Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI

N. Lassau, T. Estienne, P. de Vomecourt, M. Azoulay, John Cagnol, G. Garcia, M. Majer, E. Jehanno, R. Renard-Penna, C. Balleyguier, et al.

Version of Record: https://www.sciencedirect.com/science/article/pii/S2211568419300294
Manuscript_317b4c53250c82ea6912775d7f48e4a0

Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI

Authors: Nathalie Lassau[a,b*], Theó Estienne[c], Philippe de Vomecourt[a], Mikael Azoulay[d], John Cagnol[e], Gabriel Garcia[d], Michael Majer[d], Emmanuel Jehanno[b], Raphaëlle Renard Penna[f], Corinne Balleyguier[a,b], François Bidault[a,b], Caroline Caramella[a,b], Thibaut Jacques[h], Frédérique Dubrulle[h], Julien Behr[i], Nicolas Poussange[j], Jonathan Bocquet[k], Sarah Montagne[l], François Cornelis[m], Marie Faruch[n], Bertrand Bresson[o], Serge Brunelle[p], Aurélie Jalaguier-Coudray[p], NicolasAmoretti[q], Alain Blum[r], Anita Paisant[s],Victor Herreros[t], Olivier Rouviere[u], Salim Si-Mohamed[v], Lucy Di Marco[w], Marc Garetier[x], Frédéric Pigneur[y], Antonin Bergère[z], Catherine Cyteval[aa], Laure Fournier[bb], Caroline Malhaire[cc], Jean-Luc Drape[dd], Edouard Poncelet[ee], Corinne Bordonne[ff], Hugo Cauliez[gg], Jean-François Budzik[hh], Martine Boisserie[ii], Thibault Willaume[jj], Sébastien Molière[kk], Noémie Peyron Faure[ll], Septimu Caius Giurca[mm], Valérie Juhan[nn], Thomas Caramella[oo], Antoine Perrey[pp], Florain Desmots[qq], Matthieu Faivre-Pierre[rr], Martine Abitbol[ss], Raïssa Lotte[tt], Diana Istrati[uu], Daphné Guenoun[vv], Alain Luciani[ww], Marc Zins[ww], Jean-François Meder[ww], Anne Cotten[ww]

Affiliations:

[a] Department of Radiology, Institut Gustave Roussy, 94805, Villejuif, France.

[b] IR4M, CNRS, Université Paris-Sud, Université Paris-Saclay 91400, Orsay, France.

[c] Centre de Vision Numérique, CentraleSupélec, Université Paris-Saclay - Radiothérapie Moléculaire, Université Paris-Sud, Gustave Roussy, Inserm, Université Paris-Saclay, France

[d] Department of Radiology, Institut, Gustave Roussy, 94800 Villejuif, France.

[e] Fédération de Mathématiques de CentraleSupélec et Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France.

[f] Services de Radiologie, Hôpitaux Tenon et Pitié-Salpêtrière, AP-HP, GRC-UPMC n°5 Oncotype-URO, Sorbonne Université, 75020 Paris, France.

[g] Service de Radiologie et Imagerie Musculosquelettique, Centre de Consultations et d'Imagerie de l'Appareil Locomoteur, 59037 Lille, France.

[h] Department of Radiology, Hôpital Huriez, CHU Lille, 59037 Lille, France,

[i] Service de Radiologie Viscèrale, CHRU Jean Minjoz Besançon, 25030 Besançon, France,

[j] Department of Radiology, Clinique du Sport, 33700 Merignac, France.

[k] Department of Radiology, Clinique Saint Jean, 83100 Toulon, France

[l] Department of Radiology, Tenon and Pitié-Salpêtrière Academic Hospital, APHP, Sorbonne-Université, Paris, France.

[m] Department of Radiology, Sorbonne Université -ISCD / APHP - HUEP, Tenon Hospital, 75020 Paris, France.

[n] Department of Radiology, Hôpital Pierre Paul Riquet CHU Purpan, 3100, Toulouse, France.

[o] Department of Radiology, Hôpital Bicêtre, 94270 Le Kremlin Bicetre, France.

[p] Department of Radiology, Centre Paoli Calmettes, 13009 Marseille, France.

[q] Radiologie Ostéoarticulaire, Hôpital Pasteur, CHU de Nice, 06000 Nice, France.

[r] Department of Radiology Guilloz, CHRU Nancy, 54000 Nancy, France.

[s] Department of Radiology, CHU Angers, 49933, Angers, France.

[t] Department of Radiology, Hôpital de La Croix Rousse, 69004 Lyon, France.

[u] Department of Radiology, Pavillon B, Hôpital Edouard Herriot, 69003, Lyon, France.

[v] Service d'Imagerie Cardio Thoracique et Vasculaire, Hôpital Cardiologique Louis Pradel, 69500 Bron, France.

[w] Department of Radiology, Centre Hospitalier Lyon-Sud, 69310 Pierre-Bénite, France.

[x] Department of Radiology, Hôpital d'Instruction des Armées Clermont-Tonnerre, 29240 Brest, France.

[y] Service d'Imagerie Médicale, AP-HP, Hôpitaux Universitaires Henri Mondor, 94010 Creteil, France.

[z] Department of Radiology, Hopital Saint Joseph, 75014 Paris, France.

[aa] Service d'Imagerie Ostéo Articulaire, Hôpital Lapeyronie -UM Montpellier, 34295 Montpellier, France.

[bb] Université Paris Descartes, Department of Radiology, Hôpital Européen Georges Pompidou, 75015 Paris, France.

[cc] Department of Radiology, Institut Curie, 75005 Paris, France.

[dd] Department of Radiology, Hopital Cochin, 75014 Paris, France.

[ee] Service Imagerie de la femme, Centre Hospitalier de Valenciennes, 59300 Valenciennes, France.

[ff] Department of Radiology, Hôpital Hôtel-Dieu, APHP, 75004 Paris, France.

[gg] Department of Radiology, Centre Oscar Lambret, 59037 Lille, France.

[hh] Service d'Imagerie Musculosquelettique GHICL, Hôpital Saint Vincent de Paul, Université Catholique de Lille, 59000 Lille, France.

[ii] Department of Medical Imaging, Institut Bergonié, 33000 Bordeaux, France.

[jj] Radiologie Ostéoarticulaire, Hôpital de Hautepierre, 67200 Strasbourg, France.

[kk] Radiologie Viscérale, Hôpital de Hautepierre, 67200 Strasbourg, France.

[ll] Clinique Val d'Ouest, 69130 Écully, France.

[mm] Centre d'Imagerie du Chinonais, 37500 Chinon, France.

[nn] Department of Radiology, Hôpital Européen, 13331 Marseille, France.

[oo] Institut du Sein, 06000 Nice, France.

