Are estimations of radiomic image markers dispensable due to recent deep learning findings?

Martin Obert

Affiliation: Dept of Radiology, University Hospital Giessen, Justus-Liebig University, Giessen, Germany.

Correspondence: Martin Obert, Dept of Radiology, University Hospital Giessen, Justus-Liebig University, Klinikstrasse 33, 35392 Giessen, Germany. E-mail: martin.obert@radiol.med.uni-giessen.de

In light of recent deep learning findings, the question arises whether estimations of radiomic markers may soon be dispensable. In this editorial I would, therefore, like to compare two computed tomography (CT) lung studies to discuss this question: a deep learning (DL) cancer screening analysis, recently published in *Nature Medicine*, and a sparse data radiomic marker sarcoidosis investigation, published in this issue of the *European Respiratory Journal*. In this context, I would like to touch on three particular aspects: When are radiomic image markers still helpful? What effect does the volume of available data have on the statistical method of choice? How dogmatic should statistical analysis be in clinical applications?

In their DL lung cancer screening study based on CT images, published in May 2019, ARDILA *et al.* [1] reported that DL software outperformed radiologists, with reductions of 11% in false-positives and 5% in false-negatives, when no prior CT imaging data was available from the same patient. As a software development strategy, the authors decided to use deep convolutional neural networks (DCNN) instead of radiomic image features (RIF) as a feature learning tool. They based this decision on recent findings from open computer vision competitions that showed that DCNN methods are superior to those using hand-engineered features [2, 3] and on studies by LIAO *et al.* [4], who won the Kaggle 2017 Data Science Bowl for the use of DCNN to identify pulmonary nodules, and by Walsh *et al.* [5], who used DL for a fibrotic lung classification task. Methodologically, ARDILA *et al.* [1] divided the data from 42 579 cases into three subgroups: 70% training, 15% evaluating and 15% testing. From a statistics point of view, this procedure was unquestionably the method of choice in this case, given that the volume of available data was large.

Despite the superior performance of DCNN methods in information extraction in comparison to RIFs or, sometimes, that of human specialists, DCNN methods cannot, however, be applied successfully when sparse data of only a few hundred images are available, as documented by VALDES *et al.* [6]. Although there are available remedies to this problem, like transfer learning or data augmentation, the degradation of performance when networks were trained with smaller datasets was obvious. In such classification contexts, radiomic image markers thus remain indispensable. Moreover, because images contain more information than current quantitative markers can extract, it is still reasonable to try to improve RIF quality [7].

The study conducted by RYAN *et al.* [8] study, published in this issue of the *European Respiratory Journal*, followed a classical radiomic feature study design. Its principal aim was the development of a radiomic biomarker for sarcoidosis treatment and progression. RIF markers were evaluated from CT images (skewness, kurtosis, fractal dimension, Geary's C, Moran's I) and lung function tests (forced expiratory volume at 1 s (FEV_1), forced vital capacity (FVC), and the ratio FEV_1/FVC). Control and sarcoidosis groups were retrospectively collected from two different studies (control group: Genetic epidemiology of COPD (COPDGene), n=108, published 2010; sarcoidosis group: Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS), n=79, published 2015). Sarcoidosis progression was monitored according to Scadding stages [9] in lung sarcoidosis [10–12]. Staging on the basis of radiomic feature investigation provided a level of objectivity and reproducibility that overcame the intra- and inter-rater variability. This finding was, indeed, a major benefit of the study. The authors also verified their hypothesis that the image markers correlated with clinical findings. This is meaningful, since such correlations do not necessarily exist, and, so far, the evidence for correlations was based on detailed verbal descriptions [9, 13]. Overall, the authors concluded that radiomic feature investigations enable an objective, numerical characterisation of pulmonary parenchymal abnormalities in patients with sarcoidosis. Because the study included a fairly small number of patients, reflective of the low incidence of lung sarcoidosis, RYAN *et al.* [8] did not split their data into test and re-test groups. In a study including only 79 sarcoidosis patients, with different Scadding stages, it would have been unreasonable to assume that two patient groups of similar size could be generated that would still represent the normal variability in sarcoidosis patients. Although, from a strictly statistical standpoint, a test and re-test design would have been better, under the sparse data circumstances, the study design can, nevertheless, be regarded as acceptable.

I think that these two studies nicely illustrate why one should be careful not to falsely compare different study designs with different tasks, which may potentially require different statistical analyses due to different volumes of available data. Although the studies by ARDILA *et al.* [1] and RYAN *et al.* [8] both extracted numerical information from images, the research topics and the volume of investigated data in the two studies differ. In big data screening classification tasks, it thus seems advisable to train DCNNs directly from images. When, on the other hand, the development of radiomic markers for an orphan disease with sparse data, like sarcoidosis, is the topic, RIF analysis still plays a key role and is, in my opinion, certainly not dispensable. Moreover, in clinical contexts, when diseases with sparse data are studied, re-test designs should perhaps not be treated too dogmatically.

References

1 Ardila D, Kiraly AP, Bharadwaj S, *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; 25: 954–961.

2 Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115: 211–252.

3 Lin TY, Maire M, Belongie S, *et al.* Microsoft COCO: Common Objects in Context. *Comput Vis ECCV* 2014; 2014: 740–755.

4 Liao F, Liang M, Li Z, *et al.* Evaluate the malignancy of pulmonary nodules using the 3D deep leaky noisy-or network. 2017; preprint [https://arxiv.org/abs/1711.08324].

5 Walsh SLF, Calandriello L, Silva M, *et al.* Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.

6 Valdes G, Interian Y. Comment on 'Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study'. *Phys Med Biol* 2018; 63: 068001.

7 Obert M, Kampschulte M, Limburg R, *et al.* Quantitative computed tomography applied to interstitial lung diseases. *Eur J Radiol* 2018; 100C: 99–107.

8 Ryan SM, Fingerlin TE, Mroz M, *et al.* Radiomic measures from chest high-resolution computed tomography associated with lung function in sarcoidosis. *Eur Respir J* 2019; 54: 1900371.

9 Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J* 1961; 5261: 1165–1172.

10 Baughman RP, Culver DA, Judson MA. A concise review of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2011; 183: 573–581.

11 Nunes H, Uzunhan Y, Gille T, *et al.* Imaging of sarcoidosis of the airways and lung parenchyma and correlation with lung function. *Eur Respir J* 2012; 40: 255–263.

12 Vis R, Malviya G, Signore A, *et al.* ^{99m}Tc-anti-TNF-α antibody for the imaging of disease activity in pulmonary sarcoidosis. *Eur Respir J* 2016; 47: 1198–1207.

13 Culver DA, Baughman RP. It's time to evolve from Scadding: phenotyping sarcoidosis. *Eur Respir J* 2018; 51: 1800050.