DOI:10.14753/SE.2025.3142

SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3142.

STOLLMAYER RÓBERT

Gasztroenterológia

című program

Programvezető: Dr. Molnár Béla, kutatóprofesszor, az MTA doktora

Témavezető: Dr. Kaposi Novák Pál, egyetemi docens

DOI:10.14753/SE.2025.3142

Deep learning in the magnetic resonance imaging-based diagnostics of focal liver lesions using hepatocyte-specific contrast agents

PhD thesis

Róbert Stollmayer

Károly Rácz Conservative Medicine Division

Semmelweis University

Supervisor: Pál Novák Kaposi, MD, Ph.D.

Official reviewers: Zsuzsanna Földes-Lénárd, MD, Ph.D.

András Horváth, Ph.D.

Head of the Complex Examination Committee: Béla Molnár, MD, D.Sc.

Members of the Complex Examination Committee: Andrea Ferencz, MD, Ph.D.

Werling Klára, MD, Ph.D.

Kristóf Dede, MD, Ph.D.

Budapest

2024

DOI:10.14753/SE.2025.3142

Table of contents

DOI:10.14753/SE.2025.3142

List of abbreviations

2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
AI	artificial intelligence
ANN	artificial neural network
AUC	area under the receiver operating characteristic curve
BCE	binary cross-entropy
cc.	carcinoma
CLD	chronic liver disease
CNN	convolutional neural network
CRC	colorectal cancer
CT	computer tomography
DLM	deep learning model
ECA	extracellular contrast agent
FLL	focal liver lesion
FNH	focal nodular hyperplasia
GI	gastrointestinal
GT	ground truth
HAP	hepatic arterial phase
HBP	hepatobiliary phase
HCC	hepatocellular carcinoma
HSC	hepatocyte-specific contrast agent
LI-RADS	Liver Imaging Reporting and Data System
LLM	large language model
MET	metastasis
MONAI	Medical Open Network for Artificial Intelligence
MRI	magnetic resonance imaging
ms	milliseconds
NPV	negative predictive value
PPV	positive predictive value
pred	prediction

DOI:10.14753/SE.2025.3142

prob	probability
PVP	portal venous phase
ROC	receiver operating characteristic curve
ROI	region of interest
SD	standard deviation
SPAIR	spectral-attenuated inversion recovery
T	Tesla
T1w	T1-weighted
T2W, T2w	T2-weighted
TE	echo time
TR	repetition time
VEN	transitional phase
VLM	vision language model

1. Introduction (with the scientific background and the relevant literature)

1.1. Artificial intelligence and deep learning

Artificial intelligence (AI) is the field of study of algorithms mimicking human intelligence, which encompasses a wide range of sub-fields, including machine learning (ML), which deals with the development of learning algorithms, and deep learning (DL), a sub-section of machine learning algorithms which are developed by mathematical modeling of the human nervous system. DL uses artificial neural networks (ANNs) with multiple layers to extract and compress information from their input data (1). While simple ML algorithms are widely used in medical research among them in radiomics, their applicability and accuracy may be limited in comparison to DL algorithms. (1).

Convolutional neural networks (CNNs) are a specialized type of ANNs developed for image analysis, that can extract and compress features from radiological imaging data with configurable convolutional layers. Although CNNs are inherently more complex than traditional ML methods, their use does not require additional steps for feature extraction (1–3).

CNNs most commonly fall into the category of supervised learning algorithms, meaning that to be able to change the parameters of the network, an assigned ground truth (GT) annotation is required (3). CNNs in radiology may be used for image segmentation, classification, object detection, or image generation, among other various tasks in medical imaging (2). Depending on the required network output, the GT annotations should also match this output. Self-supervised and unsupervised approaches have also been explored for radiological image analysis but have not yet reached the level of accuracy of supervised learning methods (1,2).

Transformers are newer neural network architectures which are designed primarily for natural language processing, due to their ability to capture global context and long-range dependencies within data. In many cases, these have been shown to have superior performance to CNNs in computer vision tasks (1). Large language models (LLMs) are a recent development built on the transformer DL architecture enabling the large-scale automatic processing of digital text. Although LLMs are still young, so-called vision language models (VLM) have already been built on them enabling the joint processing of medical imaging and text information (1,4). In the long run, multi-modal large models, such as VLMs, are a goal of DL research, as well as the development of

generally adaptable, so-called foundation models which can either be used for various tasks without modification or require minimal computation for adjustment (fine tuning) for a specific task (1).

DL studies establishing new models usually start with data collection and annotation steps, which is currently also the most time-consuming and difficult part of these studies due to the low availability of high-quality clinical data (even in highly curated datasets and repositories). Collected data is organized into a data structure that is then used for DL model training and evaluation. This is commonly done by splitting the data into training, validation ("tuning"), and testing partitions or by k-fold cross-validation (usually with five folds). The data compartment used for model training can be "enriched" via data augmentation methods, and generalization can be improved through various regularization methods (2).

1.2. Analysis of focal liver lesions using deep learning

Radiological imaging as a tool for the detection and diagnosis of focal liver lesions (FLLs) has steadily grown worldwide in the recent years, while delays in treatment and unnecessary procedures and costs can only be lowered by apt and accurate diagnoses of these pathologies. In comparison to computer tomography (CT), magnetic resonance imaging (MRI) provides the most detailed non-invasive method for the characterization of FLLs, with specificity above 82% and sensitivity above 94%, thanks to its excellent soft tissue contrast, without exposure to ionizing radiation, especially when performed with intravenous contrast agents (5).

Contrast agents, that are specifically designed for the imaging of physiologically functioning hepatocytes, such as gadoxetate disodium or gadobenate dimeglumine, have proven to be highly useful in the detection and differentiation of benign and malignant foci of the liver. These contrast agents, also called hepatocyte-specific contrast agents (HSCs) are taken up by hepatocytes and secreted into the bile (6). HSCs therefore can distinguish between liver lesions composed of normal hepatocytes and those made up of poorly differentiated hepatocytes or non-hepatocytic cells. Additionally, these agents facilitate the detection of small lesions by creating a stark contrast between the enhancing parenchyma and areas lacking normal hepatocytes during the hepatobiliary phase (HBP) (6).

Common focal liver lesions (FLLs) like hepatocellular carcinoma (HCC) and liver metastasis (MET) pose a considerable diagnostic challenge. Patients with these conditions often only have curative options through surgical resection or image-guided ablation, making early detection crucial. METs are the most frequently occurring malignancies in the liver, and it is reported that up to 70% of all individuals with colorectal cancer will develop METs at some stage during their lifetime (7).

HSC-enhanced MRI has been shown to have the highest sensitivity (73.3%) for METs smaller than 10 mm in diameter among all imaging modalities, which also translates to a significantly better survival rate of patients imaged with MRI (70.8%) compared to those imaged with CT (48.1%) (8). HCC is the fifth most common solid malignancy worldwide, and the mortality rate from HCC is predicted to rise in the coming decades (9). HCC typically develops in the background of decades of chronic liver disease (CLD). MRI findings of an arterial-enhancing mass with subsequent washout and enhancing capsule on delayed interstitial phase images are diagnostic for HCC (10). Focal nodular hyperplasia (FNH) is the second most common benign solid FLL after hemangioma. FNH is a common incidental finding in imaging studies, and it is a frequent source of differential diagnostic dilemmas of malignant lesions. A definitive diagnosis of FNH can be established in patients who do not have CLD when typical features such as arterial phase and HBP hyperenhancement and a central scar are detected with HSC MRI (11). Standardized data collection and reporting systems have also been developed, such as the Liver Imaging Reporting and Data System (LI-RADS), to improve CT and MRI diagnosis by reducing variability in the interpretation of imaging studies (12). However, due to the complex nature of these systems, their integration into the clinical workflow can be cumbersome.

AI techniques have been introduced in growing numbers to facilitate lesion detection and classification, assess the patients' prognosis, or identify risk factors of FLLs based on imaging studies (13). Some of these studies extracted large numbers of image features from dynamic contrast-enhanced MRI to build mathematical models for the automatic classification of FLLs (14). Deep learning models (DLMs) are state-of-the-art image processing algorithms predominantly based on CNNs. DLMs have been tested for analysis of all known imaging modalities and achieved excellent results in image-based detection and classification of various diseases (15). A handful of studies have also

applied DLMs to classify FLLs in MRI images and demonstrated that the performance of the DLMs is excellent and comparable to the human observers' diagnostic rate (16–19). Among different DLM architectures, models using 3D convolutions could be efficiently trained on a relatively small number of cases for differentiating between the most common types of FLLs (16,17).

The processed data required for DLMs often include two-dimensional (2D) slices or three-dimensional (3D) image volumes; moreover, in the case of magnetic resonance imaging (MRI) studies, the different sequences can be condensed into a multi-channel input. Due to their efficiency, convolutional neural networks (CNNs) have replaced other ML approaches in most image classification and segmentation tasks (20,21).

Meanwhile, current CNN classification models have limited value in clinical practice as these have been trained to diagnose only a handful of liver pathologies based only on a small set of MRI images. Current DLMs cannot recognize many FLLs belonging to less common diagnoses and lesions with atypical image features or with post-treatment changes (22). A clinically useful DLM must be able to analyze multiple image sequences to identify a comprehensive set of image features that can be used for the characterization of FLLs and the generation of a differential diagnosis (23). For the transparency of the AI-driven classification process, it is essential to know how many of the detected image features support the diagnosis and to validate the localization of these features via network visualization techniques, such as activation and occlusion sensitivity maps. Such sophisticated models are better suited for the systematic evaluation of FLLs and can increase the efficiency and the reproducibility of the imaging diagnosis as well as facilitate and standardize the development of novel imaging methods, for example new MRI reconstruction methods.

DOI:10.14753/SE.2025.3142

2. Objectives

Considering recent advancements in abdominal radiology and radiomics the aim of the current thesis is to demonstrate the feasibility of DL-based focal liver lesion analysis using freely available open-source software.

The objective of the first study presented in the current study was to show that focal liver lesions frequently imaged using HSC MRI in the Hungarian population can automatically be categorized into classes approximating the opinion of a board-certified radiologist. Furthermore, to evaluate whether a single axial slice of an FLL contains enough information for this categorization in comparison to using three-dimensional images.

In our second study aimed to quantify the accuracy with which radiological features of FLLs can automatically be identified based on HSC MRI and to evaluate the agreement between the created model, a radiologist trainee, and an experienced abdominal radiologist.

DOI:10.14753/SE.2025.3142

3. Methods

3.1. Differentiation of tumor types

3.1.1. Patient and MRI study selection

In the first retrospective single-center study multi-phasic MRI studies of patients with FNHs, HCCs or METs were collected. These studies were conducted using gadoxetate disodium, an HSC and they were collected from the picture archiving and communication system of our institute. The need for written patient consent was waived by the Institutional Research Ethics Committee (SE-RKEB 136/2019.) due to the retrospective nature of the study. For image acquisition between November 2017 and October 2020 a Philips Ingenia 1.5 T scanner (Philips Medical Systems, Eindhoven, The Netherlands) was used. T2-weighted (T2w) spectral-attenuated inversion recovery (also known as SPAIR), arterial phase (HAP), portal venous phase (PVP), and HBP scans of each eligible patient were collected for further analysis. Lesions that were included were either histologically confirmed or exhibited typical radiological characteristics of the given lesion type with this type of MRI examination based on the opinion of an abdominal radiologist. Patients younger than 18 years of age at the time of imaging were excluded from the study. Patient demographics, collected and analyzed lesion properties, and in the case of metastatic lesions the histological types of the original tumor are shown in Table 1.

Table 1. *Patient demographics, imaging properties per lesion class, and details of metastatic lesion origin. (17)*

	FNH	HCC	MET	Total
Number of patients	42	13	14	69
Age in years at imaging, mean ± SD	45 ± 12	66 ± 5	57 ± 10	54 ± 14
Sex				
Male	11	8	8	27
Female	31	5	6	42
Lesion properties				
Number	71	69	76	216
Primary type				
CRC			21	
Leiomyosarcoma			18	
GI adenocc. or cholangiocc.			15	
Breast cc.			11	
Pancreas cc.			7	
Neuroendocrine ileum cc.			3	
Papillary thyroid cc.			1	

cc.: Carcinoma; CRC: Colorectal cancer; FNH: Focal nodular hyperplasia; GI: Gastrointestinal; HCC: Hepatocellular carcinoma; MET: Metastasis; SD: Standard deviation; T: Tesla.