[pp] Department of Radiology, Hôpital Montpied, CHU de Clermont Ferrand CHU, 63003 Clermont-Ferrand, France.

[qq] Department of Radiology, HIA Laveran, 13013 Marseille, France.

[rr] Clinique du Pont saint Vaast, 59500 Douai, France.

[ss] CSE Imagerie Médicale, 75010 Paris, France.

[tt] Department of Radiology, Hôpital Saint-Antoine, 7501, Paris, France.

[uu] Department of Radiology, Centre Hospitalier d'Armentières, 59280 Armentières, France.

[vv] APHM, Hôpital Sainte-Marguerite, Institute for Locomotion, 13009 Marseille, France.

[ww] Société Française de Radiologie, 75013 Paris, France.

***Corresponding author:** nathalie.lassau@gustaveroussy.fr

Abstract

Purpose: The goal of this data challenge was to create a structured dynamic with the following objectives: 1) teach radiologists the new rules of General Data Protection Regulation (GDPR), while building a large multicentric prospective database of ultrasound, computed tomography (CT) and MRI patient images, 2) build a network including radiologists, researchers, start-ups, large companies, and students from engineering schools, and 3) provide all French stakeholders working together during 5 data challenges with a secured framework, offering a realistic picture of the benefits and concerns in October 2018.

Materials and Methods: Relevant clinical questions were chosen by the Société Francaise de Radiologie. The challenge was designed to respect all French ethical and data protection constraints. Multidisciplinary teams with at least one radiologist, one engineering student, and a company and/or research lab were gathered using different networks, and clinical databases were created accordingly.

Results: Five challenges were launched: detection of meniscal tears on MRI, segmentation of renal cortex on CT, detection and characterization of liver lesions on ultrasound, detection of breast lesions on MRI, and characterization of thyroid cartilage lesions on CT. A total of 5,170 images within 4 months were provided for the challenge by 46 radiology services. Twenty-six multidisciplinary teams with 181 contestants worked for one month on the challenges. Three challenges, meniscal tears, renal cortex, and liver lesions, resulted in an accuracy > 90%. The fourth challenge (breast) reached 82% and the lastone (thyroid) 70%.

Conclusion: Theses five challenges were able to gather a large community of radiologists, engineers, researchers, and companies in a very short period of time. The accurate results of three of the five modalities suggest that artificial intelligence is a promising tool in these radiology modalities.

Keywords: Artificial intelligence (AI); Meniscal tears; Breast lesions; Thyroid cartilage; Liver lesions; Renal cortex.

Introduction

World interest in artificial intelligence (AI) is growing rapidly due to the availability of large and scalable datasets in many industries, advances in computing power particularly central processing unit (CPU) and graphics processing units (GPI), and the never-ending release of new algorithms. New standards of machine learning, such as deep learning, have tremendous impact on radiologic activities [1]. This major change seems to enable radiologists to leverage their value, efficiency, accuracy, and personal satisfaction [2]. Thrall et al. have analyzed key success factors of such changes in radiology based on value created, including increased diagnostic certainty, faster turnaround, better outcomes for patients, and better quality of work life for radiologists [3]. The strategic positioning of all participants (patients, radiologists, AI expert, information technology department) is crucial for a successful transition [4]. The Canadian Association of Radiologists (CAR) and the French Society of Radiology (SFR) published a white paper discussing the importance of AI and the probable impact on the community of radiologists in the near future [3, 5, 6]. Private and public radiologists have worked together to include all modalities such as ultrasound, computerized tomography (CT), and magnetic resonance imaging (MRI) [7]. At the same time, the European Union is reforming data protection legislation via the General Data Protection Regulation (GDPR), with new legislation implemented in May 2018 [8].

The goal of this data challenge, organized by the SFR during the 2018 Journées Francophones de Radiologie (JFR), was to create a structured dynamic with the following objectives: 1) teach radiologists the new rules of GDPR while building a large multicentric prospective database of ultrasound, CT, and MRI patient images, 2) build a network from May to September 2018 including public and private radiologists, researchers, start-ups, large companies, and students from engineering schools, and 3) provide all French stakeholders working together during 5 data challenges with a secured framework, offering a realistic picture of the benefits and concerns in October 2018.

Material and methods

Clinical questions

The feasibility of the challenge was assessed by a team of radiologists and data scientists, based on a bibliography of existing challenges and the state of the French ecosystem of radiology and AI. The goal was set for a minimum database of 600 2D images. The organ

societies of the SFR were asked to propose questions for the challenge, considering four criteria. First, the clinical relevance was judged by the radiologists. A literature search was performed using Pubmed, Kaggle website and grand-challenge website for this kind of competition in order to avoid reproducing an already existing challenge. Data-scientists from a top French engineering school studied the feasibility of the challenge with the data provided. The number of exams available per year for each question was estimated to assure it was possible to gather the images in a very short period of time. Only the validated challenges on a shortlist that would reach the 600-image database would be launched and available for teams. Five clinical questions regarding selection criteria were chosen by the SFR. The inclusion criteria for each medical question were defined by the organ society responsible of the question.

Security and data protection

Regarding GDPR, a French regulation office, Commission Nationale Informatique et Libertés (CNIL), was consulted to assure the project design was compliant. A methodology reference was chosen in order to ease the data collection. Patients were provided a letter with all necessary information. The data collected could only be used for the aim of this challenge, and only one image per exam was selected to avoid patient identification. An ethical chart was written by the AI group of the SFR and required the signature of every radiologist uploading exams to the database. The data was stored within the Gustave Roussy Institute server to guarantee the same security as to clinical patient data. Two separate phases were designed for the project: a collection phase, where data could only be uploaded to the server by users identified, and a competition phase, where only the treated and checked data could be downloaded by identified team leaders.

Communication and team gathering

The teams were to have at least one radiologist, one engineering student, and a research lab and/or company. The networks of the JFR and SFR were used to gather radiologists. For students and research labs, the networks of French Graduate Schools and Universities and Life Imaging were used.

Companies that subscribed to JFR and data analytics startups were informed. Each could subscribe online by completing a form with the information required. The organization team published the contestant's contact information and the teams were allowed to gather at

their convenience. After three months, the challenge staff proposed team compositions to contestants as an icebreaker. Only the teams with full composition were allowed to participate and access the datasets. The inclusion criteria and tutorials were published on the challenge website. Data were uploaded by each radiologist directly to the website. Contestants could access the event details, the team's contact information, and the link to download datasets online. WordPress was used to develop the website, with different plug-ins for uploading data and user management.

Inclusion and challenge phases

When the inclusion phase began on May 15th 2018, radiologists used several networks to participate in the data collection: the SFR, JFR website and newsletters, the challenge website, and the French organ societies.

