



HAL
open science

GUIDING THE CLASSIFICATION OF HEPATOCELLULAR CARCINOMA ON 3D CT-SCANS USING DEEP AND HANDCRAFTED RADIOLOGICAL FEATURES

Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, Christophe Aubé,
Maxime Ronot, Pietro Gori, Isabelle Bloch

► **To cite this version:**

Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, Christophe Aubé, Maxime Ronot, et al.. GUIDING THE CLASSIFICATION OF HEPATOCELLULAR CARCINOMA ON 3D CT-SCANS USING DEEP AND HANDCRAFTED RADIOLOGICAL FEATURES. IEEE ISBI 2025 : 2025 IEEE International Symposium on Biomedical Imaging, Apr 2025, Houston, United States. hal-04886299

HAL Id: hal-04886299

<https://telecom-paris.hal.science/hal-04886299v1>

Submitted on 14 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

GUIDING THE CLASSIFICATION OF HEPATOCELLULAR CARCINOMA ON 3D CT-SCANS USING DEEP AND HANDCRAFTED RADIOLOGICAL FEATURES

E. Sarfati^{1,2} A. Bône¹ M-M. Rohé¹ C. Aubé⁴ M. Ronot^{5,6} P. Gori² I. Bloch^{2,3}

¹ Guerbet Research, Villepinte, France

² LTCI, Télécom Paris, Institut Polytechnique de Paris, Saclay, France

³ Sorbonne Université, CNRS, LIP6, Paris, France

⁴ Département de Radiologie, Centre Hospitalier Universitaire d'Angers, Angers, France

⁵ Service de Radiologie, HUPNSV, Hôpital Beaujon, Clichy, France

⁶ INSERM UMR 1149, Université de Paris, Paris, France

ABSTRACT

Hepatocellular carcinoma is the most spread primary liver cancer across the world (~80% of the liver tumors). The gold standard for HCC diagnosis is liver biopsy. However, in the clinical routine, expert radiologists provide a visual diagnosis by interpreting hepatic CT-scans according to a standardized protocol, the LI-RADS, which uses five radiological criteria with an associated decision tree. In this paper, we propose an automatic approach to predict histology-proven HCC from CT images in order to reduce radiologists' inter-variability. We first show that standard deep learning methods fail to accurately predict HCC from CT-scans on a challenging database, and propose a two-step approach inspired by the LI-RADS system to improve the performance. We achieve improvements from 6 to 18 points of AUC with respect to deep learning baselines trained with different architectures. We also provide clinical validation of our method, achieving results that outperform non-expert radiologists and are on par with expert ones.

Index Terms— Deep Learning, CT imaging, Image Classification, Hepatocellular Carcinoma, Liver, LI-RADS.

1. INTRODUCTION

Liver cancer is the sixth most common cancer worldwide and represents the fourth cause of mortality in 2023 [1]. In particular, hepatocellular carcinoma (HCC) accounts for 80% of the primary liver cancers [1, 2]. Diagnosing the presence of hepatocellular carcinoma is obtained with a biopsy, which is a risky and invasive operation. On the other hand, radiological assessment can provide a non-invasive diagnosis, based on typical features on dynamic CT or MRI according to strict criteria described in international guidelines [3, 4, 5]. This procedure is considered a solid proxy of the presence or absence of the tumor, and often totally replaces liver biopsy in clinical routine. To perform this assessment, radiologists use a standardized scoring system called LI-RADS (Liver Imaging Reporting and Data System) [6] to characterize suspicious lesions on CT-scans. The latter uses the four phases of the traditional injection protocol for CT imaging, and is based on five main visual features. However, radiological diagnosis suffers from an inherent variability between experts, which can be reduced using automatic methods. Several methods have been proposed in the literature to improve the diagnosis of hepatocellular carcinoma using deep learning [5, 7, 8, 9, 10, 11]. The authors in [8, 12] predict directly the pathology - without using any intermediate features - but exploit very large medical pri-

vate datasets [12] (~12,000 patients in their study) that present a huge size difference between non-HCC and HCC lesions, which is a well-known feature to characterize HCC but can induce a bias in the problem setting as their HCC lesion can be considered “easy” or un-useful in the clinical process, due to the advanced stage of the tumor. In this work, we propose to evaluate an automated HCC diagnosis method on a challenging dataset of patients with risk factors of HCC and small-size nodules. As a baseline approach, we train deep learning baseline models inspired by state-of-the-art results proposed in the literature [7, 13]. To improve those results on our datasets, we propose a two-step approach, directly inspired by the radiologists' reading grid [6]. This approach uses the LI-RADS radiological criteria for a preliminary learning task in order to improve the final prediction of HCC. To the best of our knowledge, no methods have been proposed in the literature to use the LI-RADS major features for histological HCC prediction on CT-scans. Our contributions are the followings:

- We propose to improve the baseline classification of HCC using weak radiological labels (LI-RADS major features), used for a preliminary learning task.
- We propose new manually handcrafted features well-designed for HCC classification, as well as LI-RADS based deep learning radiological features, which can be combined with handcrafted features to enhance the classification performance.
- We train and evaluate (cross-validation) our method on a challenging database of small and difficult HCC tumors [9, 14], and transfer it to one private test set.
- Eventually, we compare our method to deep learning baseline methods and to liver expert and non-expert radiologists' LI-RADS diagnosis.

2. METHOD

Deep learning features. The proposed approach builds on a baseline, *i.e.* a simple supervised classifier using as ground-truth the histological confirmation of HCC, by adding expert knowledge. We propose to follow the LI-RADS scoring system decision tree [9, 15, 16] as a preliminary learning task. More precisely, we propose to combine the prediction of the three LI-RADS major features with the baseline HCC prediction in order to guide the final HCC classification. To do so, we introduce three deep learning sub-models, based on the same backbone architecture as the baseline one. We train our deep models on the three LI-RADS criteria separately, which gives

us three probability outputs, one for each criterion. Furthermore, we add the size feature s , which can be automatically computed from the lesion segmentation.

Handcrafted features. To increase the predictive power of the networks outputs, we propose to combine them with handcrafted radiomics features that are directly inspired of the radiologists' reading grid. As represented in Figure 1, the LI-RADS binary criteria are



Fig. 1: Schema of the LI-RADS five main visual features. The threshold growth, which is obtained with longitudinal scans, is not available in the datasets used in this study.

visually assessed by observing the contrast differences of the lesions with respect to the liver parenchyma, or with the lesion borders. In this sense, it is intuitive to propose simple radiomics features computed on the input images. We propose three different formulas, based on the definition of each major feature [17]:

- **APHE:** the arterial phase hyper-enhancement is defined by a contrast difference between the lesion and the liver parenchyma. This contrast is observed on the arterial phase and can be seen on the portal venous phase, as shown in Figure 2, lines one and two.
- **EC:** the enhancing capsule is typically characterized by a thin enlightened contour with a darker inner lesion surface, as shown in Figure 2, at the second and third lines.
- **NPW:** the non-peripheral washout corresponds to a decrease in attenuation or intensity from earlier to later phase, resulting in hypoenhancement in the portal venous or delayed phase, *e.g* a darkening of the lesion surface. It is observed at every line of Figure 2.

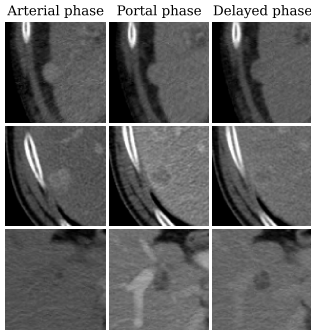


Fig. 2: Examples of center slices extracted from \mathcal{D}_1 .

Following the previous criteria description, we propose to characterize the three features by three simple formulas. The APHE feature f^{APHE} computes the difference of the median voxel values between the lesion and the parenchyma on the arterial phase. The same formulation is proposed for the NPW f^{NPW} , but on the portal venous phase. Finally, for the EC f^{EC} , we first propose to automatically compute the inner border of the lesion. Then we compute the difference of energy between the inner surface of the lesion and its border, which is a measure of the magnitude of voxel values in an image. Denoting L the lesion, P the liver parenchyma and B the lesion bor-

der, we obtain:

$$\begin{aligned} f^{APHE} &= \text{median}(L_{arterial}) - \text{median}(P_{arterial}) \\ f^{EC} &= \text{energy}(L_{arterial}) - \text{energy}(B_{arterial}) \\ f^{NPW} &= \text{median}(L_{portal}) - \text{median}(P_{portal}) \end{aligned}$$