3.1.2. Data preparation and dataset creation

MRI scans were exported as DICOM files, that were anonymized to remove the patients' social security numbers, birth date, sex, age, body weight, and date of the imaging study. Anonymized PVP and HBP files were resampled and spatially aligned to the corresponding T2w volume using BSpline as a non-rigid registration method via 3D

Slicer, an open-source visualization and medical image computing software, which was also used for annotation, cropping using the area inside the annotation and file conversion (24,25). Lesions were annotated by cubic regions of interest (ROIs). The lesions were then cropped from the aligned HAP, PVP, HBP, and T2w volumes using the same ROI. The cropped volumes were converted to NIfTI (26) file format. The saved NIfTI files were combined into one four-dimensional (4D) input data for each lesion (Figure 1). Cropped lesions were randomly sorted into datasets. 10-10 lesions were added to the test and validation dataset from each class, and the remaining tumors were added to the training dataset. For the analysis of axial slices NIfTI files were sliced up into axial PNG images. The resulting T2w, HAP, PVP, and HBP PNG files were concatenated (Figure 2) using a custom-written computer program in Python. The training and validation datasets contained three axial slices of each lesion (i.e. three most representative axial slices of the lesion within the NIfTI files). The test dataset contained one slice from each lesion.

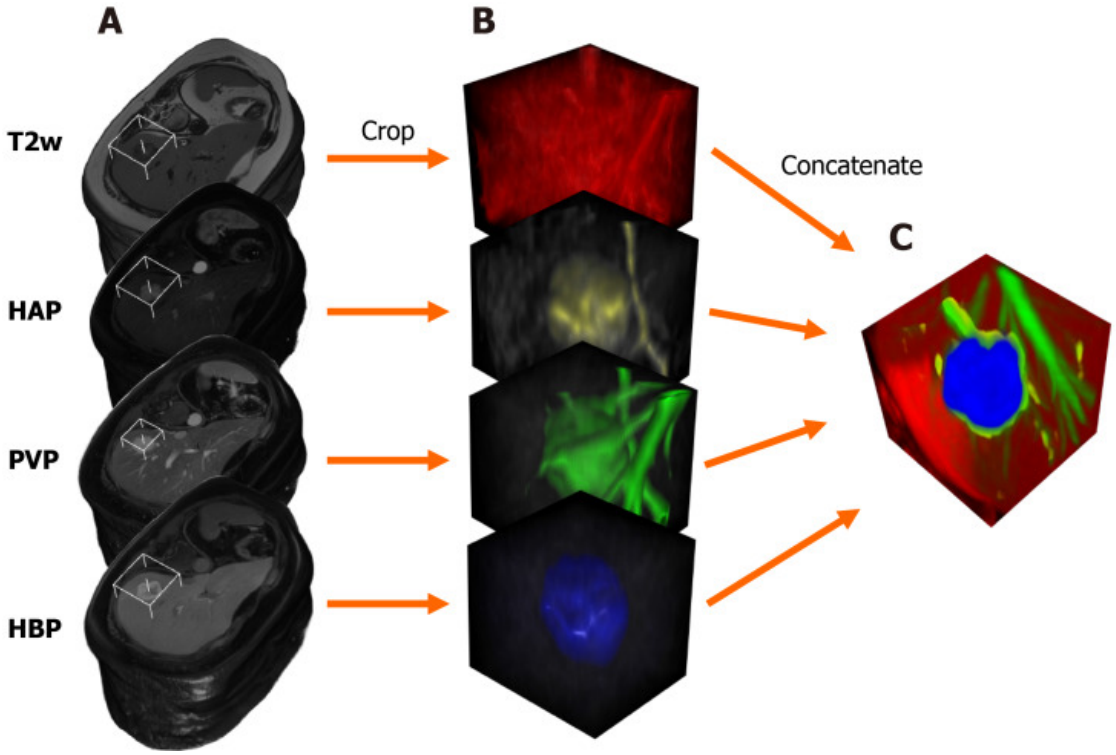

Figure 1. *Steps of input data preparation for the three-dimensional densely connected convolutional neural network. A: Three-dimensionally rendered whole volumes at the level of the lesion (indicated by the white frame); B: Cropped cubic volumes containing the lesion; C: The four cubic volumes are concatenated into one four-dimensional file;*

each volume is represented by a different color. T2w: T2-weighted; HAP: Hepatic arterial phase; PVP: Portal venous phase; HBP: Hepatobiliary phase. (17)

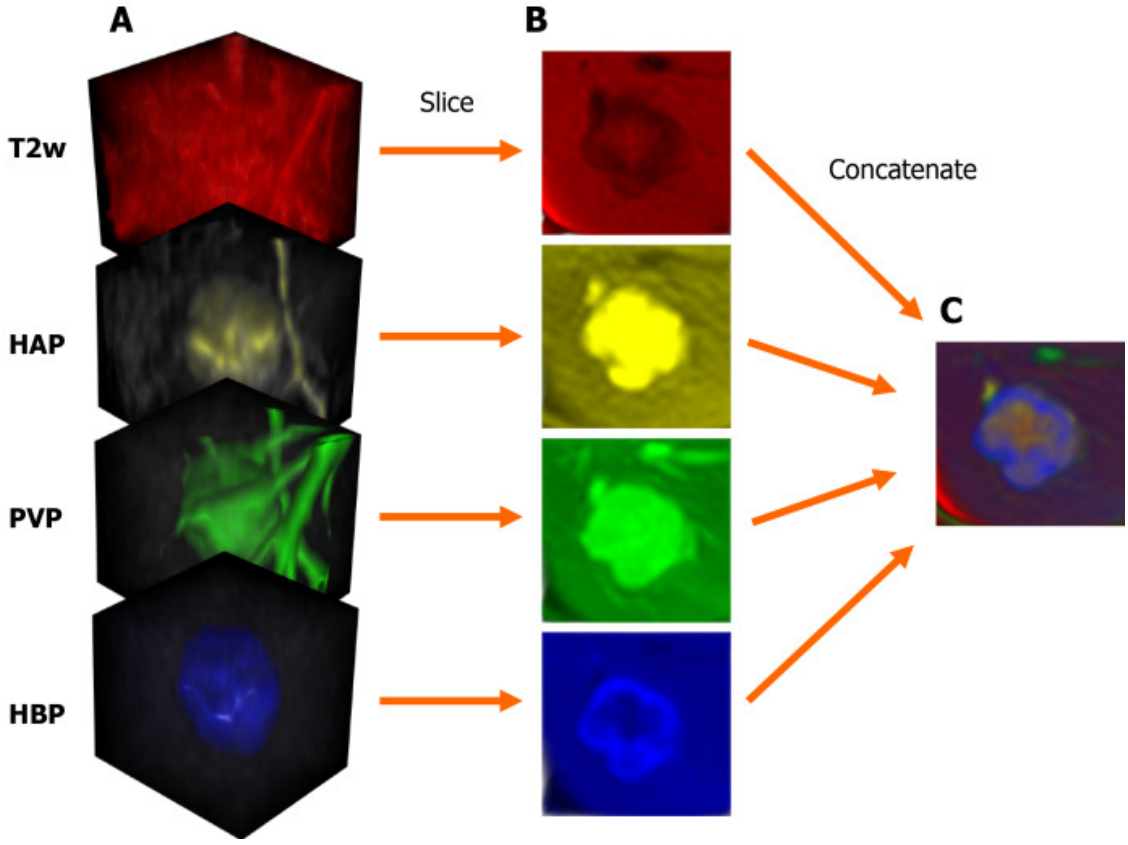

Figure 2. *Illustration of multi-channel input data for the 2D-DenseNet. Colors are changed for the purpose of illustration. A: Cubic magnetic resonance image volumes containing the lesion; B: Axial slices acquired from the cropped volumes; C: The four axial slices are concatenated into a single image; each slice corresponds to a respective image. T2w: T2-weighted; HAP: Hepatic arterial phase; PVP: Portal venous phase; HBP: Hepatobiliary phase. (17)*

3.1.3. Data processing, training, and testing

Concatenated files were modified via transform functions during the training process. Image pixel intensity was scaled between -1.0 minimum and 1.0 maximum values. Data augmentation transforms were applied to the training samples, including random rotation (70° range along two axes) and zoom (0.7–1.4 scaling) to enrich training data. Furthermore, the axial slices (2D images) were resized to 64 × 64 resolution.

Transformed images were converted to tensors, which were then fed to a DenseNet264 that used 2D convolutional layers (27).

In the case of the 3D-DenseNet264 network, NIfTI voxels were resampled to an isovolumetric shape, voxel intensities were rescaled between -1.0 minimum and 1.0 maximum value and NIfTI files were resized to 64 × 64 × 64 spatial resolution. Input resolution for both 2D and 3D networks was determined through manual hyperparameter tuning with the intention to find the lowest image resolution that does not decrease model performance metrics. The four NIfTI files were concatenated (T2w, HAP, PVP, HBP) to be used as multi-channel input for the 3D CNN. 90° rotation along two spatial axes, 60° rotation (x, y axes), zoom (0.8 - 1.35), and flipping were randomly applied to the training samples. MR volumes were converted to 4D tensors (number of channels, x-, y- and z-dimensions) that were used as network input. We used DenseNet264 models through the Medical Open Network For Artificial Intelligence (MONAI) framework (28). During the training of the networks categorical cross-entropy loss was used as a loss function and the Adam optimizer (29). Network weights were randomly initialized. Networks were trained for 70 epochs. The area under the receiver operating characteristic curve (AUC) was used for the selection of the highest performing model weights. AUC was calculated after each epoch, and the model with the highest average AUC value was saved as the final model. Statistical evaluation is done in detail on the hold-out test dataset containing 10 lesions from each class. The tumor type with the highest probability, according to the last softmax layer of the CNNs, was selected as the predicted lesion type, encoding the predicted diagnosis as 1, while the predicted incorrect classes as 0. The calculation of metrics, such as specificity, sensitivity, f1 score, positive predictive value (PPV) and negative predictive value (NPV) for each class was based on these predictions.

Classification performance is also measured using AUC values of each class, calculated from the softmax layer probability outputs. 95% confidence intervals are reported after each AUC value. To evaluate statistical significance between the test performance of the 2D and 3D classifiers DeLong's test was used [14].

3.2. Identification of radiological features

3.2.1. Clinical Dataset

For our retrospective study, 99 patients were included (Table 2) who underwent abdominal MRI with gadoxetate disodium, an HSC, between 29 September 2017 and 11

August 2021, at our institution. As this is a retrospective study, the need for written patient consent for this retrospective analysis itself was waived by the Institutional Research Ethics Committee (SE-RKEB 136/2019.). However, all patients gave written informed consent for the MRI examination. The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional review board of our university. Inclusion criteria of the study covered patients who were examined for FLLs with HSC-enhanced MRI in our institution (134 patients, 175 examinations) using the same 1.5 T MRI machine, a Philips Ingenia 1.5 T scanner (Philips Medical Systems, Eindhoven, The Netherlands), and whose liver lesions could be unequivocally diagnosed based on histology sampling or typical imaging findings as it has been recommended by international guidelines. Exclusion criteria included age under 18 years at the time of the imaging, pregnancy, incomplete or inadequate quality scans, data collection errors, examination performed on a different MRI machine, and lesions with an equivocal diagnosis. Fifteen examinations were excluded as they were performed on a different scanner, 21 studies were excluded due to incomplete or inadequate scan quality and data collection errors, while 4 studies were performed on underage patients, and 4 studies did not contain or only contained lesions with an equivocal diagnosis.

Table 2. *Patient demographics and types of lesions analyzed in the study. Some patients were diagnosed with multiple lesion types; therefore, the number of included patients is not equal to the sum of the number of patients diagnosed with different lesion types. (30)*

	FNH	HCC	MET	Other	All Patients
Number of patients	52	23	17	16	99
Male	15	16	9	6	42
Female	37	7	8	10	57
Average age at the time of imaging	44	64	57	53	54

FNH: focal nodular hyperplasia, HCC: hepatocellular carcinoma, MET: liver metastasis.

