Generative adversarial networks (GAN)-based data augmentation of rare liver cancers: The SFR 2021 Artificial Intelligence Data Challenge

Sébastien Mulé, Littisha Lawrance, Younes Belkouchi, Valérie Vilgrain, Maité Lewin, Hervé Trillaud, Christine Hoeffel, Valérie Laurent, Samy Ammari, Eric Morand, et al.

Generative adversarial networks (GAN)-based data augmentation of rare liver cancers: The SFR 2021 Artificial Intelligence Data Challenge

Authors

Sébastien Mulé [a,b*], Littisha Lawrance [c], Younes Belkouchi [c,d], Valérie Vilgrain [e,f], Maité Lewin [g,h], Hervé Trillaud [i], Christine Hoeffel [j], Valérie Laurent [k], Samy Ammari [c,l], Eric Morand[m], Orphee Faucoz [m], Arthur Tenenhaus [n], Anne Cotton [o,p], Jean-François Meder [q], Hugues Talbot [d], Alain Luciani [a,b], Nathalie Lassau [c, l]

Affiliations

[a] Medical Imaging Department, AP-HP, Henri Mondor University Hospital, 94000 Créteil, France

[b] INSERM, U955, Team 18, 94000 Créteil, France

[c] Laboratoire d'Imagerie Biomédicale Multimodale Paris-Saclay, BIOMAPS, UMR 1281, Université Paris-Saclay, Inserm, CNRS, CEA, 94800 Villejuif, France

[d] OPIS–Optimisation Imagerie et Santé, Université Paris-Saclay, Inria, CentraleSupélec, CVN-Centre de vision numérique, 91190 Gif-Sur-Yvette, France

[e] Department of Radiology, APHP, University Hospitals Paris Nord Val de Seine, Hopital Beaujon, 92110 Clichy, France.

[f] Université Paris Cité, CRI INSERM, 75018 Paris, France

[g] Department of Radiology, AP-HP Hôpital Paul Brousse, 94800 Villejuif, France.

[h]Faculté de Médecine, Université Paris-Saclay, 94270 Le Kremlin-Bicêtre, France.

[i]CHU de Bordeaux, Department of Radiology, Université de Bordeaux, 33000 Bordeaux, France

[j] Department of Radiology, Reims University Hospital, 51092 Reims, France; CRESTIC, University of Reims Champagne-Ardenne, 51100 Reims, France

[k] Department of Radiology, Nancy University Hospital, University of Lorraine, 54500 Vandoeuvre-lès-Nancy, France

[l] Department of Imaging, Institut Gustave Roussy, Université Paris-Saclay, 94800 Villejuif, France.

[m] Centre National d'Etudes Spatiales–CNES, Centre Spatial de Toulouse, 31401 Toulouse Cedex 9 - France

[n] Université Paris-Saclay, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190 Gif-sur-Yvette, France

[o] Division of Musculoskeletal Radiology, Lille University Hospital Center, Centre de Consultations Et D'imagerie de L'appareil Locomoteur, 59037 Lille, France.
[p] Lille University School of Medicine, Lille, France
[q] Department of Neuroimaging, Sainte-Anne Hospital, 75013 Paris, France; Université Paris Cité, 75006 Paris, France

[*]**Corresponding author:** sebastien.mule@aphp.fr

Generative adversarial networks (GAN)-based data augmentation of rare liver cancers: The SFR 2021 Artificial Intelligence Data Challenge

Abstract

Purpose The 2021 edition of the Artificial Intelligence) Data Challenge was organized by the French Society of Radiology (SFR) together with the Centre National d'Études Spatiales and CentraleSupélec with the aim to implement generative adversarial networks (GANs) techniques to provide 1000 magnetic resonance imaging (MRI) cases of macrotrabecular-massive (MTM) hepatocellular carcinoma (HCC), a rare and aggressive subtype of HCC, generated from a limited number of real cases from multiple French centers.

Materials and methods A dedicated platform was used by the seven inclusion centres to securely upload their anonymized MRI examinations including all three cross-sectional images (one late arterial and one portal-venous phase T1-weighted images and one fat-saturated T2-weighted image) in compliance with general data protection regulation. The quality of the database was checked by experts with manual delineation of the lesions performed by the expert radiologists involved in each center. Multidisciplinary teams competed between October 11th, 2021 and February 13th, 2022.

