Three artificial intelligence data challenges based on CT and MRI

Nathalie Lassau, Imad Bousaid, Emilie Chouzenoux, Jean-Philippe Lamarque, Benoit Charmettant, Michael Azoulay, François Cotton, Antoine Khalil, Olivier Lucidarme, Frédéric Pigneur, et al.

Three artificial intelligence data challenges based on CT and MRI

N. Lassau[a,b,1] , I. Bousaid[c], E. Chouzenoux[d], JP. Lamarque[a], B. Charmettant[a], M. Azoulay[c], F. Cotton[e], A. Khalil[f], O. Lucidarme[g], F. Pigneur[h], Y. Benaceur[i], A. Sadate[i], M. Lederlin[j], F. Laurent[k], G. Chassagnon[l], O. Ernst[m], G. Ferreti[n], Y. Diascorn[o], PY. Brillet[p], M. Creze[q], L. Cassagnes[r], C. Caramella[a,b], A. Loubet[s], A. Dallongeville[t], N. Abassebay[u], M. Ohana[v], N. Banaste[w], M. Cadi[x], J. Behr[y], L. Boussel[z] , L. Fournier[aa], M. Zins[bb], JP. Beregi[cc], A. Luciani[h], A. Cotten[dd], JF. Meder[ee]

[a] Université Paris-Saclay, Inserm, CNRS, CEA, Laboratoire d'Imagerie Biomédicale Multimodale Paris-Saclay, 94800 Villejuif, France
[b] Departement of Imaging, Institut Gustave Roussy, 94800 Villejuif, France
[c] Direction de la Transformation Numérique et des Systèmes d'Information, Gustave Roussy, 94800 Villejuif, France
[d] CVN, Inria Saclay, 91190 Gif-sur-Yvette, France
[e] Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France
[f] Department of Radiology, Hôpital Bichat, 75018, Paris, France
[g] Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale (LIB), Service de Radiologie, APHP, 75013 Hôpital Pitié–Salpêtrière, Paris, France
[h] Department of Radiology, AP-HP, Groupe Henri Mondor-Albert Chenevier, 94010 Créteil, France
[i] Department of Radiology, CHU Nîmes, 30189 Nîmes, France
[j] Department of Radiology, CHU Rennes, 35033 Rennes, France
[k] Department of Radiology, CHU de Bordeaux, 1045 Bordeaux, France
[l] Department of Radiology, Cochin Hospital, Assistance Publique-Hopitaux de Paris, 75014 Paris, France and Université de Paris, Descartes Paris 5, 75006, Paris, France
[m] Department of Radiology, CHU de Lille, Hôpital Huriez, 59037 Lille, France
[n] Service de radiologie diagnostique et interventionnelle, CS 10217, 38043 Grenoble, France
[o] Department of Radiology, CHU de Nice, Hôpital Pasteur, 06000 Nice, France
[p] Université Paris 13, INSERM UMR 1272 Hypoxie et Poumon, Assistance Publique-Hopitaux de Paris, Hôpital Avicenne, Service de Radiologie, 93022 Bobigny, France
[q] Department of Radiology, Hôpital Bicêtre, 94270 Le Kremlin-Bicêtre, France
[r] Department of Radiology, CHU de Clermont-Ferrand, Hôpital Montpied, 63003 Clermont-Ferrand
[s] Department of Radiology, CHU Montpellier, 34295 Montpellier, France
[t] Department of Radiology, GHSPJ, 75014, Paris, France
[u] Department of Radiology, CH Douai, 59507 Douai, France
[v] Radiology Department, Nouvel Hôpital Civil, 1 place de l'Hôpital, 67000, Strasbourg, France
[w] Department of Radiology, Centre Leon Berard, 69008 Lyon, France
[x] Radiologie PARIS ouest, Clinique Hartmann 92200 Neuilly sur seine
[y] Department of Radiology, CHRU de Jean-Minjoz Besançon, 25030 Besançon, France
[z] Hospices Civils de Lyon, Université de Lyon, 69000 Lyon, France.
[aa] Department of Radiology, Université Paris Descartes, Hôpital Européen Georges-Pompidou, 75015 Paris, France
[bb] Department of Radiology, Groupe Hospitalier Paris Saint Joseph, 75014 Paris
[cc] Collège des enseignants de radiologie de France (CERF, French College of Radiology Teachers), 47, rue de la Colonie, 75013 Paris, France
[dd] Lille Regional University Hospital, Musculoskeletal Imaging Department, 59000 Lille, France
[ee] Department of Neuroradiology, Centre Hospitalier Sainte-Anne, 75014 Paris, France

[1] Corresponding author at: Department of Imaging, Institut Gustave Roussy, 114, rue Édouard-Vaillant, 94805 Villejuif, France.E-mail address: nathalie.lassau@gustaveroussy.fr (N. Lassau).