Naming f the generic architecture used for the deep learning experiments, we denote by f_ω , f_θ , f_ν and f_ϕ the four distinct neural nets with parameters ω , θ , ν and ϕ and we propose to aggregate the outputs of all the models using a logistic regression giving us the final following model, with β being the 7-dimensional vectors of the regression coefficients:

$$\begin{aligned} \hat{y}_{HCC} = \beta \left[p_\omega^{HCC}, p_\theta^{APHE}, p_\nu^{EC}, p_\phi^{NPW}, f^{APHE}, f^{EC}, f^{NPW} \right]^T \\ + \beta_8 s + \alpha \end{aligned} \quad (1)$$

An overview of our method is presented in Figure 3.

Image processing. CT-scans processing. We worked with original images of full CT-scans of dimensions 512×512 with anisotropic z-spacings. Images were first selected according to the presence of the three last injection phases: arterial phase, portal venous phase and delayed phase. First, a simple registration was made between all phases on the portal venous phase by performing a linear vertical translation (z-axis) based on the liver segmentations. Registered images were then resampled to match their portal venous phase scan using a nearest neighbor interpolation. To ensure homogeneity in all our dataset, we then used the nn-UNet pre-processing which is well adapted for z-anisotropic 3D images [18], which allowed us to resample all our images to their median $0.76 \times 0.76 \times 2.00 \text{mm}^3$ voxel size.

Patch sampling. For lesion type classification, 3D patch inputs are commonly preferred [8, 13], notably due to the limited computational resources which do not allow using the full CT-scans as inputs with a large enough batch size. Lesion-centered patches were sampled on the pre-processed 3D images so that 95% of the segmented lesion fit in the boxes. Patches of size $96 \times 96 \times 24$ voxels were finally selected, corresponding to a $72.96 \times 72.96 \times 48.00 \text{mm}^3$ field of view. According to the predicted major feature, the final inputs of our networks have a size of $(N, C, 24, 96, 96)$ where N is the batch size and C the number of channels, with $C \in \{3, 4\}$ if we use arterial/venous phases + lesion segmentation (APHE) or arterial/venous/delayed phases + lesion segmentation (HCC, EC and NPW).

Datasets. Two datasets were leveraged in this study, coming from different centers of acquisition. One was used for training and validation, the two other datasets were only used for testing purpose.

\mathcal{D}_1 . Our first dataset contains a total number of 244 lesions, coming from 182 distinct patients. Among these lesions, 161 are histologically-proven hepatocellular carcinoma and 83 are other types of lesion. This dataset is the only one of our datasets that contains the three LI-RADS radiological criteria, for each lesion. Each criterion is provided by a binary number corresponding to their presence or absence, evaluated by an expert radiologist.

\mathcal{D}_2 . Our second dataset contains 1012 lesions corresponding to 543 patients and 602 no-HCC/410 HCC. This dataset presents cases of HCC that are mostly common [12, 19] as there exists a size gap between non-HCC and HCC lesions that we do not have in the first

dataset.

One major feature for HCC characterization remains the size of the lesion [2, 6]. A larger lesion has indeed more chance of being an HCC, hence the huge differences of lesion sizes between HCCs and non-HCCs amongst datasets used for deep learning methods in the literature [10, 12, 13]. In this study, we work on a challenging database of *small* hepatocellular carcinomas, which is a private set also used in [9]. Average lesion sizes for HCC and non-HCC tumors are reported in Table 1. While we can note a thin shift of ~ 4 millimeters between no HCCs and HCCs in \mathcal{D}_1 , the gap reaches 15 millimeters for \mathcal{D}_2 , which proves the difficulty of HCC characterization of our training dataset.

	No HCC	HCC
\mathcal{D}_1	15.4 ± 5.6	19.3 ± 5.6
\mathcal{D}_2	17.6 ± 12.7	32.6 ± 14.8

Table 1: Average lesion diameters for HCCs and no-HCCs in each dataset. Reported sizes are in *mm*.