Results A total of 91 MTM-HCC datasets of three images each were collected from seven French academic centers. Six teams with a total of 28 individuals participated in this challenge. Each participating team was asked to generate one thousand 3-image cases. The qualitative evaluation was done by three radiologists using the Likert scale on ten cases generated by each participant. A quantitative evaluation was also performed using two metrics, the Frechet inception distance and a leave-one-out accuracy of a 1-Nearest Neighbor algorithm.

Conclusion This data challenge demonstrates the ability of GANs techniques to generate a large number of images from a small sample of imaging examinations of a rare malignant tumor.

Keywords

Artificial intelligence;

Deep learning;

Generative adversarial networks;
Liver cancer;
Magnetic resonance imaging;

Abbreviations

AI: Artificial intelligence
CNES: Centre National d'Études Spatiales
CNN: convolutional neural network
DL: Deep learning
FID: Frechet inception distance
GAN: Generative adversarial networks
GDPR: General Data Protection Regulation
HCC: Hepatocellular carcinoma
MRI: Magnetic resonance imaging
MTM: Macrotrabecular-massive
SFR: French Society of Radiology

1. Introduction

The recent success and exponential use of artificial intelligence (AI) in medical imaging is largely due to the successful application of deep learning to labeled big data. Notably, convolutional neural networks (CNNs) have great capabilities in image recognition tasks [1]. Since 2018, a total of 11 data challenges have been organized, with two on ultrasound images, three on magnetic resonance imaging (MRI) and six on computed tomography images [2–4]. The previous data challenges led by the French Society of Radiology (SFR) demonstrated high performances of CNNs in lesion detection, segmentation and classification tasks, applied to cervical lymphadenopathies [5], pulmonary nodules [6] or breast nodules [7].

However, the performance of CNNs may be limited when data is sparse, as it is the case for a rare disease. In such situations, a model can be trained to the point of perfectly predicting

labels on the training data but poorly on independent test data, reflecting an over fitted and poorly generalizable model [8]. In that context, generative adversarial networks (GANs) have recently gained great interest. GANs have the potential to increase the number of training images by creating fake images that look like real images [9].

The 2021 edition of the Artificial Intelligence Data Challenge organized by the French Society of Radiology together with the Centre National d'Etudes Spatiales (CNES) focused on the macrotrabecular-massive (MTM) subtype of hepatocellular carcinoma (HCC), a rare and aggressive type of primary liver cancer with poor prognosis [10,11]. MTM-HCC displays suggestive imaging features on contrast-enhanced MRI including substantial necrosis and diffuse hypovascular component [12,13].

The purpose of this data challenge was to create a synthetic dataset of 1000 MTM-HCC cases from a limited number of real cases using GAN-based data augmentation techniques.

2. Material and methods

2.1 Clinical questions

HCC represents the majority of primary liver cancers and is the fourth leading cause of cancer-related mortality worldwide. HCC is a heterogeneous group of tumors, not only in terms of clinical and molecular features but also prognosis. The MTM subtype of HCC was recently introduced in the fifth edition of the World Health Organization classification of digestive tumors [14]. MTM-HCC represents an aggressive form of HCC and is associated with poor survival. Interestingly, it can be identified with high specificity on contrast-enhanced MRI. Substantial necrosis, defined as tumor necrosis occupying at least 20% of the tumor, may predict MTM-HCCs with 90% specificity [12]. Hence, its low incidence, poor prognosis and specific imaging findings make MTM-HCC the perfect candidate for this data challenge focused on data augmentation techniques in oncology. Since substantial necrosis can be detected using T1-weighted images obtained during arterial and portal-venous phase of enhancement, and T2-weighted images, it was decided to select for each case included in the data challenge three cross-sectional images, namely one image from each of these three MRI sequences.