To motivate the radiology centers, inclusion status was sent twice a week to all the radiologists registered online and graphs showing the inclusion dynamic were included. The teams received access to the first part of the datasets on September 15th, and the second part on October 12th. Work stations with internet access were available for each team within the congress center of the JFR. The competition was launched on October 14th 2018, along with the publication of the validation dataset. Each team had one hour to send the result of their work. Winning teams were announced on October 15th.

Results

Renal cortex (CT)

Because glomerular filtration, an important clinical assessment of renal functions [9], is the main function of the renal cortex, there has been a considerable interest assessing renal cortex volume and thickness [9, 10]. Indeed, renal cortical volume and thickness have been proven to be effective biomarkers for renal function in many clinical situations. Applications include identifying patients with kidney that function well in the context of evaluating potential kidney donors, choosing the best resection plane in case of partial nephrectomy for a better preservation of renal function for urological treatment, and the assessment of clinical outcomes post-operatively. A non-invasive evaluation of cortex thickness is therefore critical and automatic segmentation of the renal cortex could play a key role for functional and morphological renal assessments [11, 12]. However, renal cortex segmentation is complex

because the anatomy of renal cortex varies in shape and size, and is not easily distinguished from neighboring tissues (i.e. vessels and renal columns) because of similar intensity.

For this question, participants only had one task of performing automatic segmentation of the clinical cortex. The images were given in Niifti files as the segmentation mask. For the test set, participants returned a binary image with the same size and format as the images in the trial set. One scanner image in the coronal plane, oblique along the major axis, of the kidney before and after injection of contrast material was requested for upload.

Breast lesions (MRI)

Breast MRI is a key imaging tool for breast lesion diagnosis and characterization [13]. Major indications of breast MRI are cancer staging, clarifying findings in equivocal lesions, treatment evaluation after neoadjuvant chemotherapy, and cancer screening in high-risk women [13]. Breast MRI is known to have high sensitivity but sometimes lacks specificity. A challenge specific to staging breast cancer is assessing which contrast uptake is malignant vs. benign, as many types of contrast-uptakes are visible, and when normal breast tissue may also show multiple contrast-uptakes. In addition, breast cancer increasingly appears as multifocal lesions, especially in high-grade breast tumors or those with positive Her2 receptors. Thus, there is a need to improve breast lesion characterization on MRI, despite the current use of multimodal sequences such as T2-weighted imaging and diffusion-weighted imaging to improve breast lesion characterization [14].

The goal was to evaluate whether techniques of AI would improve breast lesion characterization on MRI [15]. For this question, there were two different tasks: the first was to predict the benign or malignant nature of the lesion, and the second was to classify the lesion among 17 types of lesions. Due to a small number of images for some lesions, we decided to have only 4 groups: glandular tissue, infiltrating ductal carcinoma, other benign lesions, and other malignant lesions. One slice of 3D native dynamic sequence with injection with the two breasts visible was provided. The lesion, proven histologically or with an anterior MRI, was requested for upload.

Liver lesions (ultrasound)

Because liver ultrasound is operator-dependent, anything that increases the standardization and robustness is useful. Until now, few studies in ultrasound were published using AI, and have been focused on breast ultrasound, fatty detection on liver, and thyroid nodules [16-18].

For this reason, the radiologists on SFR explored the detection and characterization of liver nodules. This is often the primary concern during an abdominal examination and has not been incorporated into a challenge to date.

For this question, the participants had three different tasks: detect the presence of a lesion in the image, characterize the lesion as benign or malignant, and classify among 5 types of lesions (cyst, angioma, focal nodular hyperplasia [FNH], hepatocellular carcinoma [HCC] and metastasis). One cross-sectional image of the liver with the edge apparent and without rib artifacts or calipers was requested, the gold standard being one injected imaging exam or one biopsy.

Meniscal tears (MRI)

Meniscal tears have an annual incidence estimated between 9 and 16 per 10,000 patients [19]. Tears can be either traumatic or degenerative in origin and affect the lateral and/or medial meniscus, and the orientation can be classified as vertical (longitudinal, radial, or oblique) or horizontal. Precise detection and characterization is crucial for proper orthopedic management [20]. MRI is an important noninvasive diagnostic tool and has a reported accuracy above 85%, although direct examination in arthroscopy remains the gold standard [21]. Proton density or intermediate-weighted sequences with fat suppression are the standard of care for meniscal tear detection [22]. Due to a complex orientation in space, three planes (sagittal, axial and coronal) or 3D sequences are required to examine every part of the menisci. However, meniscal tears can be subtle and certain kind of lesions are frequently missed [22]. Few peer-reviewed articles have focused on meniscus analysis in MRI using machine learning to date, with no prospective cohort focusing on meniscal tears published at the time of the data challenge [23, 24].

One MRI slice per meniscus was included for the data challenge (proton density or intermediate-weighted sequences with fat suppression). To enable analysis of both the anterior and posterior horn of each meniscus, the sagittal plane was chosen, with quality control instructions to avoid partial volume effects on meniscus edges. Patients who underwent previous knee surgery were not included. Both menisci were included for each patient. These technical choices enabled high reproducibility in the inclusion process. The goal of this task was to detect and characterize meniscus lesions, with three different tasks: detect the presence of a lesion in the meniscus, localize the fissure (anterior horn or posterior horn), and characterize the orientation of the fissure (horizontal or vertical). MRI examination

has to be performed in the sagittal plane, obtained at 1.5- or 3T, passing through the anterior horn of the meniscus, and the anterior and posterior horn should be clear and well separated.

Thyroid cartilage (CT)

Tumor invasion of the thyroid cartilage by laryngeal carcinoma worsens the prognosis after radiotherapy treatment and may lead to a preference for primary surgery [25]. Conversely, the absence of tumor invasion of the cartilage leads to the choice of radiotherapy, chemoradiotherapy, or chemotherapy followed by radiotherapy [26-28]. On CT, the presence of lysis and/or condensation is indicative of tumor invasion of thyroid cartilage and tumor contact with cartilage. The performance of CT (*i. e.*, sensitivity and specificity for the diagnosis of thyroid cartilage invasion) differs widely in clinical studies depending on the T stage of the tumor and the presence of imaging review [29, 30]. As a result, it is interesting to ask AI to identify abnormalities of thyroid cartilage so as to advise the radiologist and the multidisciplinary team. The objective of the challenge was to evaluate the feasibility of an analysis of thyroid cartilage abnormalities by artificial intelligence.

For this question, there was only one task: characterize the aspect of the thyroid cartilage as normal or abnormal (lysis or condensation). One scanner image of a patient with an initial check-up that warranted suspicion for laryngeal or hypopharyngeal tumors, was requested for upload. The slice should be at cartilage level and with a thickness of at least 1 mm.