Training and evaluation. In order to predict the three LI-RADS major features, we train our deep network on \mathcal{D}_1 in a stratified 5-fold cross-validation fashion. Each feature corresponds to an independent training. In parallel, we compute our handcrafted features and the lesion size, resulting in a 8-dimensional vector. We then optimize the logistic regression regularization coefficient on \mathcal{D}_1 , and for each deep architecture (row). Table 2 reports the cross-validation results on \mathcal{D}_1 .

We train our models using four different architectures. The first two backbones are inspired by [13] that provide state-of-the-art results of 92% AUC on their database. The first one, called “Tiny” in Table 2 and Table 3, is composed of five convolutional layers and two dense layers with 8 channels at the first layer and progressively doubled channels to reach a representation space of 128. Two dense layers are finally added, to reach 2 neurons, with a softmax function to predict probabilities (HCC vs no-HCC). This architecture contains 1.5M of parameters. The second “Small” architecture is the same architecture with doubled number of channels at each convolutional layer, resulting in a total number of 5.5M of parameters. This architecture is also presented in [20]. We also provide performances of a ResNet-18 [21], trained from scratch and initialized with weights that were pretrained on 23 different medical imaging datasets [22], as well as pretrained ResNet-50 with frozen first layers for fine-tuning purposes, reducing the number of trainable parameters to 24.9M for ResNet-18 and 29.2M for ResNet-50. We use the AdamW optimizer with a batch size of 32 and trained for 600 epochs, a learning rate of 10^{-5} and a weight decay of 10^{-3} , and for fine-tuning experiments we reduce the learning rate to 10^{-6} . For data augmentation, we use horizontal and vertical flips, rotations and affine transformations. Training experiments were realized on a Tesla V100 mono-GPU.

We evaluate our method on a private test set, \mathcal{D}_2 . To do so, we first infer the deep models of each criteria as well as for the HCC on the full set \mathcal{D}_2 to obtain predicted probabilities of HCC, APHE, EC and NPW. Then, we apply our handcrafted functions on the base to obtain features of each criteria, which allows us to obtain a feature vector of size 8 with the size feature, as for training. As we regularize with respect to the architecture in line, very small changes were observed in the HF column. For clarity, we report the average HF performance over the regularization used by line.

We report results of a cross-validated linear evaluation procedure,

i.e. we keep the same stratified cross-validation as for \mathcal{D}_1 , but we replace each validation with the full test set, hence the smaller standard deviations that are reported in Table 3. For each backbone, we report the results of choosing only the deep learning features + size (DLF), only the handcrafted features + size (HF), and the concatenation of DLF and HF + size (DLF+HF). Finally, we report the AUC from a non-specialist and a liver-pathology specialist on \mathcal{D}_1 . The AUC for radiologists is computed by normalizing the LI-RADS score to obtain HCC probability (LR-1,2,3,4,5 are divided by 5 to obtain values between 0 and 1).

3. RESULTS

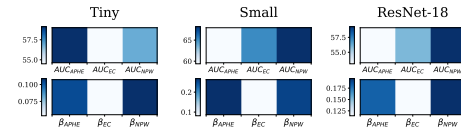


Fig. 4: First row: AUC metric for each deep learning model of LI-RADS major features. Second row: absolute values coefficients of the LI-RADS DLF in the logistic regression, for three backbones.

In Table 2, we can note that the proposed method outperforms deep learning baselines, with a boost of 6 to 18 points. For \mathcal{D}_1 , the best AUC is obtained combining our deep and handcrafted features, reaching an AUC of 75%, which represents a gain of 15 points with respect to a non specialist radiologist, and only 2.7 points away from a liver-pathology specialist. The performance of our method is robust to the backbone used. As a comparison with the state-of-the-art, the authors in [7] propose a 3D ResNet-18 computed on four-phases CT-scans 3D patches and obtain an AUC of 95.8%. With the exact same architecture, we obtain a baseline performance of 58.8% of AUC, which proves the inherent difficulty of our database and the small amount of data available. Pretrained ResNet-18 [22] helps however improving the performances, as shown in Table 2. In the same sense, the authors in [13] propose to use a deep learning architecture similar to our small encoder and obtain a groundbreaking AUC of 92%, while the same architecture obtains 64.4% on our dataset. It can be noticed that smaller encoders lead to better results in our case, given the small amount of data available.