2.2 Security and data protection

Each step of the challenge, from data uploading to the challenge phase, was performed according to the General Data Protection Regulation of the European Union. The French Commission Nationale Informatique et Libertés was consulted, and the SFR assumed the role of Data Protection Officer. DICOM images were pseudo-anonymized before uploading to the dedicated platform not conserving any data on the patient. The patient ID was replaced with an anonymized ID to identify the image slices that belonged to the same patient. This process ensured the protection of patients' data before the participants had access to it. Every radiologist involved was asked to send an information letter to patients about the use of their medical examinations, with the option of refusal of consent. A data chart was also sent to each radiologist to help them abide by the GDPR rules, as well as guidelines on the terms of use of the data for the participants to the challenge. The data collected could only be used for the aim of this challenge by the participants. A charter was signed by the participating teams to ensure this.

2.3 Communication and uploading

The data uploading phase began on October 4th, 2021. Seven French academic centers were identified by the French Society of Abdominal Imaging (SIAD): CHU Henri Mondor, AP-HP, Créteil; CHU Beaujon, AP-HP, Clichy; CHU Paul Brousse, AP-HP, Villejuif; Institut Gustave Roussy, Villejuif; CHU Nancy, Nancy; CHU Reims, Reims; CHU Bordeaux, Bordeaux. The technical and clinical specifications of the question and the format of the medical data to be uploaded were communicated to the radiologists of these centers. The radiologists had to register on the platform if they had cases of MTM-HCC and upload cases to the dedicated platform as specified per the requirements. Our aim was to have an all-in-one interface through the platform for radiologists, to both upload and annotate while maintaining a uniform format throughout the dataset. They were sent a tutorial if required and provided with support emails and phone numbers. A follow-up of their progress was also done to ensure a non-erroneous dataset. The uploaded data was monitored on a daily basis to check the conformity of the medical image examinations and the datasets were checked by an expert before sending it to the participants.

2.4 Team gathering and challenge phase

Each team was requested to have a multidisciplinary member background with at least one radiologist, one engineer/data scientist and an engineering/PhD student. The team could have been a startup, a big company or a research lab. Each team was required to register on the platform with all the members' details. Three datasets were sent to them for each challenge: the first dataset batch was sent on October 11th, 2021 after the launch of the Data Challenge during the JFR 2021 edition and the second on December 13th, 2021 for the teams to train their algorithms. Finally, the validation dataset was sent on January 12th, 2022. The deadline for the submission of the 1000 generated cases back to the platform was February 13th, 2022 at 1 pm. The jury included three radiologists for the qualitative analysis and researchers from CentraleSupélec for the quantitative analysis, and the winners were announced on April 7th, 2022. A prize of €3000 was awarded to the winner of the challenge.

3. Results

3.1. Communication and team gathering

Six teams participated in the data challenge. Twenty-eight team members participated in this challenge, including six PhD scholars, two radiologists and 20 employees. The upload phase began in October, and the first images were communicated to the teams by mid-October, 2021. A total of 279 images were uploaded from seven French academic centers (Table 1).

3.2. Cases generated

From the six participating teams, one did not provide any generated cases. The other participants successfully generated three MR images for each case. Only one group generated 100 patients, while the rest succeeded in generating 1000 patients. An example of generated cases is presented in Figure 1.

3.3. Score computation

The evaluation was split into two categories. Firstly, a quantitative evaluation based on large-scale statistics to score the generated cases created by the contestants. Secondly, a qualitative score based on the 10 best cases (30 MR images) was selected using mathematical models, and scored by three expert radiologists. This was done to evaluate the validity of the image from a radiologist's perspective. The mean of the scores obtained from each evaluation was taken to be the final score.

3.3.1. Quantitative analysis

The quantitative scoring relied on two methods to establish the final quantitative score [15,16], which were the Frechet inception distance (FID) [17] and a diversity measure using the 1-nearest neighbor algorithm [16].

Simple preprocessing was done on the images before running the evaluations. Max-min normalization was applied on each image to bring it to the range (0, 1), and they were resized using nearest neighbor interpolation to preserve intensities to a size of 299x299 pixels (Input size of the Inception v3 [17] model). Each case was evaluated by creating a three-channel tensor composed of the three MRI images generated as per the challenge, T1-weighted arterial phase, T1-weighted portal venous phase and T2-weighted images, successively. The same concatenation was also done for the real images. The tensors were then fed to pre-trained Inception v3 convolution layers to obtain the X_r and X_r real cases and generated case embeddings respectively**,** which were used for both methods (Figure 2).