Communication and team gathering

A total of 323 persons registered online for the challenge, comprised of 45 researchers, 81 students, 78 engineers, and 119 radiologists. A total of 26 teams were complete and validated with 181 persons: 27 researchers, 52 students, 59 engineers, and 43 radiologists. The website generated 5,230 sessions from June to October 2018. Inclusions began on May 15th, and the first images were uploaded at the beginning of June. A total of 5,130 images were uploaded (Fig. 2) from 46 radiology services (30 public, 13 private, 3 cancer centers). The meniscus and renal cortex challenges collected more images than the others. The five challenge groups submitted more than 600 images each before September 15th, and hence were selected for the challenge.

Data processing

For each medical question, we defined a score function (Fig. 1). Except for the renal cortex which was evaluated with the Dice Score, these score functions were linear combinations of the binary AUC of each label. For instance, for the breast question, we calculated 5 AUC (one for the benign/malignant and 4 for each type of lesions). The linear combinations of each clinical question were defined in agreement with the referent radiologist. The formula of the score as well as the Python file used were provided to each team.

When possible, all images were resampled to the same pixel size and cropped to the same size. The pixel size and the image size are shown in Table 1. To resample the data, dicom field PixelSpacing (0028, 0030) and the python library SimpleITK were used. A linear interpolator was used to do the resampling. Some images were not resampled due to two reasons: 1) The dicom field PixelSpacing was empty due to anonymization of the image by the radiologist, and 2) for ultrasound images, the dicom field Pixel Spacing is not relevant. The images include markers which are used to determine the size of the image, and this does not allow an automatic process to standardize a dataset.

Image size was an important question: the bigger the images, the more difficult the questions were. Figure 3 shows a comparison of the same image centered on the meniscus. For each clinical question, we performed tests in order to find an optimal size of images. To crop the image, two different methods were used. The first was based on a manual Python application developed specifically for the project, which was used for breast, thyroid and meniscus. For each image, the user clicks on the centers of the lesion and a box was drawn centered on the click. The cropped image was saved as a Nifti file. The Python application accelerates the process: keyboard keys were associated with functions like cropping, saving, or loading the image. This allowed us to process hundreds of images per hour. For the renal cortex, the crop was made automatically: there is a binary mask of the renal cortex, so that the crop can be made around a centroid. The ultrasound images were not cropped; all the information from the scan written on the images had to be deleted (Fig. 4). For this, a Python tool was used to detect the contour of the ultrasound inside the images and put 0 value outside the ultrasound. This code was adapted to different ultrasound systems as they do not use the same encoding. As a result, the participants had only raw data. The final step of data preparation was the pre-processing of the label files (Fig. 5). When the data was uploaded to the web site, the labels were saved in a unique CSV file. To have a clean label file, the following steps were needed: one unique label file was created for each clinical question, then

all non-relevant information (date of upload) was deleted. Categorical labels were converted to binary labels, and all false characters were cleaned. The file quality and inclusion criteria were checked by a data-scientist and the referent radiologist for each challenge. The senior radiologist also replaced the incorrect labels. A total of 4,170 images could be used for the dataset (Table 2).

For each medical question, data were split into three datasets: train set, validation set, and test set (Table 3). The labels (lesion or not, malignant/benign) were equally distributed between the datasets in order to have the same proportion. Of the 26 teams, 23 submitted results (Table 4). Six prizes were announced on October 15th: LyPhTe team for renal cortex, RadioAdvisor and LyPhTe for meniscus fissure, Owkinautes for liver lesions, SynovIA for thyroid, and Owkinautes for breast. For each challenge, the winning team and its members were invited on the stage to present their method and algorithm and to receive their price: 3,000 euros and a publication in a medical journal.

Discussion

The JFR data-challenge covered ultrasound, CT and MRI with 5 clinical questions in parallel. The contestants had to solve different type of tasks, including segmentation, detection, and characterization. The rate of inclusion of 5,170 images within 4 months was efficient with large mobilization of 46 centers. The 26 multidisciplinary teams with 181 contestants demonstrated impressive gathering of the four communities: radiologists, researchers, students, and companies. It can be noted that three challenges had very good results, with an AUC over 0.90.

Between 6-24% of the images received had to be excluded from the datasets. Publications from other challenges usually do not describe the database cleaning process. The amount of images cleaned out of the database can be explained by the team of radiologists and data scientists who rigorously analyzed the data quality. The data quality may have been increased by regular checkup of the data during the inclusion phase and by giving regular feedback to each medical center regarding the data uploaded.

There have been numerous data challenges with various formats, datasets, and prices. A number of these challenges can be found in the Grand Challenge website. Most have the same characteristics: one modality (MRI for BRATS, CT for LUNA, Histology for BACH), one organ (Liver for LITS Challenge, Breast for CAMELYON17, and one task (segmentation for BRATS, Detection for RNSA Challenge) [31-34].

Automatic segmentation is currently a main topic in medical images and machine learning. In regards to kidney, there are not as many publications compared to other organs like lungs or brain. The main difference is that most of the articles are based on a machine learning algorithm such as a graph or random forest [35-38], and not a deep learning-based method, which were used by the participants of our challenge. Some articles also use 3D MRI [37], although they have few images (under 100 patients).

Concerning breast lesions, the main technique used for data challenges is the mammography, the gold standard for detecting early-stage breast cancer before lesions are clinically detected. Several challenges have focused on breast cancer (Digital Mammography DREAM Challenge [39], CAMELYON [32], or BACH [31]), but they all focus only on 3 modalities: mammography, histopathology and ultrasound. To our knowledge, this is the first challenge assessing breast lesions using MRI images [40-42].

For liver lesions, no data challenge was organized using US. The current theme is focused on detecting nonalcoholic fatty liver, a factor for the development of hepatocellular carcinoma. The researchers used a hepatic renal index defined as a ratio of average brightness level of the liver and the kidney cortex with a good AUC of 0.97. The author used 540 images from 54 patients (10 per patient), but until now, no publication was available on detection and characterization of focal lesion [17]

Contrary to lung lesions and breast lesions, recent algorithms for machine learning have not been frequently used for the knee. In Bien et al. study, deep learning was used to detect anterior cruciate ligament tears and meniscal tears on knee MRI [23]. This approach is similar to that used by the participants in our challenge except they did not predict the position and the orientation of the tears. Other articles focus on cartilage lesion or predicting osteoarthritis [43].

These five challenges gathered a large community of radiologists, engineers, researchers, and companies in a short period of time. The results on the three modalities with an AUC > 0.90 show AI is a promising field on the three radiology modalities. In the future, it could be useful to increase the number of patients with an equal distribution between normal and abnormal images.

Conflict of interests

The authors declare that they have no conflicts of interest relating to this paper.

References

[1] T. Syeda-Mahmood, Role of big data and machine learning in diagnostic decision support in radiology, J Am Coll Radiol 15 (2018) 569-76.