The final study cohort included 131 scans of 99 subjects diagnosed with 105 FNHs, 121 HCC, 121 METs, and 32 other lesions belonging to various groups (such as hemangiomas and adenomas).

3.2.2. Image Acquisition and Processing

All MRI scans were acquired using a Philips Ingenia 1.5 T scanner (Philips Medical Systems, Eindhoven, The Netherlands) and 5–20 mL intravenous gadoxetate disodium contrast with a dosage of 0.025 mmol/kg. The scans were performed according to our institutional guidelines. For the current study the T2w SPAIR, native T1-weighted 3D mDIXON (NAT), HAP, PVP, equilibrium phase (VEN) T1-weighted 3D mDIXON, as well as HBP, standardly acquired at 15–20 s (HAP), 70–80 s (PVP), 2–3 min (VEN) and 20 min (HBP) after contrast administration, images of each lesion were collected from the institutional picture archiving and communications system. Both T2w and T1w scans were acquired in breath-hold. Standard 3D mDIXON and T2 SPAIR sequences were used. 3D mDIXON: 390 × 390 × 106 average image resolution, 0.942 mm × 0.942 mm average pixel spacing, 2.5 mm average spacing between slices, 4–6 mm slice thickness, 5.8 ms repetition time (TR), 1.8/4.0 ms echo time (TE), 15° flip angle and 552–616 Hz/pixel receiver bandwidth. T2 SPAIR: 398 × 398 × 100 average image resolution, 0.935 mm × 0.935 mm average pixel spacing, 2.8 mm average spacing between slices, 3–6 mm slice thickness, 1000–6742 ms TR, 100 ms TE, 90° flip angle, and 325–666 Hz/pixel receiver bandwidth.

Each scan was anonymized, and personal identifiers, such as patient name, birth date, social security number, and date of imaging were removed. All scans were converted to 3D NIfTI (26) image format, HBP scans were resampled with linear interpolation to isotropic, 1 × 1 × 1 mm voxel spacing, and all other scans were coregistered to the corresponding HBP scan. For image registration, the ITKElastix toolbox (31) was used with the rigid default parameter map. Misalignments resulting from image registration were manually corrected when necessary. Each FLL reported was marked with a single point marker placed in the HBP or HAP scan, and a radiologist with 13 years of experience in abdominal imaging marked the diameters of the lesions. Lesion smaller than 5 mm (largest axial diameter) were excluded from the study). Lesions were cropped from each scan based on their largest diameter, to which a 2-mm (2 voxels)-wide zone was added in each direction to account for misalignments between the 6 scans. For lesion marking and manual correction of misalignment 3D Slicer (24) was used. The steps of the analysis are shown in Figure 3.

DOI:10.14753/SE.2025.3142

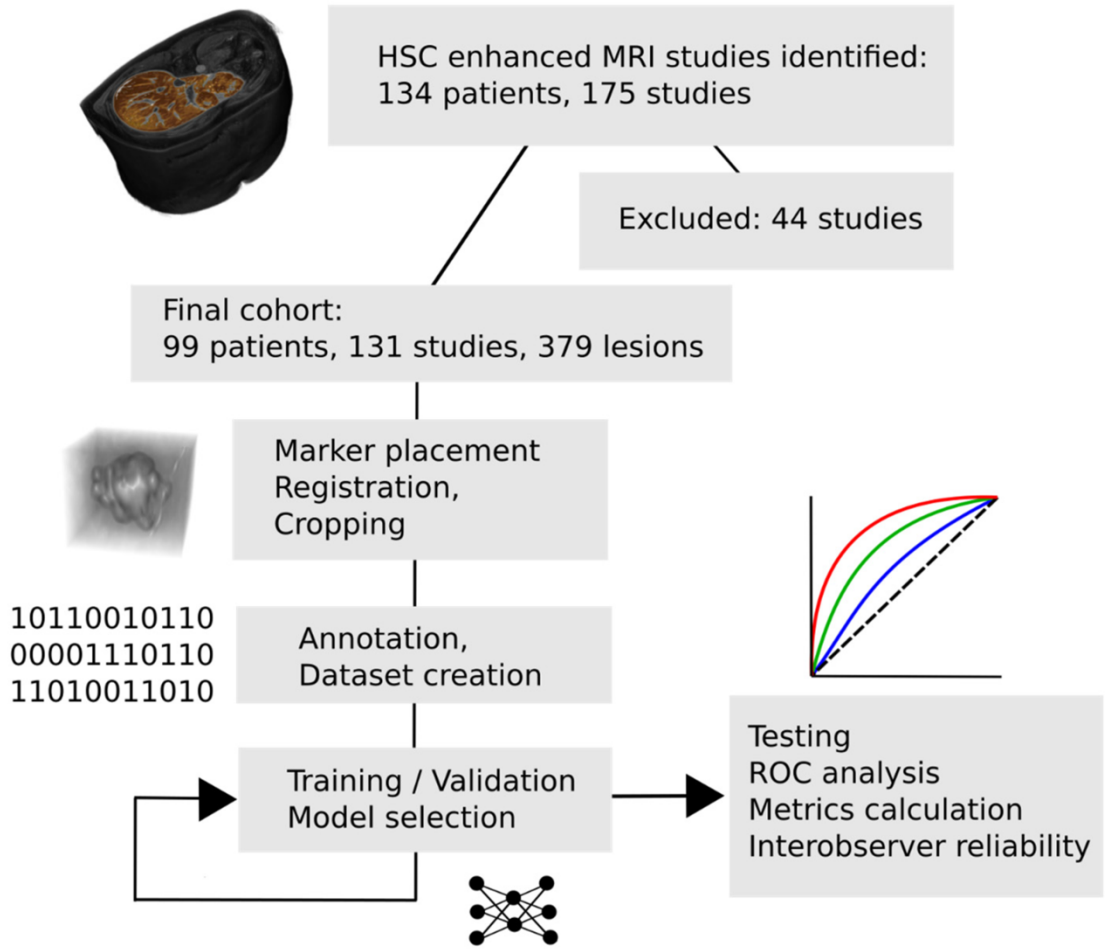

Figure 3. *Steps of the analysis. HSC: hepatocyte-specific contrast-enhanced, MRI: magnetic resonance imaging. (30)*

Cropped lesions were randomly sorted into training, validation, and test datasets for DLM training and testing (Table 3) in a ratio of 53:17:30. In the MRI scans of 99 patients, the total number of focal liver lesions was 379, 202 lesions were used for training, 65 for validation, and 112 for testing. All scans and lesions belonging to the same patient were assigned to the same dataset. Each lesion was evaluated by an expert radiologist with 13 years of experience in abdominal imaging, as well as a radiology resident with 4 years of experience. Annotators had to decide whether the tumor belonged to the four provided tumor types: FNH, HCC, MET, or other; and whether or not the lesion contained the following radiological features: early (arterial phase) contrast enhancement, washout, delayed phase enhancement, peripheral enhancement, central scar, capsule, T2 hyperintensity compared to the surrounding liver tissue, iso- or hyperintensity compared to the surrounding liver tissue on venous phase, hypoenhancing

core, hemorrhage/siderosis. For AI training, the expert-reported labels were used as GT, as well as for results calculation. Table 3 contains the detailed distribution of expert annotations among datasets. The human observers had no access to the final diagnosis or the opinion of the other reader. The detailed distribution of the final annotations among datasets is shown in Table 3.

Table 3. *Distribution of lesions and characteristics among datasets. (30)*

Tumor Type	Train	Validation	Test	Total
FNH	53	16	36	105
HCC	62	22	37	121
MET	72	19	30	121
Other	15	8	9	32
Radiological features				
Early enhancement	99	36	65	200
Washout	41	8	32	81
Delayed phase enhancement	65	28	36	129
Peripheral enhancement	53	21	31	105
Central scar	37	11	19	67
Capsule	27	6	18	51
T2 hyperintensity	88	39	42	169
Iso- or hyperintensity on venous phase	64	28	37	129
Hypoenhancing core	110	28	50	188
Hemorrhage/Siderosis	36	17	17	70

FNH: focal nodular hyperplasia, HCC: hepatocellular carcinoma, MET: liver metastasis.

3.2.3. Deep Learning Methods

To be able to automate radiological feature generation, multiple deep learning algorithms were trained with different hyperparameter setups. Models implemented in MONAI (28), an open-source framework for deep learning in healthcare imaging were used. MONAI provides multiple 3D neural network implementations that can perform classification tasks on medical images. Each trained convolutional neural network had 6

input channels, one for each 32 × 32 × 32 resolution cubic input MRI scan showing the liver lesion. Input resolution was determined through manual hyperparameter tuning with the intention to find the lowest image resolution that does not decrease model accuracy. We trained DenseNet121, DenseNet169, DenseNet201, DenseNet264 (32), EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, EfficientNetB7 (33) models with various batch sizes, dropout rates and weight decay, among other hyperparameters. All models, including EfficientNets, were trained from scratch using random initialization. All models were trained for at least 300 epochs. Each model was modified so that its last classifier layer would be a sigmoid layer to be able to perform multi-label classification. Each model has 10 probability outputs (numeric values ranging from 0 to 1), one for each radiological feature that it is trained to predict. By replacing (occluding) a part of the input volume with the mean voxel intensity of the image, the model prediction for each radiological feature changes. If important parts of the image are occluded, the prediction probability decreases, which can be visualized for each input channel and output feature. More negative values indicate higher importance in the decision process. This visualization technique will be referred to as ‘occlusion sensitivity map’. These maps can be useful for the interpretation of network predictions, highlighting the areas that played a more important role in the prediction of a feature (34). Accordingly, binary cross-entropy (BCE) loss was calculated and used for model weight adjustment during training.

DenseNets were trained with dropout probabilities of 0, 0.25, 0.5, and 0.75. Each model was trained using an Adam optimizer (29) and a learning rate of 0.0001. All models were trained from scratch, without pretraining. To improve the generalizability of the model, different data augmentation methods (such as rotation of the images) were applied during training. Images were then resized to a 32 × 32 × 32 input shape. Image intensities were normalized and scaled between −1 and 1. The best-performing model was defined as the one achieving the highest mean AUC of the 10 predicted features on the validation dataset. During training the performance on the validation dataset was evaluated after every 20 epochs. Detailed evaluation of the final selected model was done on a hold-out test dataset (Table 3).

3.2.4. Statistical Evaluation

ROC analysis was performed on the test dataset, via which AUC values were calculated, and cut-off values were set for each feature separately based on Youden's index. At the given thresholds, sensitivity, specificity, PPV, NPV, and f1 score were calculated for each feature. Reported measures are calculated in comparison to the expert radiologist's opinion (GT). Statistical power was calculated according to Obuchowski's method using the 'pROC' (35,36) R package. Inter-rater reliability was calculated via Cohen's Kappa between the expert opinion, annotations of a radiology resident, and the final selected machine learning model.

DOI:10.14753/SE.2025.3142

4. Results

4.1. Differentiation of tumor types

The 2D model achieved the highest average validation set AUC after 46 epochs, while the best average AUC value of the 3D network was reached after 62 epochs. Learning curves and AUC metrics for model selection are shown in Figure 4.