The FID between two multidimensional Gaussian distributed samples was computed using the following formula:

$$FID = \left\|\mu_r - \mu_g\right\|^2 + Tr(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}})$$

where μ_r and μ_g represent the mean (magnitude), Σ_r and Σ_g represent the covariances of the embedding X_r and X_g respectively. Tr is the algebraic application trace. For the computation of the FID [16,17], X_r and X_g are considered gaussians i.e. $X_r \sim N(\mu_r, \Sigma_r)$ and $X_g \sim N(\mu_g, \Sigma_g)$.
Being a distance, the FID has no upper bound ($FID \in [0, +\text{inf}[$) and is hence unusable as is to provide a score for the contestants. In order to transform it into a score, a simple normalization transformation was used:

$$FID_{score} = 1 - \frac{FID}{FID_{max}}$$

where FID_{max} is equal to the maximal FID obtained by all the contestants.

As for the diversity measure, the dataset was divided into 10 batches of 100 case embeddings. The euclidean pairwise distance matrix for each sample in a batch and the Leave One Out (LOO) accuracy using a 1-nearest neighbor algorithm were computed. An accuracy of 50% implied a high diversity in the cases generated, whilst an accuracy of 100% or 0% either meant a mode collapse (*i.e.*, the same case was always generated, or the model learned to generate the real images only) [16]. The score was modified to measure how close the accuracy was to 50%, using the following transform:

$$Diversity_{score} = 1 - 2|Accuracy - 0.5|$$

where $Accuracy$ is the average accuracy over all batches.

In order to mitigate the sampling bias from the creation of batches, this method was repeated 100 times, the final score being the average $Diversity_{score}$ over all samples.
Thus, the quantitative score was the mean of the $Diversity_{score}$ and the FID_{score}, or simply:

$$Quantitative_{score} = \frac{1}{2}(Diversity_{score} + FID_{score})$$

3.3.2. Qualitative analysis

For the qualitative assessment of the generated cases, three radiologists viewed 10 selected cases per contestant and assigned a score of 0 to 4 to the five defined criteria presented in Table 2. All the cases were randomized prior to the radiologist's assessments.

The 10 cases were selected using the embeddings obtained from the quantitative assessment. A K-Means clustering method was fit to produce 10 clusters of images for each contestant, and the points closest to the centers of the clusters were selected for evaluation (Figure 3).

Each case had a score in the range [0, 20], and the average of all cases made up the score of the contestant team for a single radiologist. The final qualitative score was calculated as the average of all three marks given by the radiologist. In summary:

$$Quanlitative_{score} = \frac{1}{3}(\sum_{i=0}^{2} Qualitative_{score,i})$$

Where $Qualitative_{score,i}$ is the average of the score of all 10 cases for radiologist i.

3.4. Data processing

Only a few preprocessing operations were performed on the data, such as automatic anonymisation of the examinations and annotation. The aim was to train the AI models on original data without much preprocessing to not lose further information from the images. A total of 92 cases were uploaded among which one unusable case was excluded. All the three images per case were annotated with a mask for each image specifying the lesion. The images were converted as python numpy images and the associated mask was also sent to the participants. The dataset was sent in three different batches, containing 30, 30 and 32 cases respectively. On April 7th, 2022 the results were announced based on the scores computed and the team from Philips had the best score combining the qualitative and qualitative analyses (Score = 0.64).

4. Discussion

Data augmentation is a widely used technique in every learning method, especially in CNN for image processing [18]. Its main purpose is to increase the amount of data to expand the number

of samples in a training dataset and thus increase the capability of learned models to generalize better. Studies suggest that with a large dataset, supervised learning performs and generalizes very well [19]. Therefore, one of the data augmentation solutions for using small datasets to be used in supervised learning is data generation [18]. The goal of this data challenge was to prove the feasibility of generating a large and diverse dataset (n = 1000 cases) from a very small sample (n = 91). This could have a major impact on the medical field, where collecting data is a very difficult task, considering the time it takes to build a database and the regulations that teams face to use patient data, or even for rare disease, as in this data challenge, making the availability of patient data scarce. The use of such a technique could very well revolutionize the medical domain if only a small sample of labeled data is needed to obtain a model that can generalize well.