In Table 3, we observe that for \mathcal{D}_2 the combination of our DLF and HF vectors outperforms both methods taken separately in most of the cases, as well as the deep learning baseline. One can note that performances on this dataset are much higher than on \mathcal{D}_1 , which reflects the lesion size gap evoked in Table 1. Our deep learning baseline transfers well enough on this base, proving the robustness of the training.

Figure 4 allows visualizing the importance of building an efficient model to predict the LI-RADS major features, before using them to predict HCC. We can indeed draw a parallel between the best individual performance of each predicted feature (by their AUC) and the absolute values of the deep learning major features in the logistic regression for HCC prediction. The logistic regression for HCC naturally selects the most accurately predicted major feature (NPW). However, as we can see with the small encoder performance, a smarter deep features combination could be beneficial for our model performance.

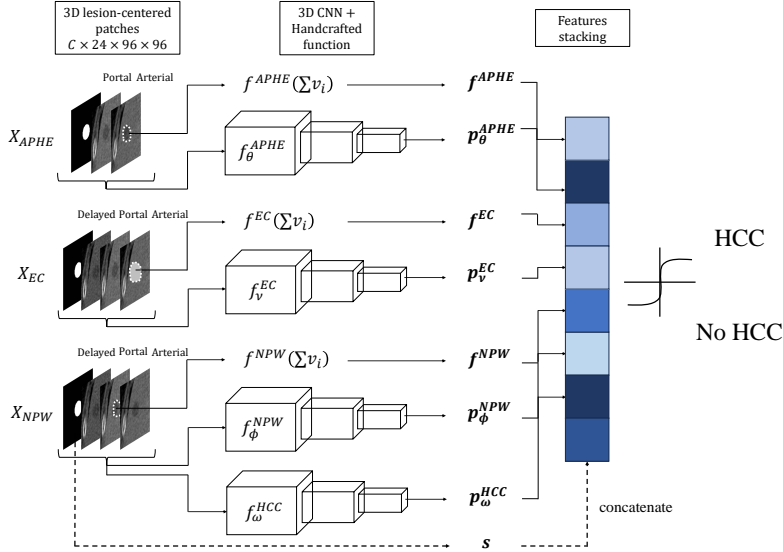


Fig. 3: Overview of the proposed method. s corresponds to the lesion diameter (in mm). The architecture leads to a one-dimensional vector of 8 components, being either a handcrafted feature or a predicted probability of presence of each criterion.

Train \rightarrow Test	Model	DL baseline	DLF	HF	DLF+HF	\uparrow w.r.t baseline
$\mathcal{D}_1 \rightarrow \mathcal{D}_1$	Radiologists	Non specialist: 60.4 / Specialist: 76.7 / AVG: 68.6				
	Tiny	68.3 ± 4.3	74.2 ± 4.7	74.4 ± 3.8	74.9 ± 3.8	+6.6
	Small [13]	64.4 ± 9.4	74.7 ± 3.9	74.4 ± 3.8	75.0 ± 3.8	+10.6
	ResNet-18 [7]	58.8 ± 4.5	70.6 ± 2.8	74.4 ± 3.8	70.5 ± 2.9	+15.6
	ResNet-18 (Med3D)	59.4 ± 3.7	73.7 ± 3.0	74.4 ± 3.8	74.2 ± 2.9	+15.0
	ResNet-50 (Med3D)	56.5 ± 3.5	71.0 ± 1.9	74.4 ± 3.8	71.2 ± 1.9	+17.9

Table 2: 5-fold cross-validation AUCs. Reported average AUCs and std are computed across the five folds. The small and ResNet-18 architectures obtain SOTA results on other datasets, as reported in [7, 13].