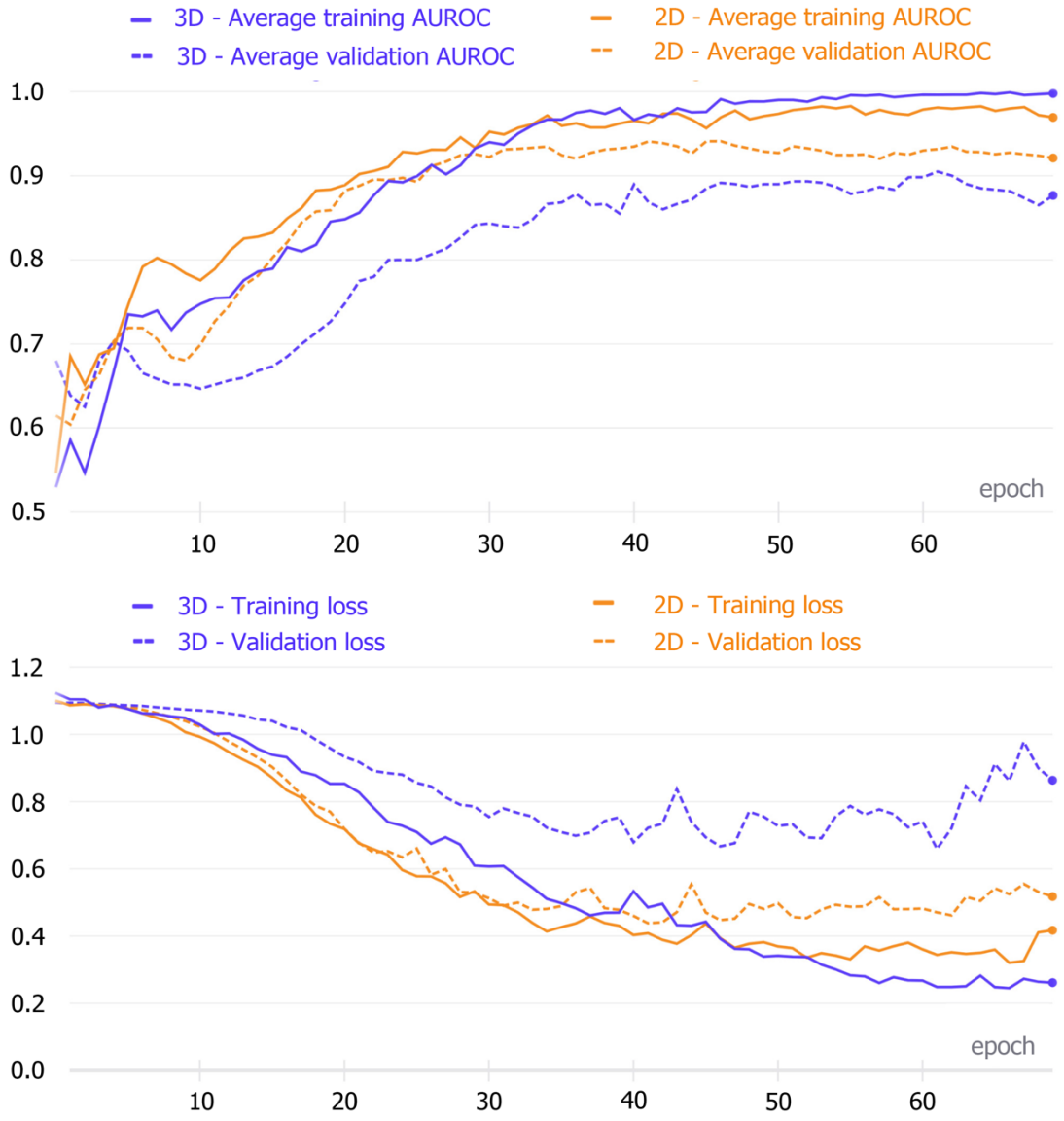

Figure 4. *Training metric curves and loss curves. The upper figure shows the area under the receiver operating characteristic curve (AUC) values after each training epoch of the two-dimensional and three-dimensional densely connected convolutional neural networks (DenseNets). The lower figure indicates the loss values for each training epoch of the two networks. 2D: Two-dimensional; 3D: Three-dimensional; AUC: Area under the receiver operating characteristic curve. (17)*

These model weights were saved and then used to make test set predictions (Figure 5). The finalized 2D and 3D networks were evaluated on the same test set, consisting of 10 lesions from each tumor type. On this test dataset, the finalized 2D model achieved 0.9900 (0.9664–1.0000), 0.9600 (0.8786–1.0000) and 0.9950 (0.9811–1.0000) AUC values for FNH, HCC and MET. The average AUC was 0.9783 (0.9492–1.0000). The AUC of the finalized 3D model was 0.9700 (0.9077–1.0000), 0.9050 (0.7889–1.0000) and 0.9550 (0.8890–1.0000) for FNH, HCC and MET diagnosis. The average AUC value was 0.9433 (0.8942–0.9924) on the test dataset (Figure 5). No statistically significant difference was found between the diagnostic performance of the 2D, and 3D classifiers based on the ROC curve comparison for the three classes ($p = 0.4835$ for FNH; $p = 0.4347$ for HCC; $p = 0.1913$ for MET). The 2D CNN achieved comparable results to the 3D network (Table 4). The highest diagnostic accuracy was achieved by both networks for FNH and MET, while both networks demonstrated lower AUC values for HCC (Table 4). PPV, sensitivity, f1 score, specificity and an NPV of 0.9091, 1.0000, 0.9524, 0.9500, 1.000 values were achieved by the 2D model for FNH diagnosis. The 3D network performed FNH classification with similar PPV (0.9000), sensitivity (0.9000), f1 score (0.9000), specificity (0.9500) and NPV (0.9500) values as the 2D network. During HCC classification both the 2D and 3D models reached acceptable metrics with PPVs of 1.000 and 0.8750, sensitivities of 0.8000 and 0.7000, f1 scores of 0.8889 and 0.7778, specificities of 1.000 and 0.9500, lastly NPVs of 0.9091 and 0.8636. For the differentiation of METs from FNHs and HCCs the use of the 2D DenseNet resulted in a PPV of 0.9091, sensitivity of 1.000, f1 score of 0.9524, specificity of 0.9500 and NPV of 1.000, while the 3D DenseNet achieved values of 0.7500, 0.9000, 0.8182, 0.8500 and 0.9444 for PPV, sensitivity, f1 score, specificity and NPV respectively. On average, both the 2D and 3D trained models could distinguish FNHs, HCCs and METs reliably with PPVs of 0.9394 and 0.8417, sensitivities of 0.9333 and 0.8333, f1 scores of 0.9312 and 0.8320, specificities of 0.9667 and 0.9167, NPVs of 0.9697 and 0.9194.

DOI:10.14753/SE.2025.3142

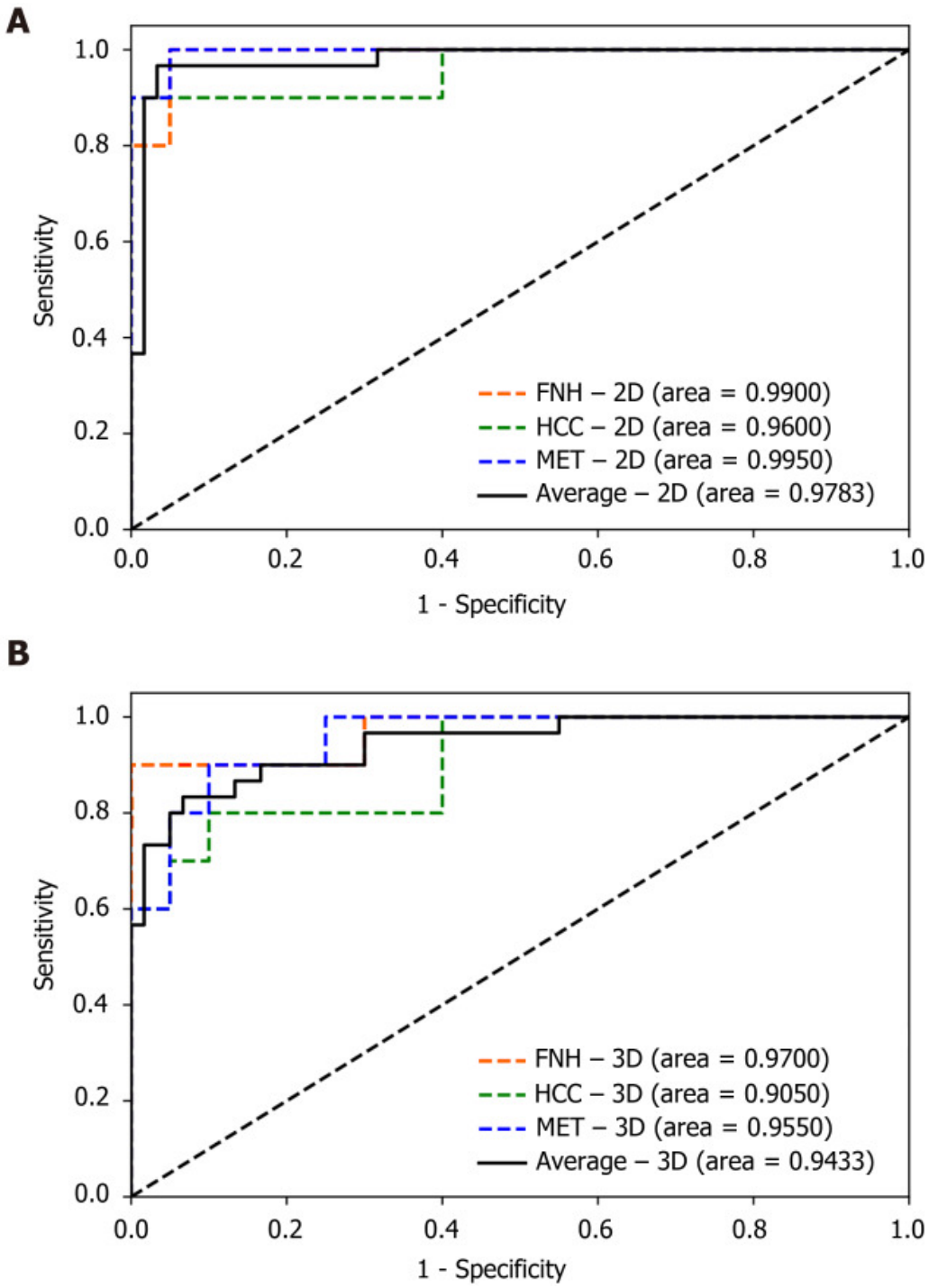

Figure 5. *Receiver operating characteristic curves. A: Two-dimensional; B: Three-dimensional. 2D: Two-dimensional; 3D: Three-dimensional; FNH: Focal nodular hyperplasia; HCC: Hepatocellular carcinoma; MET: Metastasis. (17)*

Table 4. *Evaluation metrics of the two-dimensional and three-dimensional densely connected convolutional neural networks. (17)*

Input data	PPV	Sensitivity	F1 score	Specificity	NPV
FNH 2D	0.9091	1.0000	0.9524	0.9500	1.0000
3D	0.9000	0.9000	0.9000	0.9500	0.9500
HCC 2D	1.0000	0.8000	0.8889	1.0000	0.9091
3D	0.8750	0.7000	0.7778	0.9500	0.8636
MET 2D	0.9091	1.0000	0.9524	0.9500	1.0000
3D	0.7500	0.9000	0.8182	0.8500	0.9444
Mean 2D	0.9394	0.9333	0.9312	0.9667	0.9697
3D	0.8417	0.8333	0.8320	0.9167	0.9194

2D: Two-dimensional; 3D: Three-dimensional; FNH: Focal nodular hyperplasia; HCC: Hepatocellular carcinoma; MET: Metastasis; NPV: Negative predictive value; PPV: Positive predictive value.

In addition, these results are supported by the extraction of attention maps from the trained models using test set images. We used an open-source software (M3d-CAM) to visualize the most important regions for diagnosis-making (37). The extracted attention maps may correlate with the certainty with which a model classifies FLLs (Figure 6).