[2] M. Recht, R.N. Bryan, Artificial intelligence: threat or boon to radiologists? J Am Coll Radiol 14 (2017) 1476-80.

[3] J.H. Thrall, X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer, et al., Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success, J Am Coll Radiol 15 (2018) 504-8.

[4] C. Liew, The future of radiology augmented with artificial intelligence: a strategy for success, Eur J Radiol 102 (2018) 152-6.

[5] Artificial intelligence and medical imaging 2018: French Radiology Community white paper, Diagn Interv Imaging 99 (2018) 727-42.

[6] A. Tang, R. Tam, A. Cadrin-Chenevert, W. Guest, J. Chong, J. Barfett, et al., Canadian Association of Radiologists white paper on artificial intelligence in radiology, Can Assoc Radiol J 69 (2018) 120-35.

[7] J.P. Beregi, M. Zins, J.P. Masson, P. Cart, J.M. Bartoli, B. Silberman, et al., Radiology and artificial intelligence: an opportunity for our specialty, Diagn Interv Imaging 99 (2018) 677-8.

[8] F. Pesapane, C. Volonte, M. Codari, F. Sardanelli, Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States, Insights Imaging 9 (2018) 745-53.

[9] L.A. Stevens, J. Coresh, T. Greene, A.S. Levey, Assessing kidney function--measured and estimated glomerular filtration rate, N Engl J Med 354 (2006) 2473-83.

[10] M.D. Beland, N.L. Walle, J.T. Machan, J.J. Cronan, Renal cortical thickness measured at ultrasound: is it better than renal length as an indicator of renal function in chronic kidney disease?, AJR Am J Roentgenol.195 (2010) W146-9.

[11] X. Li, X. Chen, J. Yao, X. Zhang, F. Yang, J. Tian, Automatic renal cortex segmentation using implicit shape registration and novel multiple surfaces graph search, IEEE Trans Med Imaging 31 (2012) 1849-60.

[12] F. Artunc, S. Yildiz, C. Rossi, A. Boss, H. Dittmann, H.P. Schlemmer, et al., Simultaneous evaluation of renal morphology and function in live kidney donors using dynamic magnetic resonance imaging, Nephrol Dial Transplant 25 (2010) 1986-91.

[13] H.I. Greenwood, R.I. Freimanis, B.M. Carpentier, B.N. Joe, Clinical breast magnetic resonance imaging: technique, indications, and future applications, Semin Ultrasound CT MR 39 (2018) 45-59.

[14] K.W. Kim, C.M. Kuzmiak, Y.J. Kim, J.Y. Seo, H.K. Jung, M.S. Lee, Diagnostic usefulness of combination of diffusion- and T2-weighted imaging, including apparent diffusion coefficient in breast resions: assessment of histologic grade, Acad Radiol 25 (2018) 643-52.

[15] M. Codari, S. Schiaffino, F. Sardanelli, R.M. Trimboli, artificial intelligence for breast MRI in 2008-2018: a systematic mapping review, AJR Am J Roentgenol (2019) 1-13.

[16] L. Wang, S. Yang, S. Yang, C. Zhao, G. Tian, Y. Gao, et al., Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network, World J Surg Oncol 17 (2019) 12.

[17] M. Byra, G. Styczynski, C. Szmigielski, P. Kalinowski, L. Michalowski, R. Paluszkiewicz, et al., Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images, Int J Comput Assist Radiol Surg 13 (2018) 1895-903.

[18] F. Sadoughi, Z. Kazemy, F. Hamedan, L. Owji, M. Rahmanikatigari, T.T. Azadboni, Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review, Breast Cancer 10 (2018) 219-30.

[19] S. Kim, J. Bosque, J. Meehan, A. Jamali, R. Marder, Increase in outpatient knee arthroscopy in the United States: a comparison of national surveys of ambulatory surgery, 1996 and 2006, J Bone Joint Surg Am 93 (2011) 994-1000.

[20] S.C. Mordecai, N. Al-Hadithy, H.E. Ware, C.M. Gupte, Treatment of meniscal tears: An evidence based approach, World J Orthop 5 (2014) 233-41.

[21] R. Crawford, G. Walley, S. Bridgman, N. Maffulli, Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and ACL tears: a systematic review, Br Med Bull 84 (2007) 5-23.

[22] F. Lecouvet, T. Van Haver, S. Acid, V. Perlepe, T. Kirchgesner, B. Vande Berg, et al., Magnetic resonance imaging (MRI) of the knee: Identification of difficult-to-diagnose meniscal lesions, Diagn Interv Imaging 99 (2018) 55-64.

[23] N. Bien, P. Rajpurkar, R.L. Ball, J. Irvin, A. Park, E. Jones, et al., Deep-learning-assisted diagnosis for knee magnetic resonance imaging, PLoS Med 15 (2018) e1002699.

[24] B. Norman, V. Pedoia, S. Majumdar, Use of 2D U-Net Convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry, Radiology 288 (2018) 177-85.

[25] G. Department of Veterans Affairs Laryngeal Cancer Study, G.T. Wolf, S.G. Fisher, W.K. Hong, R. Hillman, M. Spaulding, et al., Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer, N Engl J Med 324 (1991) 1685-90.

[26] A.A. Forastiere, H. Goepfert, M. Maor, T.F. Pajak, R. Weber, W. Morrison, et al., Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer, N Engl J Med 349 (2003) 2091-8.

[27] A.A. Forastiere, Q. Zhang, R.S. Weber, M.H. Maor, H. Goepfert, T.F. Pajak, et al., Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer, J Clin Oncol 31 (2013) 845-52.

[28] G. Janoray, Y. Pointreau, P. Garaud, S. Chapet, M. Alfonsi, C. Sire, et al., Long-term results of a multicenter =randomized phase III trial of induction chemotherapy with cisplatin, 5-fluorouracil, ± docetaxel for larynx preservation, J Natl Cancer Inst 108 (2016).

[29] D.M. Hartl, G. Landry, F. Bidault, S. Hans, M. Julieron, G. Mamelle, et al., CT-scan prediction of thyroid cartilage invasion for early laryngeal squamous cell carcinoma, Eur Arch Otorhinolaryngol 270 (2013) 287-91.

[30] B. Li, M. Bobinski, R. Gandour-Edwards, D.G. Farwell, A.M. Chen, Overstaging of cartilage invasion by multidetector CT scan for laryngeal cancer and its potential effect on the use of organ preservation with chemoradiation, Br J Radiol 84 (2011) 64-9.

[31] G. Aresta, T. Araújo, S. Kwok, S. Chennamsetty, M. Safwan, V. Alex, et al., BACH: Grand challenge on breast cancer histology images, submitted to medical image analysis - Publication licensed under the Creative Commons CC-BY-NC-ND 40 license. (2018).