MTM-HCC is a recently identified HCC subtype associated with poor survival [10, 11]. Developing AI-based algorithms able to identify this specific subtype during pre-therapeutic work-up may thus have strong prognostic and therapeutic implications. The low incidence of this type of tumor makes the ability to artificially generate a large number of cases of major importance.

This challenge was difficult given the fact that the participants were supposed to generate three MRI images, each corresponding to a specific sequence, for each of the 1000 cases. This implied that the body structure, as well as the tumor location, had to be preserved between sequences while modifying the contrast only. Multiple models exist that perform such transformations if masks (body mask, lesion mask) or the local region to transform is provided [20–22]. Most of these generative models thrive on a large number of training images, which means that participants had to find a way to adapt them to a small dataset. Most participants succeeded in generating these 1000 cases.

Moreover, the task of evaluating generative models is also a difficult one [16]. Multiple metrics exist in order to measure the similarity between two datasets of images, and the fidelity to the dataset used for training generative models. However, in most computer vision tasks, the most important factor is human perception. For this data challenge, we chose to couple a statistical method to evaluate a large number of images with a subjective task that relies on expert radiologists, to measure if the generated images were indistinguishable to their eyes. Coupling both methods seemed to provide the best scoring for this difficult task. Furthermore,

the best teams had a radiologist among their participants, suggesting that collaboration between data scientists and radiologists is essential to tackle problems involving medical expertise.

In conclusion, this data challenge that uses MR images of MTM-HCC demonstrates the ability of GANs techniques to generate a large number of images from a small sample of rare malignant tumor studies.

Human 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.

Acknowledgements

We would like to thank the Société Française de Radiologie for the opportunity to organise this challenge and for its support. We would also like to thank Laure Boyer and Marie Laure Gouzy (MEDES-IMPS, France) for their continuous involvement in the SFR/CNES partnership initiated in 2020.

Informed consent and patient details

The authors declare that they obtained 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.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Cheng PM, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón Romero F, et al. Deep learning: an update for radiologists. Radiographics 2021;41:1427–1445.

[2] Lassau N, Bousaid I, Chouzenoux E, Verdon A, Balleyguier C, Bidault F, et al. Three artificial intelligence data challenges based on CT and ultrasound. Diagn Interv Imaging 2021;102:669–674.

[3] Lassau N, Bousaid I, Chouzenoux E, Lamarque JP, Charmettant B, Azoulay M, et al. Three artificial intelligence data challenges based on CT and MRI. Diagn Interv Imaging 2020;101:783–788.

[4] 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.

[5] Courot A, Cabrera DLF, Gogin N, Gaillandre L, Rico G, Zhang-Yin J, et al. Automatic cervical lymphadenopathy segmentation from CT data using deep learning. Diagn Interv Imaging 2021;102:675–681.

[6] Blanc D, Racine V, Khalil A, Deloche M, Broyelle J-A, Hammouamri I, et al. Artificial intelligence solution to classify pulmonary nodules on CT. Diagn Interv Imaging 2020;101:803–810.

[7] Evain E, Raynaud C, Ciofolo-Veit C, Popoff A, Caramella T, Kbaier P, et al. Breast nodule classification with two-dimensional ultrasound using Mask-RCNN ensemble aggregation. Diagn Interv Imaging 2021;102:653–658.

[8] Nakaura T, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. Diagn Interv Imaging 2020;101:765–770.

[9] Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 2019;9:16884.

[10] Sessa A, Mulé S, Brustia R, Regnault H, Pregliasco AG, Rhaiem R, et al. Macrotrabecular-massive hepatocellular carcinoma: light and shadow in current knowledge. J Hepatocell Carcinoma 2022;9:661–670.

[11] Ziol M, Poté N, Amaddeo G, Laurent A, Nault J-C, Oberti F, et al. Macrotrabecular-massive hepatocellular carcinoma: a distinctive histological subtype with clinical relevance. Hepatology 2018;68:103–12.

[12] Mulé S, Galletto Pregliasco A, Tenenhaus A, Kharrat R, Amaddeo G, Baranes L, et al. Multiphase liver MRI for identifying the macrotrabecular-massive subtype of hepatocellular carcinoma. Radiology 2020;295:562–71.