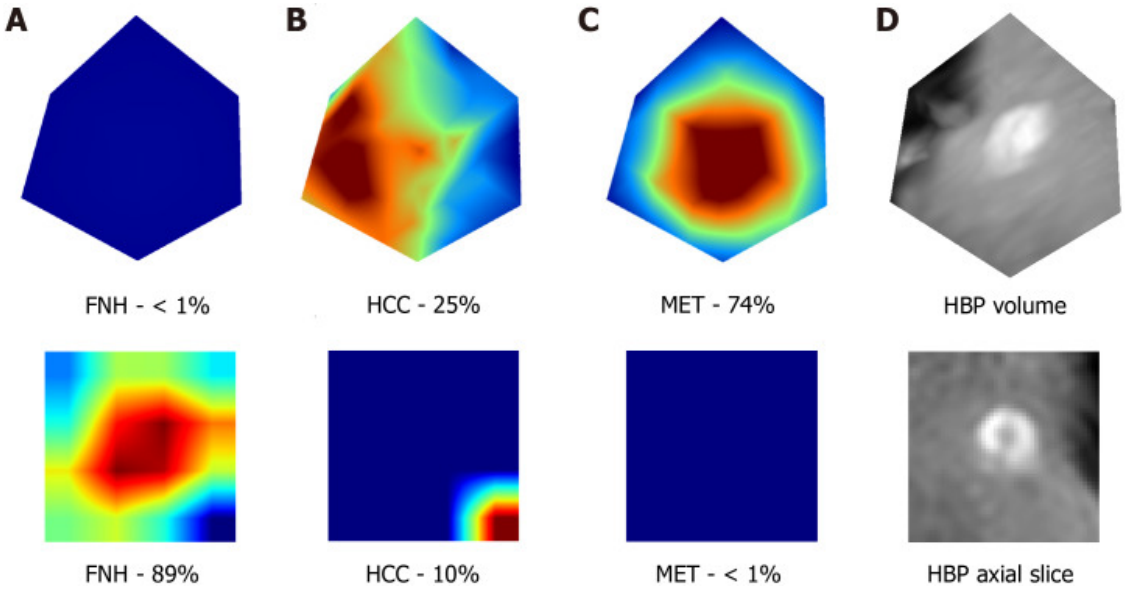

Figure 6. *Visualization of the attention maps compared to the hepatobiliary phase input images. Two-dimensional (lower row) and three-dimensional (upper row) attention maps*

(column A-C) and hepatobiliary phase (column D) images were extracted from the 3rd dense block of the trained network. A-C: Two-dimensional (lower row) and three-dimensional (upper row) attention maps; D: Hepatobiliary phase images. Column A contains the attention maps for focal nodular hyperplasia (FNH), column B for hepatocellular carcinoma, and column C for metastasis probabilities. The correct diagnosis is FNH in this case. Probabilities for different lesion classes are annotated below each attention map. The red areas are more important for the classification than other image regions. FNH: Focal nodular hyperplasia; HCC: Hepatocellular carcinoma; MET: Metastasis; HBP: Hepatobiliary phase. (17)

4.2. Identification of radiological features

After training each model with multiple hyperparameter setups, the highest validation mean AUC (0.9147) was achieved by the EfficientNetB0 model after 480 epochs. In this setting, the network was trained with a batch size of 32. We provide the training results of the other model architectures as well in decreasing order, based on validation mean AUC: EfficientNetB6 (0.9033), EfficientNetB2 (0.9033), EfficientNetB3 (0.902), EfficientNetB4 (0.8988), EfficientNetB1 (0.8922), EfficientNetB5 (0.8922), DenseNet121 (0.8807), DenseNet169 (0.8792), DenseNet201 (0.8733), DenseNet264 (0.8682), EfficientNetB7 (0.856). The final EfficientNetB0 model could identify most features with excellent metrics when tested on the independent test dataset. Table 5 summarizes the results for each feature, including all lesion types. The highest AUCs were reached for the detection of delayed phase enhancement (0.99) and iso- or hyperintensity on the venous phase (0.98). These features were only rarely detected as false positives or remained undetected. The least predictable features based on AUC were T2 hyperintensity (0.79), peripheral enhancement (0.74), and washout (0.64). ROC curves and corresponding AUC values are shown in Figure 7.

DOI:10.14753/SE.2025.3142

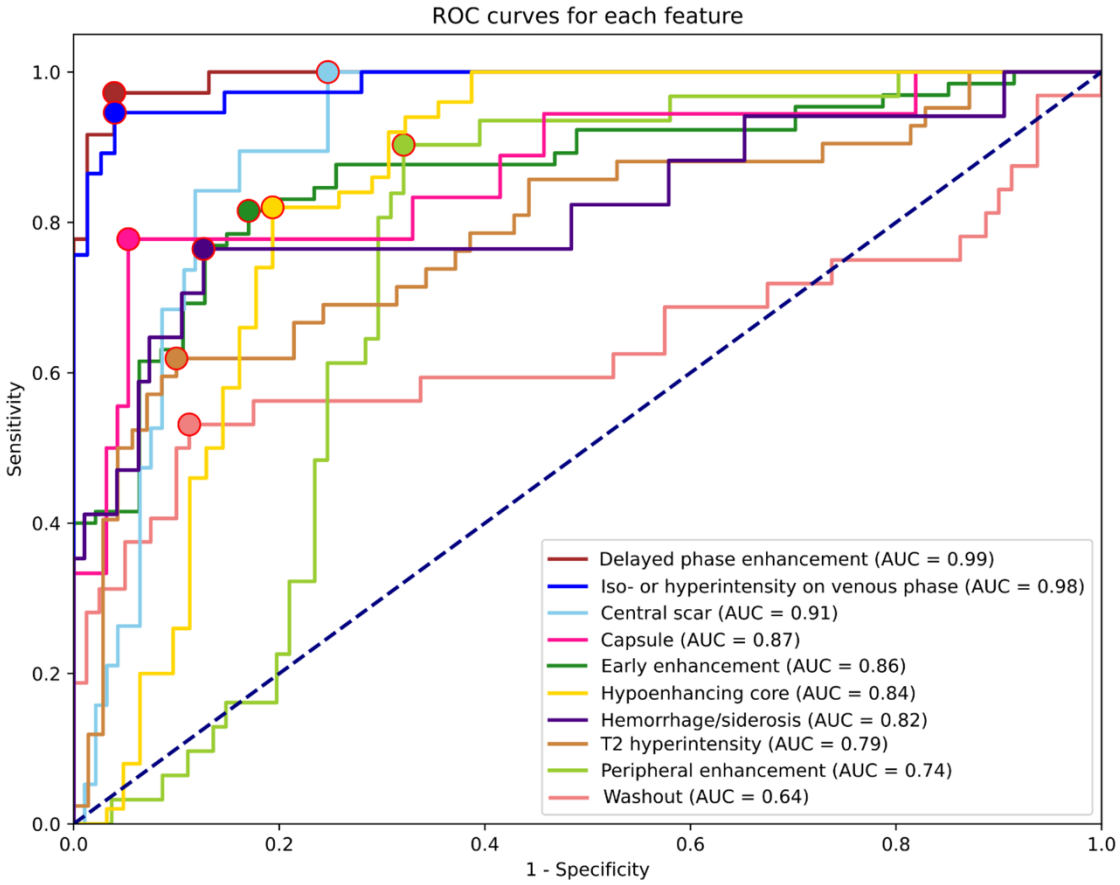

Figure 7. *Receiver operating characteristic (ROC) curves for each feature are based on the test dataset predictions. The colored dots indicate the cut-off points used to calculate metrics for the specific feature. AUC: Area Under the ROC curve. (30)*

Table 5. *Test dataset metrics. (30)*

Radiological Features	PPV	NPV	Sensitivity	Specificity	f1	AUC	Power
Delayed phase enhancement	0.92	0.97	0.94	0.96	0.93	0.99	1
Iso- or hyperintensity on venous phase	0.92	0.96	0.92	0.96	0.92	0.98	1
Central scar	0.44	0.99	0.95	0.75	0.60	0.91	1
Capsule	0.72	0.95	0.72	0.95	0.72	0.87	1
Early enhancement	0.87	0.75	0.80	0.83	0.83	0.86	1
Hypoenhancing core	0.77	0.83	0.80	0.81	0.78	0.84	1
Hemorrhage/siderosis	0.50	0.94	0.71	0.87	0.59	0.82	0.99
T2 hyperintensity	0.78	0.79	0.60	0.90	0.68	0.79	1
Peripheral enhancement	0.51	0.93	0.87	0.68	0.64	0.74	0.98
Washout	0.64	0.82	0.50	0.89	0.56	0.64	0.64
Mean values	0.71	0.89	0.78	0.86	0.73	0.84	-
SD values	0.17	0.08	0.14	0.09	0.13	0.10	-

PPV: positive predictive value, NPV: negative predictive value, AUC: area under the receiver operator characteristic curve, SD: standard deviation.

The highest and lowest PPVs were reached for delayed phase enhancement (0.92) and central scar (0.44) detection, while the best and worst sensitivities were for central scar (0.95), delayed phase enhancement (0.94), and iso- or hyperintensity on venous phase (0.92) vs. T2 hyperintensity (0.60) and washout (0.50). NPVs and specificities were higher on average (0.89, 0.86) than PPVs and sensitivities (0.71, 0,78). Apart from early enhancement (0.75) and T2 hyperintensity (0.79), all other NPVs were above 0.8. The feature with the lowest specificity was peripheral enhancement (0.68), while the most specific was delayed phase enhancement (0.96). As shown in Table 4, almost all feature AUCs were calculated with power reaching 0.98; therefore, the number of samples is more than sufficient to support these results.

Table 6. *Results for annotated features: focal nodular hyperplasia. (30)*

Radiological Features	True Positives	True Negatives	False Positives	False Negatives	f1
Delayed phase enhancement	34	0	0	2	0.97
Iso- or hyperintensity on venous phase	34	0	0	2	0.97
Early enhancement	32	1	2	1	0.96
Central scar	18	0	17	1	0.67
Washout	0	34	2	0	0
Peripheral enhancement	0	33	3	0	0
Capsule	0	34	2	0	0
T2 hyperintensity	0	32	0	4	0
Hemorrhage/siderosis	0	32	4	0	0
Hypoenhancing core	0	36	0	0	-

To be able to explore the differences in predictions between the different lesion types, results are reported for FNHs, HCCs, and METs separately as well (Table 6, Table 7, and Table 8). Since not all features are present in all lesion types, not all metrics can be calculated for all features in each case. To simplify this problem, feature predictions are ordered according to their respective f1 scores. To provide more details on false detections, non-abundant features are also listed for each lesion type. Features present in FNHs were generally well recognizable by the model. Features related to contrast enhancement that are representative of FNHs, such as early or delayed phase enhancement, had f1 scores above 0.95, while non-present features were rarely detected. Central scars were common false positive detections, but mostly if the lesion was FNH (Figure 8). If the lesion analyzed was HCC (Table 7) or MET (Table 8), the model almost never predicted the presence of a central scar.