[32] P. Bandi, O. Geessink, Q. Manson, M. van Dijk, M. Balkenhol, M. Hermsen, et al., From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge, IEEE Trans Med Imaging (2018) 1-.

[33] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans Med Imaging 34 (2015) 1993-2024.

[34] A.A.A. Setio, A. Traverso, T. de Bel, M.S.N. Berens, C.V.D. Bogaard, P. Cerello, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge, Med Image Anal 42 (2017) 1-13.

[35] C. Jin, F. Shi, D. Xiang, X. Jiang, B. Zhang, X. Wang, et al., 3D Fast automaticsegmentation of kidney based on modified AAM and random forest, IEEE Trans Med Imaging 35 (2016) 1395-407.

[36] X. Chen, R.M. Summers, M. Cho, U. Bagci, J. Yao, An automatic method for renal cortex segmentation on CT images: evaluation on kidney donors, Acad Radiol 19 (2012) 562-70.

[37] U. Yoruk, B.A. Hargreaves, S.S. Vasanawala, Automatic renal segmentation for MR urography using 3D-GrabCut and random forests, Magn Reson Med 79 (2018) 1696-707.

[38] D. Xiang, U. Bagci, C. Jin, F. Shi, W. Zhu, J. Yao, et al., CorteXpert: a model-based method for automatic renal cortex segmentation, Med Image Anal 42 (2017) 257-73.

[39] The Digital Mammography DREAM Challenge; 2016. Available from: https://www.synapse.org/#!Synapse:syn4224222/wiki/401743.

[40] T. Araujo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, et al., Classification of breast cancer histology images using convolutional neural networks, PloS One 12 (2017) e0177544.

[41] J. Arevalo, F.A. Gonzalez, R. Ramos-Pollan, J.L. Oliveira, M.A. Guevara Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput Methods Programs Biomed 127 (2016) 248-57.

[42] J. Shan, S.K. Alam, B. Garra, Y. Zhang, T. Ahmed, Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods, Ultras Med Bio 42 (2016) 980-8.

[43] F. Liu, Z. Zhou, A. Samsonov, D. Blankenbaker, W. Larison, A. Kanarek, et al., Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection, Radiology 289 (2018) 160-9.

Figure Captions

Figure 1. For each of the 5 challenges, the score was calculated following different selection criteria.

Figure 2. For the 5 challenges, the number of images sent was monitored between June and September 2018.

Figure 3. For the meniscal tears challenge, the images were resized. Figures show example of change in image resolution (a) Image of 256 × 128 pixels. (b) Image of 256 × 256 pixels. (c) Image of 384 × 256 pixels.

Figure 4. The images of the liver lesion challenge were processed to facilitate the analysis. Figures show image before and after processing. (a) Ultrasound image before pre-processing. (b) Ultrasound image after pre-processing.

Table 1. Pixel size and image size after pre-processing.

Table 2. Comparison between images received and images kept for each dataset.

Table 3. Train, validation, and test split.

Table 4. Scores of each teams for the challenges.

Renal Cortex (CT scan):

$$\text{score} = \text{Dice}_{\text{segmentation}}$$

Breast lesions (MRI):

$$\text{score} = 0.6 * \text{AUC}_{\text{bening/malignant}} + \frac{0.4}{4} \sum_{\text{lesiontype}} \text{AUC}_{\text{lesiontype}}$$

Liver lesions (US):

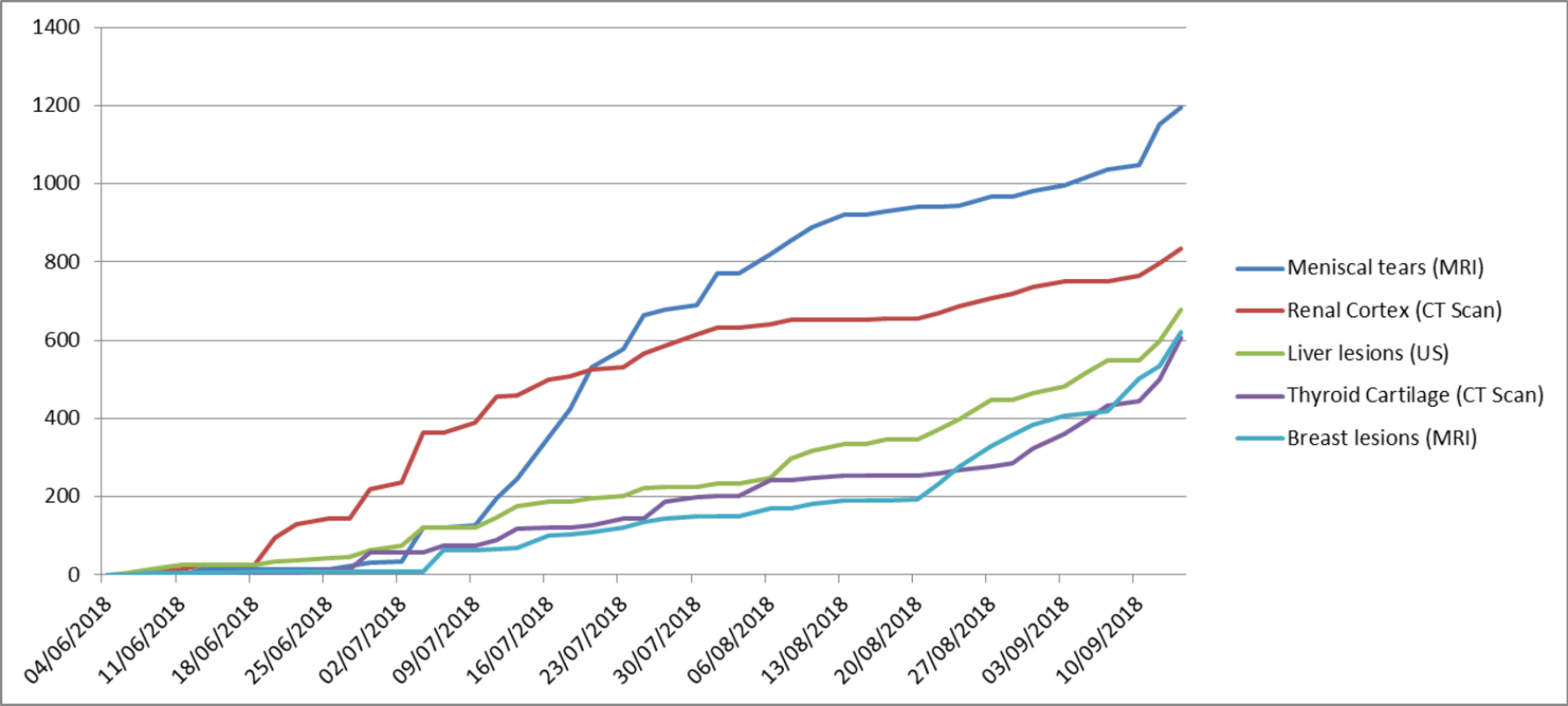
$$\text{score} = 0.5 \cdot \text{AUC}_{\text{lesiondetection}} + 0.3 \cdot \text{AUC}_{\text{benign/malignant}} + \frac{0.2}{6} \sum_{\text{lesiontype}} \text{AUC}_{\text{lesiontype}}$$