[13] Rhee H, Cho ES, Nahm JH, Jang M, Chung YE, Baek SE, et al. Gadoxetic acid-enhanced MRI of macrotrabecular-massive hepatocellular carcinoma and its prognostic implications. J Hepatol 2021;74:109–21.

[14] WHO Classification of Tumours. Digestive System Tumours. 5th ed. Lyon, France: IARC, 2019: pp 229–239.

[15] Haarburger C, Horst N, Truhn D, Broeckmann M, Schrading S, Kuhl C, et al. Multiparametric magnetic resonance image synthesis using generative adversarial networks. Eurographics Workshop on Visual Computing for Biology and Medicine 2019.

[16] Xu Q, Huang G, Yuan Y, Guo C, Sun Y, Wu F, et al. An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint arXiv:1806.07755, 2018.

[17] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500, 2017.

[18] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6:1–48.

[19] Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. IEEE Intell Syst 2009;24:8–12.

[20] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks 2018. https://doi.org/10.48550/arXiv.1611.07004.

[21] Sohn K, Lee H, Yan X. Learning structured output representation using deep conditional generative models. Adv Neural Inf Process Syst 2015;28.

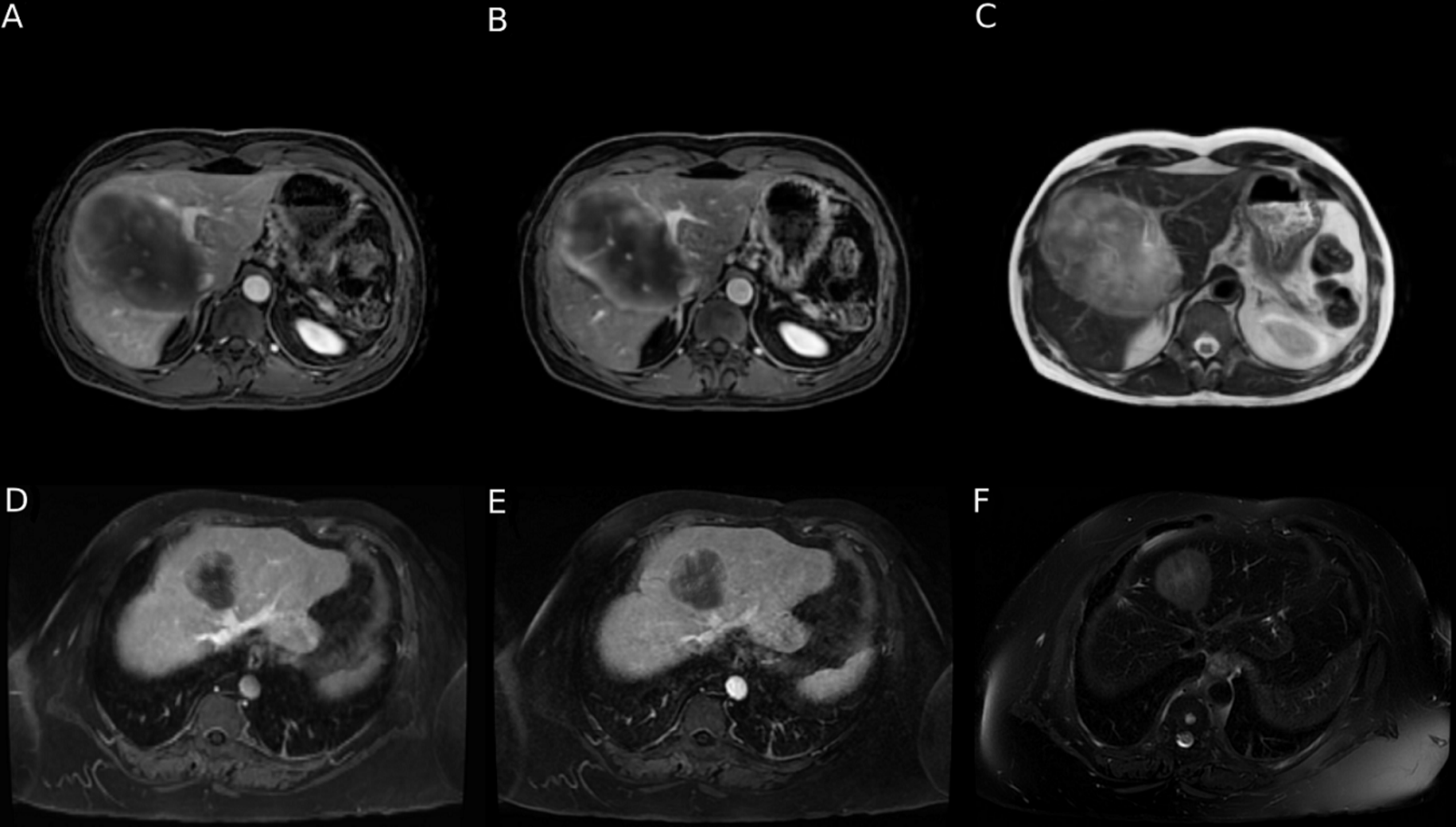
[22] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. In: International conference on machine learning. PMLR, 2017. pp: 2642–2651.

