and an Automated Clinical Decision Support System, *Infect. Control Hosp. Epidemiol.* 19 (1998) 94–100, <https://doi.org/10.2307/30141996>.
- B.E. Chapman, S. Lee, H.P. Kang, W.W. Chapman, Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm, *J. Biomed. Inf.* 44 (2011) 728–737, <https://doi.org/10.1016/j.jbi.2011.03.011>.
- A. Spandorfer, C. Branch, P. Sharma, et al., Deep learning to convert unstructured CT pulmonary angiography reports into structured reports, *Eur. Radio. Exp.* (2019) 3, <https://doi.org/10.1186/s41747-019-0118-1>.
- T. Jorg, B. Kämpgen, D. Feiler, et al., Efficient structured reporting in radiology using an intelligent dialogue system based on speech recognition and natural language processing, *Insights Imaging* (2023) 14, <https://doi.org/10.1186/s13244-023-01392-y>.
- R.C. Cury, I. Megyeri, T. Lindsey, et al., Natural language processing and machine learning for detection of respiratory illness by chest ct imaging and tracking of covid-19 pandemic in the united states, *Radio. Cardiothorac. Imaging* 3 (2021), <https://doi.org/10.1148/ryct.2021200596>.