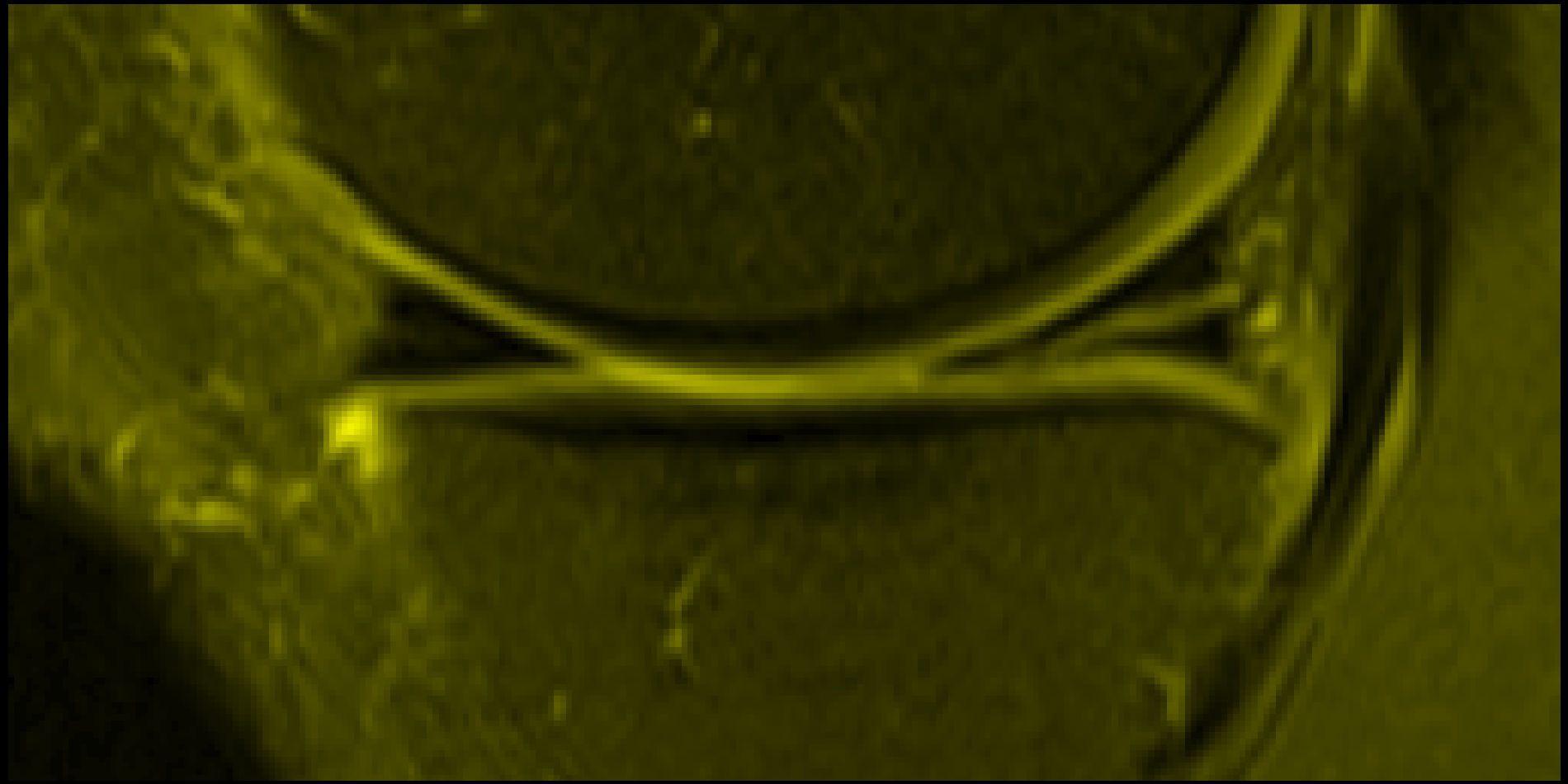
Meniscal tears (MRI):

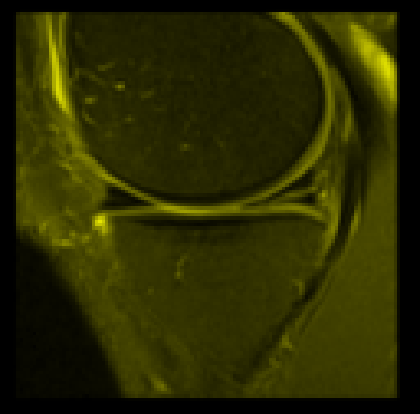
$$\text{score} = 0.4 \cdot \text{AUC}_{\text{tearsdetection}} + 0.3 \cdot \text{AUC}_{\text{tearsposition}} + 0.3 \cdot \text{AUC}_{\text{tearsorientation}}$$

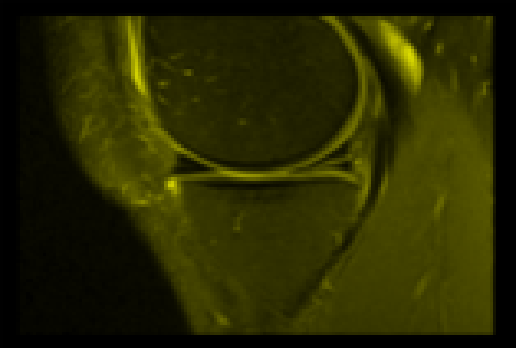
Thyroid cartilage (CT scan):

$$\text{score} = \text{AUC}_{\text{normal/abnormale}}$$


1400
1200
1000
800
600
400
200
0
04/06/2018
11/06/2018
18/06/2018
25/06/2018
02/07/2018
09/07/2018
16/07/2018
23/07/2018
30/07/2018
06/08/2018
13/08/2018
20/08/2018
27/08/2018
03/09/2018
10/09/2018
Meniscal tears (MRI)
Renal Cortex (CT Scan)
Liver lesions (US)
Thyroid Cartilage (CT Scan)
Breast lesions (MRI)