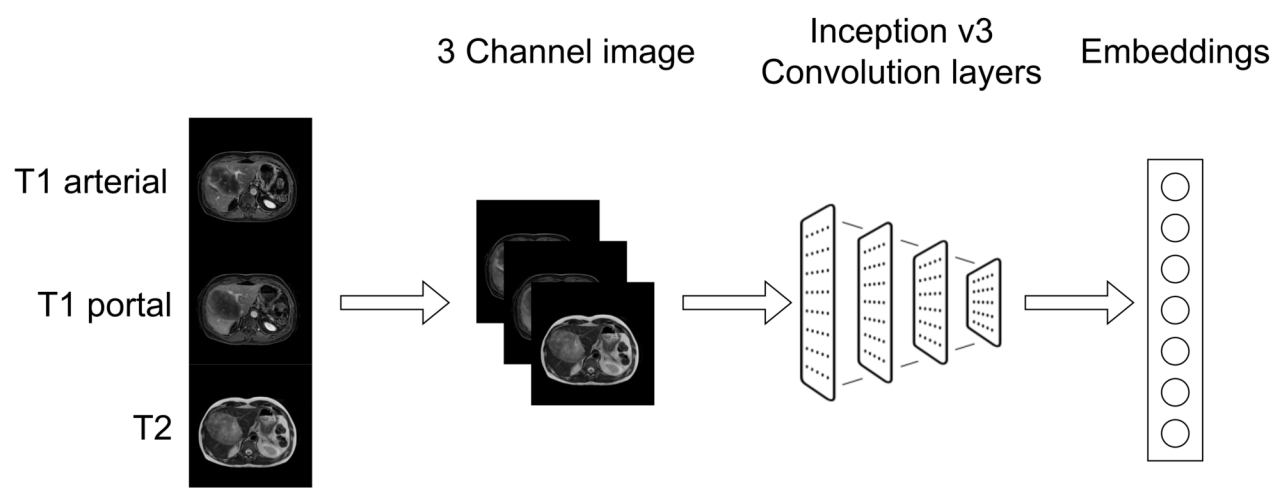
Figure legends

Figure 1: Two generated cases by the winning teams that obtained the highest scores with radiologists qualitative evaluation. Each row represents a patient, A-D) Fat-saturated T1-weighted MR images obtained during the arterial phase of enhancement; B-E) Fat-saturated T1-weighted MR images obtained during the portal-venous phase of enhancement; C-F) T2-weighted images.

Figure 2: Case embeddings obtained using a pre-trained Inception v3. The three generated images for a single patient are concatenated to create a three-channel image consisting of the T1-weighted arterial and portal-venous images, and then T2-weighted image. A feed-forward is applied on the three-channel image to obtain the embedding using Inception v3 convolution layers.

Figure 3: t-SNE representation of the embeddings of all contestants' generated cases and the real cases in the real plane. The large dots represent the data points that were selected per contestant.