KEYWORDS

Artificial intelligence (AI); Machine learning; Deep learning; Magnetic resonance imaging (MRI); Computed Tomography (CT);

ABSTRACT

Purpose: The 2019 edition of the data challenge was organized by the French Society of Radiology (SFR) during the Journées Francophones de Radiologie with the aim to: (i) work on relevant problematics of public health (ii) build large multicentric and prospective databases and (iii) boost the French AI community around radiologists. In comparison to the 2018 edition a first objective was to increase the question's complexity by including 3D information and prognostic analysis. The second objective was to improve the database quality and quantity with more balance among classes and data from at least 1000 examinations per question.

Material and method: Relevant clinical questions were proposed by organ societies of the SFR. Their feasibility was assessed by experts in the field of AI. A dedicated platform was set up for inclusion centers to safely upload their anonymized examinations in compliance with European regulation. The quality of the database was checked by experts weekly with annotations performed by radiologists. Multidisciplinary teams competed between September 11^{th} and November 13^{th} 2019.

Results: Three questions were selected using different imaging and evaluation modalities, including: pulmonary nodule detection and classification from 3D CT, prediction of expanded disability status scale in multiple sclerosis using 3D MRIs and segmentation of muscular surface for sarcopenia estimation from 2D CT. A total of 4347 examinations were gathered of which only 6% were excluded. Three independent databases from 24 individual centers were created. A total of 143 participants was split into 20 multidisciplinary teams.

Conclusion: Three data challenges with over 1200 GDPR compliant, multicentric, 2D/3D CT and MRI databases were organized for 20 multidisciplinary teams.

INTRODUCTION

In recent years the number of available medical image banks has soared [1], allowing the rapid development of artificial intelligence algorithms capable of handling different tasks such as classification, detection, or segmentation in different modalities [2]. It has also fostered the emergence of many artificial intelligence (AI) challenges competitions [3, 4]. As AI is likely to deeply impact radiology practice [5] it will arguably be of great help to offload repetitive tasks such as organ segmentation and to grasp more useful and quantitative information from images allowing more time spent on solving complex clinical problems [6]. Liew et al. anticipated that with the rise of AI the practice of radiology will drastically transform and radiologists will have to learn how to work with many new collaborators (including AI researchers, engineers, data officers...)[7].

However many pitfalls are still in the way of an efficient implementation of AI in radiologists' routine practice. Many studies have pointed out that training a deep learning model on precisely annotated databases is an essential requirement for safe deployment of such applications in the medical context [8, 9]. The recent implementation of the General Data Protection Regulation (GDPR) legislation imposes certain restrictive but essential rules.

In this context, the first edition of the data challenge was organized during the Journées Francophones de Radiologie (JFR) in 2018; its motivations were to:

(i). Help to solve common and relevant issues of public health.
(ii). Stimulate interactions within the AI community.
(iii). Build qualitative databases compliant with the new EU regulations.

These objectives have been fully completed [10] and led to 6 publications with promising results in 5 challenges: meniscus tear detection on 2D MRI [11,12], thyroid cartilage classification on 2D CT [13], breast lesion characterisation on 2D MRI [14], liver lesion classification on 2D ultrasound [15] and kidney cortex segmentation using 2D CT [16].

The second edition of the data challenge took place during the JFR in 2019. The objective was to increase the level of difficulty of the challenges offered by including 3D information and prognostic questions. The aim was also to increase the quality of the datasets with at least 1000 examinations for each question. Radiologists were directly involved in the labelling and the segmentation of the different datasets which allowed teaching good practices and ensuring the quality of the information provided to the teams.