0
5
10
14

assumed.csv | HNGR.csv | saving_train_parameter.csv | d

```
Name,Path,Upload User,Upload Time,Size,Page ID,Blo
Carnet-dadresse.xlsx,/wp-content/uploads/menisque/
61019634.txt,/wp-content/uploads/menisque/médial/1
61019634.dicom,/wp-content/uploads/menisque/médial
61019634-dcm_.dcm,/wp-content/uploads/menisque/méd
61019634.dcm,/wp-content/uploads/menisque//2/,Phil
61019634.dcm,/wp-content/uploads/menisque/médial/1
Anonymat-Seq17-Ser7_-Img82.dcm,/wp-content/uploads
Anonymat-Seq17-Ser7_-Img82.dcm,/wp-content/uploads
Anonymat-Seq19-Ser701-Img124.dcm,/wp-content/uploa
Anonymat-Seq19-Ser701-Img124.dcm,/wp-content/uploa
1-Seq19-Ser701-Img127.dcm,/wp-content/uploads/sein
2-Seq19-Ser701-Img127.dcm,/wp-content/uploads/sein
3-Seq19-Ser701-Img125.dcm,/wp-content/uploads/sein
4-Seq19-Ser701-Img125.dcm,/wp-content/uploads/sein
Anonymat-Seq5_-Ser602-Img56.dcm,/wp-content/upload
Anonymat-Seq5_-Ser602-Img59.dcm,/wp-content/upload
Anonymat-Seq5_-Ser602-Img68.dcm,/wp-content/upload
Anonymat-Seq7_-Ser602-Img47.dcm,/wp-content/upload
Anonymat-Seq8_-Ser602-Img27.dcm,/wp-content/upload
Anonymat-Seq8_-Ser602-Img31.dcm,/wp-content/upload
I1_30000.dcm,/wp-content/uploads/cortex_renal/29/,
I1_40000.dcm,/wp-content/uploads/cortex_renal/29/,
I1_40000.dicom,/wp-content/uploads/cortex_renal/29
I1T_0000.dcm,/wp-content/uploads/cortex_renal/29/,
I1000000.dicom,/wp-content/uploads/cortex_renal/28
I1000000-dcm_.dcm,/wp-content/uploads/cortex_renal
I201806121455498340001.bmp,/wp-content/uploads/foi
I201806121455501100002.bmp,/wp-content/uploads/foi
0901020008.dcm,/wp-content/uploads/foie/ss_caliper
I201806130901282940001.bmp,/wp-content/uploads/foi
0902110020.dcm,/wp-content/uploads/foie/ss_caliper
0903150031.dcm,/wp-content/uploads/foie/ss_caliper
I201806130947575610001.bmp,/wp-content/uploads/foi
I1_30000.dicom,/wp-content/uploads/cortex_renal/28
I1_40000.dicom,/wp-content/uploads/cortex_renal/28
Anonymat-Seq5_-Ser602-Img65.dcm,/wp-content/upload
Anonymat-Seq5_-Ser602-Img84.dcm,/wp-content/upload
36924660.dicom,/wp-content/uploads/sein/normal/30/
Anonymat-Seq5_-Ser602-Img47.dcm,/wp-content/upload
I1000000.dcm,/wp-content/uploads/cortex_renal/29/,
66625274.dcm,/wp-content/uploads/cortex_renal/74/,
```

assumed.csv | HNGR.csv | saving_train_parame

```
id,Bénin,Lésion,Malin,Type de lésion
foie_0,0,0,0,Foie Homogène
foie_1,0,0,0,Foie Homogène
foie_2,0,0,0,Foie Homogène
foie_3,0,0,0,Foie Homogène
foie_4,0,0,0,Foie Homogène
foie_5,0,0,0,Foie Homogène
foie_6,0,0,0,Foie Homogène
foie_7,0,0,0,Foie Homogène
foie_8,0,0,0,Foie Homogène
foie_9,0,0,0,Foie Homogène
foie_10,0,0,0,Foie Homogène
foie_11,0,0,0,Foie Homogène
foie_12,0,0,0,Foie Homogène
foie_13,0,0,0,Foie Homogène
foie_14,0,0,0,Foie Homogène
foie_15,0,0,0,Foie Homogène
foie_16,0,0,0,Foie Homogène
foie_17,0,0,0,Foie Homogène
foie_18,0,0,0,Foie Homogène
foie_19,0,0,0,Foie Homogène
foie_20,0,0,0,Foie Homogène
foie_21,0,0,0,Foie Homogène
foie_22,1,1,0,Angiome
foie_23,1,1,0,Angiome
foie_24,0,0,0,Foie Homogène
foie_25,1,1,0,Angiome
foie_26,1,1,0,Angiome
foie_27,0,1,1,Métastase
foie_28,0,1,1,Métastase
foie_29,0,1,1,Métastase
foie_30,1,1,0,Angiome
foie_31,1,1,0,Angiome
foie_32,1,1,0,Angiome
foie_33,0,1,1,CHC
foie_34,0,1,1,CHC
foie_35,0,1,1,CHC
foie_36,1,1,0,Kyste
foie_37,0,1,1,Métastase
foie_38,0,1,1,CHC
foie_39,0,1,1,Métastase
foie_40,0,1,1,Métastase
foie_41,0,1,1,Métastase
```

8: Fichier /home/theoestienne/Documents/data/CURATED_FI

	Pixel Size	**Image size**
Cortex Renal	1 mm	192 * 192
Breast	0.7 mm	440 * 440
Meniscus fissure	0.332 mm	256 * 256
Thyroid cartilage	0.45 mm	192*192
Liver	/	/

	Number of images received	Number of images kept in dataset	Percentage
Renal cortex	835	787	94%
Breast	622	504	81%
Meniscus fissure	2389	1823	76%
Thyroid cartilage	606	511	84%
Liver	678	545	80%
TOTAL	5130	4170	/

	Number of images received	Number of images kept in dataset	Percentage
Renal cortex	835	787	94%
Breast	622	504	81%
Meniscus fissure	2389	1823	76%
Thyroid cartilage	606	511	84%
Liver	678	545	80%
TOTAL	5130	4170	/

Teams	Renal Cortex	Meniscal Tears	Liver Lesions	Thyroid Cartilage	Brest Lesions
Airion		0.86			
Alphamed			0.76		
Data Med Team			0.72		
DataBC Team					0.58
GBC	0.86				
IBM-GHICL		0.52			
ICM X Rays	0.85				
Illuin in the Deep		0.85			
Incepto	0.84	0.55	0.55		
Inria-Liryc	0.87				
Kynapse		0.68			
Laveran	0.86 (Late)				0.50
Lis-Clinique St Jean Team		0.86	0.60	0.65	
LyPhte	0.87	0.91			
MinskX		0.48			
Nhance			0.80		
Owkinautes	0.86		0.90		0.82
PyRATS					0.53
Quantmetry		0.73			
R2D2		0.15			
RadioAdvisor		0.90			
RGR			0.78		
SynovIA		0.83		0.70	