A
B
C
D
E
F


3 Channel image
Inception v3
Convolution layers
Embeddings
T1 arterial
T1 portal
T2

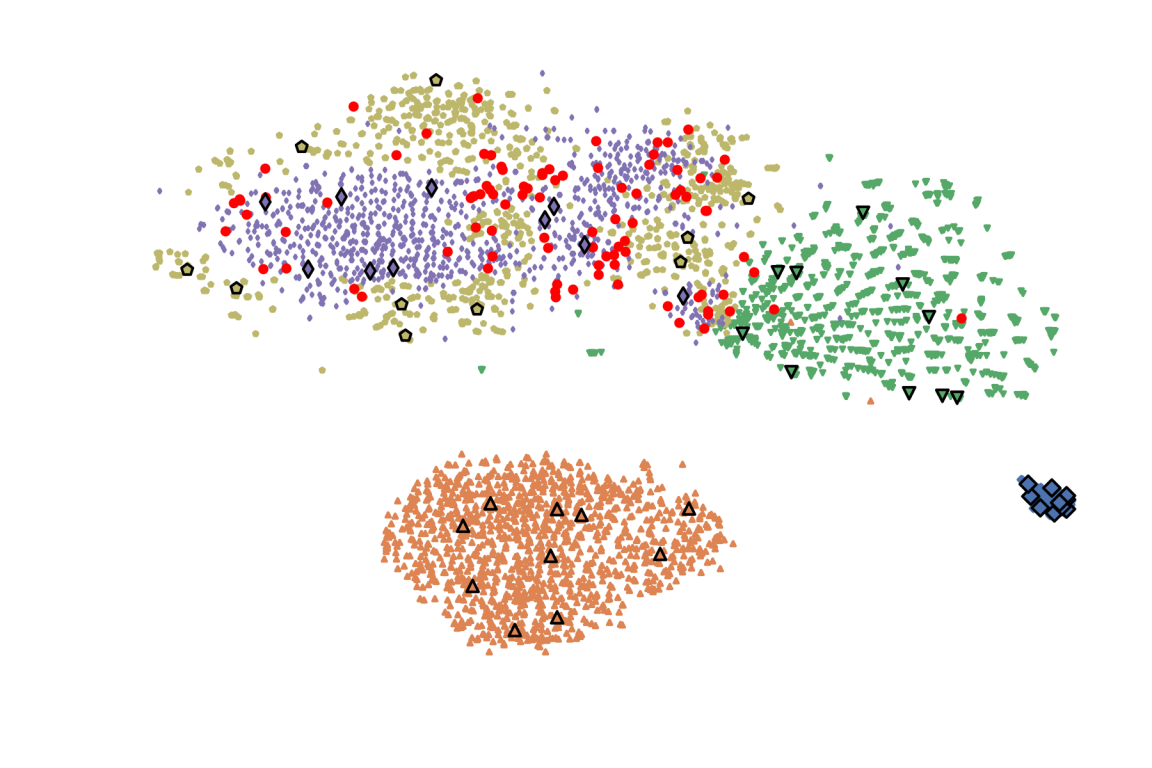


Table 1: Participating centers and numbers of included cases of macrotrabecular-massive hepatocellular carcinoma

Participating center	Number of included cases
AP-HP Henri-Mondor, University Hospital, Créteil, France	43
AP-HP Beaujon, University Hospital, Clichy, France	24
AP-HP Paul Brousse, University Hospital, Villejuif, France	13
Bordeaux University Hospital CHU, Bordeaux, France	7
Reims University Hospital, Reims, France	3
Institut Gustave Roussy, University Hospital, Villejuif, France	1
Nancy University Hospital, Nancy, France	1
Total	92

Table 2. The five criteria used by the radiologists to evaluate the selected cases. The criteria aim to assess the quality of the generated images and their conformity to a macrotrabecular-massive hepatocellular carcinoma.

Criteria	**Score**
Acceptable liver morphology	0–4
Tumor morphology consistent with that of a HCC	0–4
Lesion contrast enhancement patterns consistent with that of a HCC	0–4
Presence of substantial necrosis	0–4
Consistent appearance of each generated MRI images with the expected ones (T1-weighted arterial and portal-venous images, and T2-weighted image)	0–4
Qualitative score of a single three-MR image case by each radiologist	Σ (0-20)

HCC = Hepatocellular carcinoma; MRI = Magnetic resonance imaging