MATERIAL AND METHODS

CLINICAL QUESTIONS

The starting point for the challenge was to ask organ societies of the French Society of Radiology (SFR) to suggest relevant clinical questions that could be asked to participants. This process allows being sure of putting the benefit of the patient at the core of the work carried out by the teams, sending messages to the medical community on the hot topics to deal with and spread good clinical practices. To be kept in the challenge the questions had to meet four criteria. First, the clinical relevance was judged by organ society's experts. Second, a literature review was performed to ensure the originality of the questions (using Pubmed, Kaggle and Grand-Challenge websites). Third, the feasibility of the challenge was ensured by data-scientists (Inria Saclay, France). Finally an estimation of the number of examinations produced in France each year for each question was carried out by the organizing committee to ensure to get a sufficient number of examinations (more than 1000 per challenge). Special care was taken to define challenge questions raising different AI tasks such as classification, automatic segmentation, and regression.

SECURITY AND DATA PROTECTION

This challenge has been entirely designed following the General Data Protection Regulations (GDPR), as ensured by the French regulation office called Commission Nationale Informatique et Libertés (CNIL). This year, the SFR has taken on the role of Data Protection Officer for these three challenges. In each inclusion center, patients were provided a letter with complete description of the use of their examinations during and after the challenge. An ethical charter has been signed by members of the AI groups of the SFR and radiologists who uploaded the images. All images sent by the different centers had to be anonymized. Still, a second verification was done to remove all potentially identifying DICOM fields with a dedicated Python script.

Participating teams during the challenges only worked with anonymized, pseudonymized and processed images.

INCLUSION AND COMMUNICATION

Inclusions for the three challenges started on April 1st and ended August 30th 2019. Newsletters have been sent to radiology departments of French hospitals. A dedicated website was built used as a platform to upload images/annotations and check their quality for all challenges. It allowed sharing information such as inclusion criteria, challenge timeline and files such as tutorials for the segmentation software, “3D slicer”.

The infrastructure has been provided by the Department of digital transformation and information services (DTNSI) of Gustave Roussy: a platform to exchange examinations, a server for hosting the event's website, and the storage of all the datasets.

To maintain a sufficient inclusion rate to reach the 1000 examinations included per challenges, information newsletters were sent two times a week to participants by email. Also, a margin of 10% of additional examinations was requested to deal with poor quality examinations and damaged files.

Once received from the inclusion centers, the data was reviewed by an expert engineer from an external partner (EASYS) to ensure anonymisation and formatting. The quality was then checked by radiology experts from the organ societies. This year the data control and quality check was done in real time allowing giving rapid feedbacks.

TEAM GATHERING AND CHALLENGE

The challenge’s website was also used as a platform for teams to register and get information. The teams were asked to be multidisciplinary, including at least one radiologist, one engineering/data-scientist student, and a research lab and/or company. Different networks were used to gather such team (JFR and SFR for radiologists, French graduate schools, Universities, and life imaging networks for students and research labs).

The first batch of data was sent to the teams on September 11th, the second batch was sent a month later, on October 11th. Finally, the validation set was sent on Sunday, October 13th. The different teams had limited time (two hours for the pulmonary nodule and one hour for the other two challenges) to generate results on the validation set. After scores calculation and jury deliberations, the winning team for each challenge was announced on Monday, October 14th.

RESULTS

Three subjects were selected to constitute the 2019 data challenge: classification of pulmonary nodule on 3D CT, prediction of expanded disability status scale (EDSS) score in multiple sclerosis on 3D MRI, and estimation of abdominal sarcopenia on 2D CT.

Pulmonary nodule (CT)

The risk of malignancy of a pulmonary nodule rises with the increase in its diameter. Screening studies using semi-automated volume measurements have shown higher accuracy and reproducibility compared to diameter measurements, and it has been shown that small nodules (those with a volume <100 mm^3 or diameter <5 mm) are not predictive for lung cancer [17]. For this question, teams had different tasks: to recognize pulmonary nodule on 3D CT chest scanners, to segment them, to estimate their volume and to classify them into probable benign (volume < 100mm^3) or probably malignant nodules (volume ≥ 100 mm^3).

Neurological impairment in multiple sclerosis (MRI)

The EDSS is used in the clinical evaluation of multiple sclerosis to rate neurological impairment. Magnetic resonance imaging (MRI) has become essential in the diagnosis and disease monitoring of patients with multiple sclerosis (MS). Significant associations have been shown between T2 lesion measures and EDSS measures [18], and that automated lesion volume quantification can be applied reliably on 3D fluid-attenuated inversion recovery (FLAIR) data sets [19]. For this question, candidates had to develop a two-year predictive EDSS-score algorithm based on brain MRI of known multiples sclerosis patients. The training set included 3D FLAIR or axial T2 FLAIR brain MRI and the corresponding clinical data (patients' age, sex, EDSS-score at two years from the images).