DOI:10.14753/SE.2025.3142

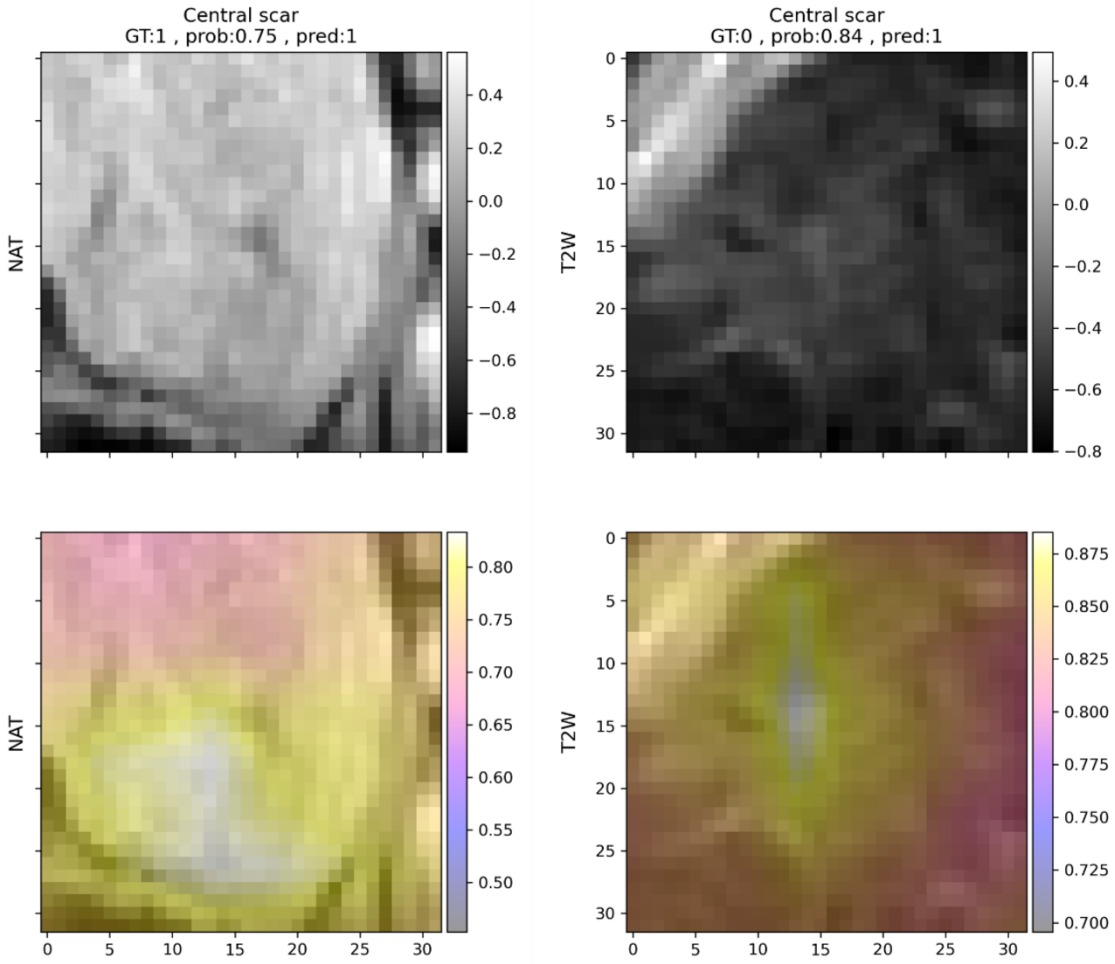

Figure 8. *Two examples (in each column) of central scar predictions in focal nodular hyperplasias. Left: correct prediction, right: incorrect prediction. Upper row: native T1-weighted (left) and T2-weighted (right) images. Rescaled voxel intensities are indicated on the y-axis. Lower row: occlusion sensitivity maps indicating the contribution of each voxel to the prediction. In the case of the T2-weighted image, the area representing the central scar presumably increases the probability of the identification of this feature. In the case of the native T1-weighted image, the areas near the central scar led to the highest increase in the prediction probability. GT: ground truth, prob: probability, pred: prediction, NAT: native T1-weighted image, T2W: T2-weighted image. (30)*

Table 7. *Results for annotated features: hepatocellular carcinoma. (30)*

Radiological Features	True Positives	True Negatives	False Positives	False Negatives	f1
Capsule	13	18	1	5	0.81
Early enhancement	19	6	1	11	0.76
Hemorrhage/siderosis	12	14	6	5	0.69
Washout	16	3	2	16	0.64
Hypoenhancing core	11	12	9	5	0.61
T2 hyperintensity	3	26	6	2	0.43
Delayed phase enhancement	0	35	2	0	0
Peripheral enhancement	0	20	17	0	0
Central scar	0	34	3	0	0
Iso- or hyperintensity on venous phase	0	34	2	1	0

Table 8. *Results for annotated features: liver metastasis. (30)*

Radiological Features	True Positives	True Negatives	False Positives	False Negatives	f1
Peripheral enhancement	27	0	0	3	0.95
Hypoenhancing core	26	0	0	4	0.93
T2 hyperintensity	18	3	1	8	0.8
Early enhancement	0	26	4	0	0
Washout	0	25	5	0	0
Central scar	0	29	1	0	0
Capsule	0	28	2	0	0
Hemorrhage/siderosis	0	28	2	0	0
Delayed phase enhancement	0	30	0	0	-
Iso- or hyperintensity on venous phase	0	30	0	0	-

Among all lesion types, HCC feature prediction yielded the least desirable results in this analysis as well. As reported in Table 7, diagnostically important features, namely washout and early enhancement, were undetected in half and nearly half of all cases. Features that are present in both HCCs and METs, such as peripheral enhancement (MET), were common false positive findings in the HCC group, but not in the MET group. Capsule was less difficult to detect, but peripheral enhancement was falsely detected in half of the analyzed cases, possibly due to the similarity between the two. Although hemorrhage was reported only in HCCs by the expert annotator, the algorithm predicted it in four cases in FNHs as well, and two cases in METs. Hemorrhage in HCCs remained undetected in one-third of cases, like hypoenhancing core (Table 7). Features related to contrast enhancement were detected less accurately in HCCs. Hypoenhancing core was falsely detected in nine cases and missed in five cases. The presence of other similar features such as early enhancement or hemorrhage might make the detection of a hypoenhancing core more difficult.

The most common mistake in the case of METs was the underdiagnosis of T2 hyperintensity (eight cases), which was most commonly marked in this group (Table 8). Features mostly present in FNHs were almost perfectly predicted (Table 8), while washout and early enhancement were the most common falsely detected features. Both peripheral enhancement and hypoenhancing core were identified with an f1 score above 0.9. For an example of hypoenhancing core prediction, see Figure 9.

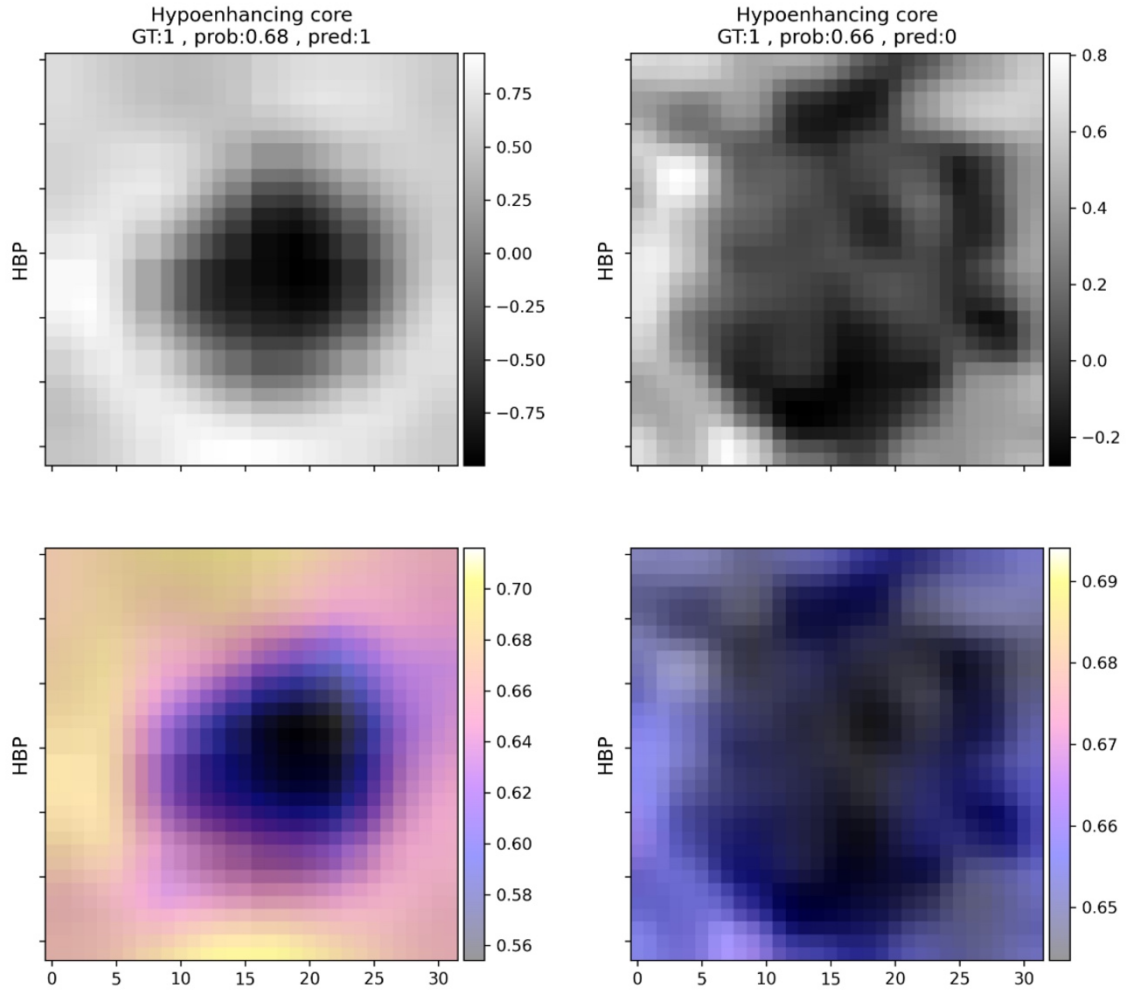

Figure 9. *Examples of hypoenhancing core predictions in liver metastasis (left) and hepatocellular carcinoma (right). Left: correct prediction, right: incorrect prediction. Upper row: processed hepatocyte-specific contrast-enhanced scans. Rescaled voxel intensities are indicated on the y-axis. Lower row: occlusion sensitivity maps indicating the contribution of each voxel to the prediction. These maps indicate the prediction probability of the model for the hypoenhancing core feature, while the corresponding part of the image is replaced by the mean intensity value of the image. In the shown cases the image area that represents the hypoenhancing core is replaced by higher values (which makes the hypoenhancing core disappear), thus decreasing the probability of the identification of this feature. GT: ground truth, prob: probability, pred: prediction, HBP: hepatocyte-specific contrast-enhanced image. (30)*

Features reported in other lesion types were variably predictable (Table 8). Peripheral enhancement might be confused with nodular enhancement, exhibited by

hemangiomas, which was not explicitly analyzed, as only a low number of cases were available. Hypoenhancing core represents a similar case, as both cysts and hemangiomas may mislead the model predictions due to their enhancement characteristics.

Table 8. *Results for annotated features: other lesions. (30)*

Radiological Features	True Positives	True Negatives	False Positives	False Negatives	f1
T2 hyperintensity	4	2	0	3	0.73
Hypoenhancing core	3	2	3	1	0.6
Early enhancement	1	6	1	1	0.5
Delayed phase enhancement	0	8	1	0	0
Peripheral enhancement	0	2	6	1	0
Central scar	0	7	2	0	0
Iso- or hyperintensity on venous phase	0	8	1	0	0
Washout	0	9	0	0	-
Capsule	0	9	0	0	-
Hemorrhage/siderosis	0	9	0	0	-

Cohen's Kappas scores for each feature in each combination are reported in (Table 9). The mean Kappa score was 0.60 for the agreement between the predictive model and the expert, similar to novice opinion compared to model predictions, indicating moderate reliability. In the case of delayed phase enhancement and venous phase iso- or hyperintensity, the agreement was almost perfect (> 0.8). Even the worst feature predictions (central scar, peripheral enhancement, capsule) showed moderate agreement (> 0.4) with the expert opinion. Features that were less accurately predicted by the network were also subject to disagreement between the two human observers. Central scar, for example, was more frequently identified by both the model and the radiology resident, while only moderate agreement was observable in the case of washout in all three comparisons.

Table 9. *Interobserver agreement between the three observers, measured by Cohen's Kappa. (30)*

Radiological Features	Model vs. Expert	Model vs. Novice	Novice vs. Expert
Delayed phase enhancement	0.90	0.76	0.82
Iso- or hyperintensity on venous phase	0.88	0.73	0.77
Capsule	0.67	0.51	0.76
Early enhancement	0.62	0.59	0.82
Hypoenhancing core	0.60	0.54	0.83
T2 hyperintensity	0.52	0.52	0.81
Hemorrhage/siderosis	0.50	0.42	0.76
Central scar	0.48	0.68	0.66
Peripheral enhancement	0.45	0.36	0.79
Washout	0.41	0.53	0.59
Mean values	0.60	0.56	0.76
SD values	0.16	0.12	0.07

SD: standard deviation.

5. Discussion

FLLs are common findings during liver imaging, and the differentiation of benign and malignant types of FLLs is a significant diagnostic challenge, as imaging signs may overlap between different pathologies which can substantially alter the therapeutic decision. Therefore, precise and reproducible differential diagnosis of FLLs is critical for optimal patient management. Today, the most accurate imaging modality to diagnose FLLs is multi-phase dynamic contrast-enhanced MRI. Extracellular contrast agents (ECAs) are commonly used to perform multi-phase dynamic post-contrast MRI studies to differentiate between lesions based on their distinct contrast enhancement patterns, such as HAP hyper-enhancement or washout in the PVP (5). In comparison to ECAs, HSCs are taken up by hepatocytes and (in part) excreted through the biliary tract; thus, they can better differentiate between those lesions that consist of functionally active and impaired hepatocytes or those that are extrahepatic in origin (38). This behavior of HSCs is utilized for making a distinction between FNH and hepatocellular adenoma, or to detect small foci of HCC and MET within the surrounding liver parenchyma (39,40).