Train \rightarrow Test	Model	DL baseline	DLF	HF	DLF+HF	\uparrow w.r.t baseline
$\mathcal{D}_1 \rightarrow \mathcal{D}_2$	Tiny	77.8 ± 1.8	82.2 ± 0.1	83.0 ± 0.2	83.2 ± 0.3	+5.4
	Small [13]	72.7 ± 4.5	82.4 ± 0.2	83.0 ± 0.2	83.0 ± 0.3	+10.3
	ResNet-18 [7]	71.0 ± 2.4	81.6 ± 0.6	83.0 ± 0.2	82.2 ± 0.6	+12.0
	ResNet-18 (Med3D)	71.4 ± 2.6	82.3 ± 0.1	83.0 ± 0.2	83.2 ± 0.2	+11.8
	ResNet-50 (Med3D)	63.6 ± 4.3	81.7 ± 0.5	83.0 ± 0.2	82.4 ± 0.5	+19.4

Table 3: 5-fold cross-validation of a model trained on \mathcal{D}_1 and evaluated on \mathcal{D}_2 . At each iteration, we use one training fold from \mathcal{D}_1 and as validation fold we use the entire test set \mathcal{D}_2 . We show average and std AUC across the five folds. This explains the lower std compared to Table 2.

4. DISCUSSION AND CONCLUSION

In this paper, we proposed a novel approach for automatic hepatocellular classification by guiding our training using the radiologists' reading grid, the LI-RADS. We first predict LI-RADS major features using a deep learning approach in order to obtain probability of feature presence. In parallel, we propose a handcrafted approach to obtain relevant statistics related to each major feature. We then combine both approaches in a simple manner to predict HCC probability, and we achieve between 3 and 20 points improvements with respect to deep learning baselines, regardless of the backbone used, even with backbones that achieved state-of-the-art results for HCC classification in the literature.

For future work, an interesting perspective would be to improve the combination between DLF and HF, as proposed for instance in [23]

in which the authors minimize the mutual information between both vectors during pretraining in order to reduce redundancy between DL and HC features. Moreover, in this study we propose a two-step approach for HCC classification: first we extract features and then we train a logistic regression for HCC prediction. A potential idea would be to combine these blocks in an end-to-end approach, *i.e.* integrating the HCC classification as a dense layer after the major features classification layer. This method would involve using a binary cross-entropy computed on the major features (multilabel problem) and the HCC. Pretraining methods could also be explored, notably using the radiological features to guide the learning of a relevant latent space [20, 24, 25]. Eventually, there is a need for new datasets in HCC classification, notably public datasets (the only existing one exclusively contains HCC cases [26]), in order to test and compare the proposed method.

Compliance with ethical standards. This research study was conducted retrospectively using human data collected from various medical centers, whose Ethics Committees granted their approval. Data was de-identified and processed according to all applicable privacy laws and the Declaration of Helsinki.

Acknowledgments. This work was supported by Région Île-de-France (ChoTherIA project), ANRT (CIFRE #2021/1735) and was granted access to the HPC resources of IDRIS under the allocation 2023-A0150314655 and 2023-AD011014090 made by GENCI.