Sarcopenia (CT)

It has been shown that assessing muscular body surface on a single axial CT slice at the height of the third lumbar vertebra was highly correlated with clinically valuable parameters of body composition [20]. For this question, teams had two tasks: to segment muscular body mass and to estimate its surface. The dataset was composed of single axial 2D slices from CT abdominal scanners. The images and segmentation received from the different centers were checked by an expert radiologists and divided into 4 classes from A to D based on the quality of the segmentation, A being the best. Only classes from A to C were kept in the dataset communicated to the teams. The training set included the 2D slices in digital imaging and communications in medicine (DICOM) format, its segmentation masks and its surface in mm^2. For the test set on which teams were evaluated only the images were provided. Participants should return a binary predicted mask with the same size and format and the estimation of the surface of interest in mm^2.

For each question, inclusion criteria were different (table 1).

A total of 4262 examinations were uploaded from 24 different inclusion centers (table 2). The 3 challenges gathered nearly the same number of examinations. Of the 4262 examinations uploaded from the inclusion centers, 4007 (94%) met the predefined inclusion criteria (table 3).

For each medical question, data was split into three datasets: training set, validation set, and test set (table 4).

Score computation

To calculate scores, a combination of Dice coefficient (measure of similarity between predicted and ground truth segmentation) and mean square error (MSE, between predicted and ground-truth surfaces) was used for sarcopenia, area under curve (AUC) between probability of abnormality and ground-truth for pulmonary nodules, and MSE for EDSS score assessment in MS were used (table 5).

- <u>Muscle area calculation (sarcopenia challenge):</u>

Combination of Sørensen–Dice coefficient (Dice) and mean square error (MSE):

Final Score = minimum score (rank (Dice) + rank (MSE)), among all teams.

Segmentation of abdominal muscles was made using a density threshold ranging from -29 to 150 Hounsfield units (HU).

The teams were evaluated based on the DICE coefficient between the predicted segmentation and the ground-truth segmentation and the MSE between the predicted surface and the ground-truth surface. The overall score is the sum of the team's ranks in the two metrics. The winning team is, therefore, the one with the lowest score. In case of equality, the winning team is the one with the best DICE score.

- <u>Pulmonary nodule classification challenge :</u>

Score = Area under the curve (AUC) of the receiver operating characteristic curve (ROC)

Score = AUC (ground truth classification, Probability of an abnormal examination)

Examinations were annotated by expert radiologist for each abnormal nodules (up to 5 nodules). The localization of each nodule was also annotated. No images resampling was done on these images. Pixel size was indicated in DICOM field PixelSpacing.
Teams were asked to provide an excel file containing the examination name of the exam and the probability of normality/abnormality. The winning team was the one with the best AUC score between the probability of abnormality established by their method and ground-truth.

- Neurological impairment prediction (MS challenge) :

Score = MSE (EDSS score)

The inclusion criteria for the EDSS prediction in multiple sclerosis patients from MRI is a 3D FLAIR or axial T2 FLAIR MRI examination. Patient data included also age, sex, examination date, and EDSS score at 2 years. The results expected by the team was a CSV file containing the exam identifier and the associated EDSS score. The score is then calculated using the MSE between the EDSS score evaluated by the team and the reference score evaluated by a neurologist.

For this 2019 year’s edition, 20 teams were participating with 143 members including: 61 engineers working in startup or in big companies, 25 students, 31 radiologists and 26 researchers or Ph.D students. The teams were ranging from 3 to 15 members. Of the 20 teams, 16 were able to submit results. Six prizes were announced on October 15th three for the best teams, and three for the best inclusion center. For sarcopenia estimation the winning team was Owkin with members from Assistance Publique des Hôpitaux de Paris (APHP) and École Polytechnique, the best inclusion center was Pitié-Salpêtrière hospital. The Pixyl team won the pulmonary nodules challenge with members from Groupement des Hôpitaux de l'Institut Catholique de Lille (GHICL), CHU of Grenoble and Grenoble University, the best inclusion center being CHU Bichat. IBM-france Cognitive team with members from Centre Jean Perrin, Quantacell, Ecole National de l’Aviation Civile (ENAC) and DataValoris won the EDSS prediction challenge with data from the OFSEP. For each challenge, the winning team members were invited on stage to present their method and to receive their price: 3,000 € provided by the SFR.