In the current study, we evaluated different AI models on liver MRI images for the prediction of FLLs compiled from three different types of lesions, namely FNHs, HCCs and METs. To ensure that the models could achieve the highest possible prediction rate, we narrowed down our data collection to only those four MRI sequences that provided the highest tissue contrast compared to the neighboring parenchyma or depicted distinctive imaging features of the lesion types. For the same reason, we used only HSC-enhanced scans for the analysis. We collected post-contrast images from HAP, PVP and HBP, and a T2w SPAIR image in the case of each lesion. A similar image analysis strategy was used by Hamm *et al* (16), who predicted 494 FLLs from six categories, including simple cyst, cavernous hemangioma, FNH, HCC, intrahepatic cholangiocarcinoma, and colorectal cancer METs using a 3D CNN model. The authors used HAP, PVP and delayed venous phase MRI images for the classification of the FLLs. They reported that the CNN model demonstrated 0.92 accuracy, 0.92 sensitivity and 0.98 specificity. The disadvantage of this study compared to ours was that it did not include HBP images, with only ECA images used for the MRI scans.

There are a handful of studies that included conventional ML methods and achieved reasonably good results. Wu *et al* (41), for example, extracted radiomics

features from non-enhanced multi-parametric MRI images of FLLs and used them in ML models to differentiate between hepatic haemangioma and HCC. The final classifier achieved an AUC of 0.89, a sensitivity of 0.822 and a specificity of 0.714. Jansen *et al* (14), in their 2019 paper, used traditional ML methods for the same problem achieving an average accuracy of 0.77 for five major FLL types.

Our models' performance in the test set was comparable to those from previous publications, as the AUC, sensitivity and specificity were excellent for both the 2D (0.9783, 0.9333 and 0.9667 respectively) and 3D (0.9433, 0.8333 and 0.9167 respectively) architectures, which demonstrates the robustness of our data collection and analysis.

The quality and quantity of input data are pivotal when training neural networks. MRI liver tumor analysis using DL methods has steeply increased, but there is evidence lacking to support the use of 2D or 3D data. The additional dimension in 3D network inputs makes them computationally more demanding and the different data augmentation methods and hyperparameters must be well chosen to avoid artifacts. Our study supports the results of Wang *et al* (23) and Hamm *et al* (16), emphasizing the need for multi-channel input volumes in order to achieve better accuracy. In contrast to these approaches, we have also utilized HBP images, thereby increasing the number of input channels to four in order to improve accuracy and additionally trained 2D CNNs, proving them to be just as effective classifiers as 3D models.

The selected architecture of the DL model can substantially alter classification accuracy. It is a novelty of our analysis that compared to previous examinations we utilized a DenseNet architecture. DenseNets contain multiple dense blocks, where each layer is connected with the residuals from previous layers. DenseNets require fewer trainable parameters at the same depth than conventional CNNs, as newly learned features are shared through all layers (27). Our results are among the first to indicate that this highly efficient network design can enhance the performance of AI models for the classification of multi-parametric MRI images of FLLs.

Our first study's limitations are the low number of patients involved, the retrospective nature of the study, that it was conducted within a single institute, and thus lack of external evaluation. Additionally, a patient-level data split, instead of lesion level split, would have increased the independence of the test dataset samples. Further

improvement may be achieved by additional data collection (including additional lesion classes) and the use of more MRI volumes and different data augmentation methods as well as the use of pre-trained networks. Classification should also possibly be extended to more lesion classes.

Our second study explores our findings on how DLMs may perform on a small, single-institutional dataset concerning a complex reporting task. Several research groups have reported excellent results on the automatic, DLM-based multi-label classification of various types of FLLs, but these put more emphasis on predicting lesion class and less emphasis on mimicking the human observers (16,17,19,42). More interpretable methods have been described in radiology in general, the most obvious one being chest X-ray reporting using deep learning methods, where multiple findings have to be identified in parallel by the AI (43). While chest X-ray interpretation is among the most advanced research areas in deep learning radiomics, other examination types and areas with less frequently performed studies and much more complex reporting tasks lack sufficient proof for the application of AI methods. Research on radiological feature descriptors is also of importance as many of the lesions are multifocal, many types may be found parallel, and histological confirmation cannot be acquired in all cases, thus, a certain diagnosis may not be possible (and necessary) for all lesions. Additionally, the described features allow a much broader extension of applications, since each may allow the user to draw different conclusions, such as whether tumor recurrence is observable (enhancement) or whether the malignant transformation of a regenerative nodule has occurred. Additionally, these models reproducibly give the same output on the same input images and because of this they can be used a metrics during research projects.

These are partly the reasons why the main emphasis of the second study is on radiological feature identification. Although the classification of different FLLs based on the identified features could seem like a straightforward task, various challenges promote it to a research topic on its own. While the majority of the lesions evaluated in our study fell into three main lesion types, the liver is host to one of the largest varieties of focal pathologies; as such, it would be worth examining diagnostic algorithms built upon the present feature identifier in a more detailed manner. As such, they should be evaluated on a larger variety of pathologies. There are multiple lesion types, for example, cholangiocellular carcinoma, that are not present in the current dataset, but in future

studies should be evaluated, considering their clinical importance. Apart from this, further evaluation in this direction could be carried out in multiple ways that did not fit within the scope of the current manuscript. A classifier model could be built solely for the diagnosis of FLLs, as well as by reusing the currently presented feature identifier, for example, via transfer learning. In this case, the training of the model would be guided to take into account the radiological features identifiable by human observers, apart from deep features. The diagnosis of the tumor could also be based on the probabilities of the predictions for each feature. In this case, the top-N features would be used to create an algorithmic approach for diagnosis making. Our current interpretation of the feature detector partly opposes this approach, as the predictions of the model are evaluated based on the calculated optimal cut-off values. Apart from these, there could be other ways to create a diagnostic model that integrates the feature identifier for better interpretability. Because of this, the automatic classification of focal liver lesions lies outside the scope of the current paper.

Abdominal imaging studies, such as HSC MRI, are less frequently approached in a similar manner due to the higher cost of imaging, the complexity of the task, the smaller amount of available data, and the more variable agreement on radiological feature abundance among professionals, as well as the need for more time-consuming data preparation and analyses. Most papers use some form of deep learning interpretation method, such as attention maps, to try to find explanations for classification predictions, while direct feature predictions have rarely been the focus of research. Wang et al., in their 2019 study, were among the first to use CNNs for focal liver lesion feature identification (23). The reported model was able to correctly identify radiological features present in test lesions with 76.5% PPV and 82.9% sensitivity, which is similar to our results, though their method was built on a precious lesion classifier, from which feature predictions were derived. Our study deliberately avoided the diagnosis of lesions and focused solely on feature identification. Sheng et al. also used deep learning to predict radiological features based on gadoxetate disodium-enhanced MRI, dedicated to LI-RADS grading in an automated and semi-automatic manner. They reported AUCs of 0.941, 0.859, and 0.712 (internal testing) for arterial phase enhancement, washout, and capsule prediction. The model was also tested on an external test set, achieving AUCs of 0.792, 0.654, and 0.568, respectively (44). Though they evaluated fewer features, similarly to our findings, arterial

phase enhancement was more accurately predictable than washout and capsule, both of which are challenging for the AI to predict. The results of Wang et al. also led to a similar conclusion, as arterial phase hyperenhancement and delayed phase hyperenhancement, among others, just as according to our results, were well predictable features, while others, such as central scar and washout, were especially difficult to accurately predict (23). Central scar and washout were also difficult to identify and were quite often false positive findings; furthermore, in our experience, circle-like features such as peripheral enhancement, which might be confused with capsule by the model, were just as common false positive findings. The difficulty in the detection of these features is consistent with previous research on gadoxetate disodium, as HCC indicative features, such as capsule and washout, are less distinguishable using gadoxetate disodium than with extracellular contrast agents (45). Delayed phase enhancement, which is related to the hepatocyte-specific nature of gadoxetate disodium, was an accordingly straightforward prediction.
In the current study various occlusion sensitivity maps are shown that attempt to visually explain the decision-making process of the neural network classifier. The maps can be helpful in explaining the decision-making process even in a very complex task and can draw attention to erroneous decision-making that may be based on, for example, image artifacts or non-task-related image areas. The modification of the padding value from the image mean intensity to specific values depending on scan type and predicted radiological feature may be a promising direction for further investigation.

As mentioned previously, features on which there might be disagreement between expert radiologists as well (e.g., central scar) are more difficult to build a model upon. In the future, it is possible that more thorough curation of training data based on the opinions of multiple experts would be necessary to optimize these methods. A promising research direction would be a more detailed examination of how each image, their quality, and the reported expert consensus could be used to construct balanced, high-quality datasets that are more representative of radiological liver lesion features. The current study has additional limitations. It was retrospectively conducted within a single institute, and only a small number of patients were included. To mitigate the consequences of these problems, further multi-institutional studies are needed. Additional methods, such as transfer learning with other, similar, multi-modal datasets may be used in addition to the previously mentioned dataset reannotation. Further data augmentation methods, such as

random cropping, also must be evaluated. Splitting the model into multiple feature predictors based on conflicting features and corresponding scans may also be examined as a potential solution for inaccurate predictions (e.g., T2 hyperenhancement and hypoenhancing core). Apart from these, the tested methodology has the potential to aid less experienced radiologists or other clinicians in understanding and interpreting HSC MRI of FLLs in an automated, controllable manner by providing predictions of radiological features in a few seconds.

DOI:10.14753/SE.2025.3142

6. Conclusions

Based on our studies, we can state that routinely acquired radiological image materials can be used for analysis with AI methods, such as CNNs. According to our results, densely connected CNNs trained on multi-sequence MRI scans can be promising new alternatives to single-phase approaches; furthermore, the use of multi-dimensional input volumes can help the AI-based diagnosis of FLLs. According to our results, 3D and 2D DenseNets can reach similar performance in the differentiation of FLLs based on MRI images.

Based on our findings regarding the automatic identification of radiological features in focal liver lesions AI model predictions are reliable, and they could provide descriptions of radiological features present in FLLs, putting more weight on the exclusion of a feature and allowing false positive predictions depending on the type of lesion and features present. Mistakes may partly be due to human uncertainty or the lack of consensus among experts on the definition of a given radiological feature, not to mention various imaging artifacts and image processing errors that may make proper predictions more difficult.

Altogether DL-based approaches could aid clinicians and medical researchers, especially when large quantities of images need to be processed or an automatically utilizable reference is required.

DOI:10.14753/SE.2025.3142

7. Summary

The current thesis discusses the development and evaluation of recently introduced deep learning (DL) methods for the automated classification and description of focal liver lesions (FLLs) based on magnetic resonance imaging (MRI) using hepatocyte-specific contrast agents (HSCs). As the first study of its kind in Hungarian literature, it demonstrates that convolutional neural networks (CNNs) can effectively differentiate focal nodular hyperplasia (FNH), hepatocellular carcinoma (HCC), and liver metastases (MET) based on both two-dimensional (2D) and three-dimensional (3D) information. Furthermore, various radiological features of focal liver lesions are automatically identifiable by CNNs, such as DenseNets and EfficientNets. These advancements enable the development of downstream imaging methods, particularly in deep learning reconstruction, which requires the definition of a ground truth/reference to which a generated image is compared.

The study methodology involved the manual collection and preparation of large quantities of MRI data. Data processing includes anonymization, resampling, and alignment of MRI scans, followed by the creation of training, validation, and test datasets. All architectures were employed within the MONAI framework for analysis and custom written computer code was written by our research group for both analyses.

Our first study found that both 2D and 3D CNNs effectively differentiate between FNH, HCC, and MET with AUCs above 0.90. In our second study EfficientNetB0 was identified as the top-performing model for radiological feature identification, achieving the highest validation mean AUC (0.9147) after 480 epochs.

The integration of such models into clinical practice faces challenges, such as the need for larger, multi-institutional datasets and further validation studies as well as more detailed large scale data annotation.

Our research concludes that DL techniques, particularly CNNs, are promising tools for enhancing MRI-based diagnosis of FLLs. The use of HSC-enhanced MRI combined with advanced DL models shows high diagnostic accuracy, aiding early and precise diagnosis of liver conditions. In summary, this dissertation demonstrates significant advancements in applying DL to medical imaging, providing a robust framework for future research and clinical integration in diagnosing focal liver lesions.