5. REFERENCES

- [1] Joo Hyun Oh and Dae Won Jun, “The latest global burden of liver cancer: A past and present threat,” *Clinical and molecular hepatology*, vol. 29,2, 09 2023.
- [2] McGlynn KA, Petrick JL, and El-Serag HB, “Epidemiology of hepatocellular carcinoma,” *Hepatology*, 2021.
- [3] Oestmann PM, Wang CJ, Savic LJ, et al., “Deep learning-assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (hcc) versus non-hcc on contrast-enhanced mri of the liver,” *European Radiology*, vol. 31, 07 2021.
- [4] Hanning Ying, Xiaoqing Liu, Min Zhang, et al., “A multicenter clinical ai system study for detection and diagnosis of focal liver lesions,” *Nature Communications*, vol. 15, 02 2024.
- [5] Wenqi Shi, Sichi Kuang, Sue Cao, et al., “Deep learning assisted differentiation of hepatocellular carcinoma from focal liver lesions: choice of four-phase and three-phase ct imaging protocol,” *Abdominal Radiology*, vol. 45, 09 2020.
- [6] Jorge A. Marrero, Laura M. Kulik, Claude B. Sirlin, et al., “Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the american association for the study of liver diseases,” *Hepatology*, vol. 68, no. 2, pp. 723–750, Aug. 2018.
- [7] Yating Ling, Shihong Ying, Lei Xu, et al., “Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase ct scans with minimum extra information,” *Frontiers in Oncology*, vol. 12, 2022.
- [8] Alexandre Bône and Marc-Michel Rohé, “Leveraging multiphase ct for the automatic diagnosis of hcc with deep learning: Impacts of lesion localization and phase inclusion,” in *IEEE ISBI*, 2023.
- [9] Sébastien Mulé, Maxime Ronot, Mario Ghosn, et al., “Automated ct li-rads v2018 scoring of liver observations using machine learning: A multivendor, multicentre retrospective study,” *JHEP Reports*, vol. 5, no. 10, pp. 100857, 2023.
- [10] Ke Wang, Yuehua Liu, Hongxin Chen, et al., “Fully automating li-rads on mri with deep learning-guided lesion segmentation, feature characterization, and score inference,” *Frontiers in Oncology*, vol. 13, 2023.
- [11] Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, et al., “Improving the automatic diagnosis of hepatocellular carcinoma with contrastive learning,” in *European Congress of Radiology*, 2024.
- [12] Hanning Ying, Xiaoqing Liu, Min Zhang, et al., “A multicenter clinical AI system study for detection and diagnosis of focal liver lesions,” *Nature Communications*, vol. 15, no. 1, pp. 1–16, December 2024.
- [13] Koichiro Yasaka, Hiroyuki Akai, Osamu Abe, and Shigeru Kiryu, “Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: A preliminary study,” *Radiology*, vol. 286, pp. 170706, 10 2017.
- [14] Christophe Aub, Frederic Oberti, Julie Lonjon, et al., “Easl and aasld recommendations for the diagnosis of hcc to the test of daily practice,” *Liver international : official journal of the International Association for the Study of the Liver*, vol. 37, 03 2017.
- [15] Yunan Wu, Gregory M. White, Tyler Cornelius, et al., “Deep learning li-rads grading system based on contrast enhanced multiphase mri for differentiation between lr-3 and lr-4/lr-5 liver tumors,” *Annals of Translational Medicine*, vol. 8, 2020.
- [16] Junghoan Park, Jae Seok Bae, Jong Min Kim, et al., “Development of a deep-learning model for classification of li-rads major features by using subtraction images of mri: a preliminary study,” *Abdominal Radiology*, vol. 48, no. 8, pp. 2547–2556, Aug. 2023.
- [17] Cynthia Santillan, Kathryn Fowler, Yuko Kono, et al., “Li-rads major features: Ct, mri with extracellular agents, and mri with hepatobiliary agents,” *Abdominal Radiology*, vol. 43, 01 2018.
- [18] Fabian Isensee, Jens Petersen, Andre Klein, et al., “nnu-net: Self-adapting framework for u-net-based medical image segmentation,” 2018.
- [19] Jin-Young Choi, Jeong-Min Lee, and Claude Sirlin, “Ct and mr imaging diagnosis and staging of hepatocellular carcinoma: Part i. development, growth, and spread: Key pathologic and imaging aspects,” *Radiology*, vol. 272, pp. 635–654, 09 2014.
- [20] Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, et al., “Learning to diagnose cirrhosis from radiological and histological labels with joint self and weakly-supervised pretraining strategies,” in *IEEE ISBI 2023*.
- [21] Kaiming He, X. Zhang, Shaoqing Ren, et al., “Deep residual learning for image recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [22] Sihong Chen, Kai Ma, and Yefeng Zheng, “Med3d: Transfer learning for 3d medical image analysis,” 2019.
- [23] Rebeca Vétül, Clément Abi-Nader, Alexandre Bône, et al., “Non-redundant combination of hand-crafted and deep learning radiomics: Application to the early detection of pancreatic cancer,” in *Cancer Prevention Through Early Detection*, Cham, 2023, pp. 68–82, Springer Nature Switzerland.
- [24] Emma Sarfati, Alexandre Bône, Marc-Michel Rohé, et al., “Weakly-supervised positional contrastive learning: Application to cirrhosis classification,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*.
- [25] Camille Ruppli, Pietro Gori, Roberto Ardon, et al., “Decoupled conditional contrastive learning with variable metadata for prostate lesion detection,” in *Medical Image Learning with Limited and Noisy Data*, Cham, 2023, pp. 95–105, Springer Nature Switzerland.
- [26] “The cancer imaging archive (tcia) dataset,” Accessed: YYYY-MM-DD.