DISCUSSION

The second edition of the JFR data challenge took in consideration the results of the previous edition [10] and increased the difficulty as requested by the competitors. Three questions were selected for their medical interest and technical feasibility: pulmonary nodule detection and classification from 3D CT, prediction of EDSS in multiple sclerosis using 3D MRI and segmentation of muscular surface for sarcopenia estimation from 2D slices of CT at L3 level. The medical questions went from diagnostic (in last year edition) to prognostic (this year edition). Participants had to handle 3D information from CT and MRI and were asked to perform various tasks.

Larger databases were collected in 2019 edition compared to the previous edition, more than 1000 examinations by challenge versus less than 500 for the previous edition from 24 different centers. Improvements have been made during the inclusion phase regarding the quality of the data, only 6% of examinations didn’t match the inclusion criteria against 19% last year (table 2). This was made possible by giving continuous feedback to the different medical centers during the inclusion phase. The challenge successfully achieved to gather radiologists from many French hospitals, large companies and startups, academics and engineering students from French top engineering schools. This emphasizes the great interest and dynamism shown by the community with regards to the development of AI.

The harsh time restriction imposed to the participants to calculate their prediction prompted the actors to regroup in larger teams up to 15. These constraints pushed many teams to give up, 6 out of 13 teams that joined the pulmonary nodule challenge didn't submit their results on time and more than half of those which joined multiple sclerosis challenge (table 6).

While numerous data challenges have been organized these past years, many of which can be found in the Grand Challenge website, a special effort has been made here to tackle original questions that were not addressed yet.

Sarcopenia is defined by a loss of muscular mass, frequently observed on oncologic patients. Assessment of muscular body mass has been revealed to be a strong prognostic indicator because sarcopenia has been associated with cancer outcomes including treatment toxicity, worse overall survival and disease-free survival [21], in multiple stages and cancer types. E. Burns et al. [22] have developed an algorithm capable of segmenting truncal musculature at multiple lumbar levels on 102 patients achieving a dice score of 0.94%. Participants of our challenge were working at only one lumbar level but reached a slightly better dice score on a much larger set of patients (1025 patients for training and 500 for testing). It is also probable that the bottleneck for this question lies in the quality of the ground truth segmentation than in the deep learning analysis as many teams got very close results in terms of dice score (table 6). Higher quality ground truth segmentation may have been achieved by keeping only segmentation of class 3 and 4 or asking 3 expert radiologists to decide by consensus of the segmentation as it is done in [23].

Multiple sclerosis is the most common immune-mediated disorder affecting the central nervous system and the most common cause of chronic neurological disability in young people. In this disease, the myelin sheath of neurons is damaged, which disrupts the ability of parts of the nervous system to communicate, resulting in a range of signs and symptoms, including physical, mental, and sometimes psychiatric problems. To our knowledge, no studies have been published concerning the prediction of the EDSS from 3D flair images using deep learning. Attempts have been made to extract biomarkers from MRI examinations correlated with disability progression [24,25] but these studies only focus on a limited number of patients with no use of AI techniques. The winning team was the fruit of a successful cooperation and collaboration between seven specialized radiologists from two hospitals, one startup, and one academic laboratory. They manage to predict EDSS with a significant 3.04 mean square error.

Lung cancer is still the leading cause of cancer-related deaths in Europe screening with CT is shown to be effective in reducing mortality from lung cancer [26]. Size is the most important morphologic criteria used to distinguish between malignant and benign nodule. Many studies have been published for automatic detection and classification of pulmonary nodules[27–29]. Ather et al. stands that thanks to the availability of such large databases, the automation of nodule classification is expected to be one of the first application of AI [30]. They nevertheless emphasize the work that has to be done by radiologists to ensure the safe implementation of those tools. Many studies are using the public Lung Imaging Database Consortium (LIDC) [31]. However, most studies perform classification over 2D crops of pulmonary nodules. Nasrullah et al. describe a model capable of detecting and classifying pulmonary nodule from 3D CT that was trained and tested on two large datasets [32]. The originality of the question asked to the participants of our data challenge was to work on a completely new multicentric dataset.

Conclusion

Three data challenge with over 1200 GDPR compliant, multicentric, 2D/3D CT and MRI databases were organized for 20 multidisciplinary teams. In future challenges, it could be stimulating not to restrain the analysis to medical images but to increase complexity by combining different modalities such as histopathological or genetic information and therefore get closer situation faced by radiologists in routine practice.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

Informed consent and patient details

The authors declare that they obtained a written informed consent from the patients and/or volunteers included in the article. The authors also confirm that the personal details of the patients and/or volunteers have been removed.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contribution of authors

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Acknowledgements

We would like to thank the Société Française de Radiologie for the opportunity to organize these challenges during the Journées Francophones de Radiologie.

We would like to thank Gustave-Roussy for the resssources mobilized and the data hosting.

We would like to thank the Commission Nationale de l'Informatique et des Libertés (CNIL) for their support.

We would like to thank Jean François Raffier and Easys for the technical support during the challenge.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Syeda-Mahmood T. Role of Big data and machine learning in diagnostic decision support in radiology. J Am Coll Radiol 2018;15:569–76.

[2] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[3] Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat Methods 2019;16:1247–53.

[4] Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA pediatric bone age machine learning challenge. Radiology 2018;290:498–503.
[5] Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018;15:504–8.
[6] European Society of Radiology (ESR). What the radiologist should know about artificial intelligence – an ESR white paper. Insights Imaging 2019;10:44.
[7] Liew C. The future of radiology augmented with Artificial Intelligence: A strategy for success. Eur J Radiol 2018;102:152–6.
[8] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402.
[9] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017;21:4–21.
[10] Lassau N, Estienne T, de Vomecourt P, Azoulay M, Cagnol J, Garcia G, et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. Diagn Interv Imaging 2019;100:199–209.
[11] Roblot V, Giret Y, Bou Antoun M, Morillot C, Chassin X, Cotten A, et al. Artificial intelligence to diagnose meniscus tears on MRI. Diagn Interv Imaging 2019;100:243–9.
[12] Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diagn Interv Imaging 2019;100:235–42.
[13] Detecting abnormal thyroid cartilages on CT using deep learning. Diagn Interv Imaging 2019;100:251–7.
[14] Detection and characterization of MRI breast lesions using deep learning. Diagn Interv Imaging 2019;100:219–25.
[15] Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imaging 2019;100:227–33.
[16] Kidney cortex segmentation in 2D CT with U-Nets ensemble aggregation. Diagn Interv Imaging 2019;100:211–7.
[17] Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol 2014;15:1332–41.
[18] Fahrbach K, Huelin R, Martin AL, Kim E, Dastani HB, Rao S, et al. Relating relapse and T2 lesion changes to disability progression in multiple sclerosis: a systematic literature review and regression analysis. BMC Neurol 2013;13:180.
[19] Egger C, Opfer R, Wang C, Kepp T, Sormani MP, Spies L, et al. MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? NeuroImage Clin 2017;13:264–70.
[20] Zopfs D, Theurich S, Große Hokamp N, Knuever J, Gerecht L, Borggrefe J, et al. Single-slice CT measurements allow for accurate assessment of sarcopenia and body composition. Eur Radiol 2019.
[21] Shachar SS, Williams GR, Muss HB, Nishijima TF. Prognostic value of sarcopenia in adults with solid tumours: A meta-analysis and systematic review. Eur J Cancer 2016;57:58–67.
[22] Barnard R, Tan J, Roller B, Chiles C, Weaver AA, Boutin RD, et al. Machine learning for automatic paraspinous muscle area and attenuation measures on low-dose chest CT Scans. Acad Radiol 2019;26:1686–94.
[23] Lin L, Dou Q, Jin Y-M, Zhou G-Q, Tang Y-Q, Chen W-L, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology 2019;291:677–86.
[24] von Gumberz J, Mahmoudi M, Young K, Schippling S, Martin R, Heesen C, et al. Short-term MRI measurements as predictors of EDSS progression in relapsing-remitting multiple sclerosis: grey matter atrophy but not lesions are predictive in a real-life setting. PeerJ 2016;4.

[25] Poonawalla AH, Datta S, Juneja V, Nelson F, Wolinsky JS, Cutter G, et al. Composite MRI scores improve correlation with EDSS in multiple sclerosis. Mult Scler Houndmills Basingstoke Engl 2010;16:1117–25.
[26] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.
[27] Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg 2017;12:1799–808.
[28] Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 2017;7:1–11.
[29] Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Comput Math Methods Med 2016;2016:1–7.
[30] Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. Clin Radiol 2020;75:13–9.
[31] Armato III SG, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, et al. Lung image database consortium: Developing a resource for the medical imaging research Community. Radiology 2004;232:739–48.
[32] Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors 2019;19.

Table 1 Images inclusion criteria for each challenge

Challenge	**Sarcopenia**	**Pulmonary nodules**	**Multiple sclerosis**
Modality	CT	CT	MRI
Image type	2D	3D	3D
Images specifications	• soft filter • slice thickness 1.25 to 7 mm • with or without contrast agent injection	• hard filter (mostly) • slice thickness <1.25mm • with or without contrast agent injection • containing between 1 and 5 nodules (density: mainly solid, but also mixed or ground glass). • malignant nodules defined by a volume > 100 mm3, or diameter > 6 mm	• 3D FLAIR , or axial 2D FLAIR

Table 2 Inclusion centers for each of the challenges

Sarcopenia	**Pulmonary Nodules**	**Multiple Sclerosis**
1557 2D CT scan	1237 3D CT	1468 3D MRI
Multicentric	Multicentric	Multicentric
•Pitié-Salpêtrière *(518)* •Gustave Roussy *(328)* •CHU de Nîmes *(252)* •Henri-Mondor *(177)* •CHU Lille *(101)* •Hospices Civiles Lyon *(68)* •Bicêtre *(58)* •GHPSJ *(30)* •Centre Léon Berard *(16)* •RPO cabinet libéral *(9)*	•CHU Bichat *(267)* •CHU de Nîmes *(143)* •CHU Rennes *(137)* •CHU Bordeaux *(105)* •CHU Cochin *(102)* •CHU Grenoble *(100)* •CHU Nice *(99)* •Avicenne *(98)* •CHU Clermont-Ferrand *(50)* •Gustave Roussy *(45)* •CHU Montpellier *(42)* •CH Douai *(25)* •CHU Strasbourg *(17)* •CHRU Besançon *(6)* •Institut A. Tzanck libéral *(1)*	•OFSEP[1] *(1468)*

[1] Observatoire Français de la Sclérose En Plaques

Table 3 Exams received and included for each challenge

Challenge	**Number of examinations received**	**Number of examinations kept in the dataset**
Sarcopenia	1557	1515 (1515/1557; 97%)
Pulmonary Nodules	1237	1031 (1031/1237; 83%)
Multiple Sclerosis	1468	1461 (1461/1468; 99%)
Total	***4262***	***4007 (4262/4007; 94%)***

Table 4 Training, validation, and testing dataset for each challenge

Challenge	Sarcopenia	Pulmonary nodules	Multiple sclerosis
Training set	513	343	480
Validation set	512	344	498
Test set	500	344	483
Total (# examinations)	1525	1031	1461

Table 5 Details of the score used to rank the teams

Score	Formula
Dice score	$D = 2\|X \cap Xc\|/ \|X\| + \|Xc\|$, with X ground truth segmentation and Xc, segmentation from teams
Mean Squared Error	$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Yn} - Yn)^2$, with $\hat{Y}$= team's value estimation, and Y= ground-truth value

Table 6 Results calculated for each team by challenge

Teams	Sarcopenia	Pulmonary nodules (AUC)	Multiple sclerosis
Aidence		0.878	
ALICEMEDICAL			5.47
Autonomous Radio	11	*N.S*	*N.S*
Axone	16		*N.S*
Azmed			4.61
biSEPs	*N.S*		
Dapsi-AI	*N.S*		
EASYPICKY	*N.S*		
GAMC	7	0.793	*N.S*
Ghicl-Pixyl			**3.04**
IBM-France-Cognitive		**0.899**	
icometrix			3.80
ILLUIN TECHNOLOGY	*N.S*		
LEVIATAN	10	*N.S*	*N.S*
LyPhTe	7	0.838	3.92
Milvue	6		
NAIS		0.681	*N.S*
ONCONEURAL		0.644	*N.S*
Owkin	**4**	0.768	*N.S*
Tripode-rouen	11		

Bold indicates the best result for each competition
N.S. indicates that the team left the challenge. Empty cell indicates not competing team.