8. References

1. Perez-Lopez R, Ghaffari Laleh N, Mahmood F, Kather JN. A guide to artificial intelligence for cancer researchers. Nat Rev Cancer. 2024 Jun;24(6):427–441.

2. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018 Aug;9(4):611–629.

3. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–444.

4. Bordes F, Pang RY, Ajay A, Li AC, Bardes A, Petryk S, Mañas O, Lin Z, Mahmoud A, Jayaraman B, Ibrahim M, Hall M, Xiong Y, Lebensold J, Ross C, Jayakumar S, Guo C, Bouchacourt D, Al-Tahan H, Padthe K, Sharma V, Xu H, Tan XE, Richards M, Lavoie S, Astolfi P, Askari Hemmat R, Chen J, Tirumala K, Assouel R, Moayeri M, Talattof A, Chaudhuri K, Liu Z, Chen X, Garrido Q, Ullrich K, Agrawal A, Saenko K, Celikyilmaz A, Chandra V. An Introduction to Vision-Language Modeling [Internet]. arXiv; 2024 [cited 2024 Jun 4]. Available from: https://arxiv.org/abs/2405.17247

5. Matos AP. Focal liver lesions: Practical magnetic resonance imaging approach. World J Hepatol. 2015;7(16):1987.

6. Neri E, Bali MA, Ba-Ssalamah A, Boraschi P, Brancatelli G, Caseiro Alves F, Grazioli L, Helmberger T, Lee JM, Manfredi R, Martì-Bonmatì L, Matos C, Merkle EM, Op De Beeck B, Schima W, Skehan S, Vilgrain V, Zech C, Bartolozzi C. ESGAR consensus statement on liver MR imaging and clinical use of liver-specific contrast agents. Eur Radiol. 2016 Apr;26(4):921–931.

7. Lafaro KJ, Roumanis P, Demirjian AN, Lall C, Imagawa DK. Gd-EOB-DTPA-Enhanced MRI for Detection of Liver Metastases from Colorectal Cancer: A Surgeon's Perspective! Int J Hepatol. 2013;2013:1–7.

8. Renzulli M, Clemente A, Ierardi AM, Pettinari I, Tovoli F, Brocchi S, Peta G, Cappabianca S, Carrafiello G, Golfieri R. Imaging of Colorectal Liver Metastases: New Developments and Pending Issues. Cancers. 2020 Jan 8;12(1):151.

9. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. Cancer Res. 2014 Jun 1;74(11):2913–2921.

10. Arif-Twari H, Kalb B, Chundru S, Sharma P, Costello J, Guessner RW, Martin DR. MRI of hepatocellular carcinoma: an update of current practices. Diagn Interv Radiol. 2014 May 2;20(3):209–221.

11. Yoneda N, Matsui O, Kitao A, Kozaka K, Kobayashi S, Sasaki M, Yoshida K, Inoue D, Minami T, Gabata T. Benign Hepatocellular Nodules: Hepatobiliary Phase of Gadoxetic Acid–enhanced MR Imaging Based on Molecular Background. RadioGraphics. 2016 Nov;36(7):2010–2027.

12. Laino ME, Viganò L, Ammirabile A, Lofino L, Generali E, Francone M, Lleo A, Saba L, Savevski V. The added value of artificial intelligence to LI-RADS categorization: A systematic review. Eur J Radiol. 2022 May;150:110251.

13. Hill CE, Biasiolli L, Robson MD, Grau V, Pavlides M. Emerging artificial intelligence applications in liver magnetic resonance imaging. World J Gastroenterol. 2021 Oct 28;27(40):6825–6843.

14. Jansen MJA, Kuijf HJ, Veldhuis WB, Wessels FJ, Viergever MA, Pluim JPW. Automatic classification of focal liver lesions based on MRI and risk factors. Deserno TM, editor. PLOS ONE. 2019 May 16;14(5):e0217053.

15. Montagnon E, Cerny M, Cadrin-Chênevert A, Hamilton V, Derennes T, Ilinca A, Vandenbroucke-Menu F, Turcotte S, Kadoury S, Tang A. Deep learning workflow in radiology: a primer. Insights Imaging. 2020 Dec;11(1):22.

16. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, Lin M, Duncan JS, Weinreb JC, Chapiro J, Letzen B. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol. 2019 Jul;29(7):3338–3347.

17. Stollmayer R, Budai BK, Tóth A, Kalina I, Hartmann E, Szoldán P, Bérczi V, Maurovich-Horvat P, Kaposi PN. Diagnosis of focal liver lesions with deep learning-based multi-channel analysis of hepatocyte-specific contrast-enhanced magnetic resonance imaging. World J Gastroenterol. 2021 Sep 21;27(35):5978–5988.

18. Wang SH, Han XJ, Du J, Wang ZC, Yuan C, Chen Y, Zhu Y, Dou X, Xu XW, Xu H, Yang ZH. Saliency-based 3D convolutional neural network for categorising common focal liver lesions on multisequence MRI. Insights Imaging. 2021 Dec;12(1):173.

19. Zhen SH, Cheng M, Tao YB, Wang YF, Juengpanich S, Jiang ZY, Jiang YK, Yan

YY, Lu W, Lue JM, Qian JH, Wu ZY, Sun JH, Lin H, Cai XJ. Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data. Front Oncol. 2020 May 28;10:680.

20. Kim J, Min JH, Kim SK, Shin SY, Lee MW. Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study. Sci Rep. 2020 Jun 11;10(1):9458.

21. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D Deep Learning on Medical Images: A Review. Sensors. 2020 Sep 7;20(18):5097.

22. Baranes L, Chiaradia M, Pigneur F, Decaens T, Djabbari M, Zegaï B, Costentin C, Laurent A, Calderaro J, Rahmouni A, Luciani A. Imaging benign hepatocellular tumors: Atypical forms and diagnostic traps. Diagn Interv Imaging. 2013 Jul;94(7–8):677–695.

23. Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, Lin M, Weinreb JC, Duncan JS, Chapiro J, Letzen B. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. Eur Radiol. 2019 Jul;29(7):3348–3357.

24. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012 Nov;30(9):1323–1341.

25. Klein S, Staring M, Pluim JPW. Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines. IEEE Trans Image Process. 2007 Dec;16(12):2879–2890.

26. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods. 2016 May;264:47–56.

27. Huang G, Liu Z, Pleiss G, Maaten LVD, Weinberger KQ. Convolutional Networks with Dense Connectivity. IEEE Trans Pattern Anal Mach Intell. 2022 Dec 1;44(12):8704–8716.

28. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, Murrey B, Myronenko A, Zhao C, Yang D, Nath V, He Y, Xu Z, Hatamizadeh A, Myronenko A, Zhu W, Liu Y, Zheng M, Tang Y, Yang I, Zephyr M, Hashemian B, Alle S, Zalbagi Darestani M, Budd C, Modat M, Vercauteren T, Wang G, Li Y, Hu Y, Fu Y, Gorman B, Johnson H,

Genereaux B, Erdal BS, Gupta V, Diaz-Pinto A, Dourson A, Maier-Hein L, Jaeger PF, Baumgartner M, Kalpathy-Cramer J, Flores M, Kirby J, Cooper LAD, Roth HR, Xu D, Bericat D, Floca R, Zhou SK, Shuaib H, Farahani K, Maier-Hein KH, Aylward S, Dogra P, Ourselin S, Feng A. MONAI: An open-source framework for deep learning in healthcare [Internet]. arXiv; 2022 [cited 2024 May 31]. Available from: https://arxiv.org/abs/2211.02701

29. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization [Internet]. arXiv; 2014 [cited 2024 May 31]. Available from: https://arxiv.org/abs/1412.6980

30. Stollmayer R, Budai BK, Rónaszéki A, Zsombor Z, Kalina I, Hartmann E, Tóth G, Szoldán P, Bérczi V, Maurovich-Horvat P, Kaposi PN. Focal Liver Lesion MRI Feature Identification Using Efficientnet and MONAI: A Feasibility Study. Cells. 2022 May 5;11(9):1558.

31. Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Trans Med Imaging. 2010 Jan;29(1):196–205.

32. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks [Internet]. arXiv; 2016 [cited 2024 May 25]. Available from: https://arxiv.org/abs/1608.06993

33. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019 [cited 2024 May 25]; Available from: https://arxiv.org/abs/1905.11946

34. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. 2016 [cited 2024 May 25]; Available from: https://arxiv.org/abs/1610.02391

35. Obuchowski NA, Lieber ML, Wians FH. ROC Curves in Clinical Chemistry: Uses, Misuses, and Possible Solutions. Clin Chem. 2004 Jul 1;50(7):1118–1125.

36. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Dec;12(1):77.

37. Gotkowski K, Gonzalez C, Bucher A, Mukhopadhyay A. M3d-CAM: A PyTorch library to generate 3D data attention maps for medical deep learning [Internet]. arXiv; 2020 [cited 2024 May 30]. Available from: https://arxiv.org/abs/2007.00453

38. Thian YL, Riddell AM, Koh DM. Liver-specific agents for contrast-enhanced

MRI: role in oncological imaging. Cancer Imaging. 2013;13(4):567–579.

39. Kim YY, Park MS, Aljoqiman KS, Choi JY, Kim MJ. Gadoxetic acid-enhanced magnetic resonance imaging: Hepatocellular carcinoma and mimickers. Clin Mol Hepatol. 2019 Sep 25;25(3):223–233.

40. Grieser C, Steffen IG, Kramme IB, Bläker H, Kilic E, Fernandez CMP, Seehofer D, Schott E, Hamm B, Denecke T. Gadoxetic acid enhanced MRI for differentiation of FNH and HCA: a single centre experience. Eur Radiol. 2014 Jun;24(6):1339–1348.

41. Wu J, Liu A, Cui J, Chen A, Song Q, Xie L. Radiomics-based classification of hepatocellular carcinoma and hepatic haemangioma on precontrast magnetic resonance images. BMC Med Imaging. 2019 Dec;19(1):23.

42. Oestmann PM, Wang CJ, Savic LJ, Hamm CA, Stark S, Schobert I, Gebauer B, Schlachter T, Lin M, Weinreb JC, Batra R, Mulligan D, Zhang X, Duncan JS, Chapiro J. Deep learning–assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver. Eur Radiol. 2021 Jul;31(7):4981–4990.

43. Çallı E, Sogancioglu E, Van Ginneken B, Van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Med Image Anal. 2021 Aug;72:102125.

44. Sheng R, Huang J, Zhang W, Jin K, Yang L, Chong H, Fan J, Zhou J, Wu D, Zeng M. A Semi-Automatic Step-by-Step Expert-Guided LI-RADS Grading System Based on Gadoxetic Acid-Enhanced MRI. J Hepatocell Carcinoma. 2021 Jun;Volume 8:671–683.

45. Allen BC, Ho LM, Jaffe TA, Miller CM, Mazurowski MA, Bashir MR. Comparison of Visualization Rates of LI-RADS Version 2014 Major Features With IV Gadobenate Dimeglumine or Gadoxetate Disodium in Patients at Risk for Hepatocellular Carcinoma. Am J Roentgenol. 2018 Jun;210(6):1266–1272.

9. Bibliography of the candidate's publications

9.1. Publications related to the present thesis

1. **Stollmayer R**, Budai BK, Rónaszéki A, Zsombor Z, Kalina I, Hartmann E, Tóth G, Szoldán P, Bérczi V, Maurovich-Horvat P, Kaposi PN. Focal Liver Lesion MRI Feature Identification Using Efficientnet and MONAI: A Feasibility Study. Cells. 2022;11(9). **(IF:6.0, 2022)**

2. **Stollmayer R**, Budai BK, Tóth A, Kalina I, Hartmann E, Szoldán P, Bérczi V, Maurovich-Horvat P, Kaposi PN. Diagnosis of focal liver lesions with deep learning-based multi-channel analysis of hepatocyte-specific contrast-enhanced magnetic resonance imaging. World J Gastroenterol. 2021;27(35):5978-5988. **(IF:5.374, 2021)**

9.2. Publications not related to the present thesis

1. Zsombor Z, Rónaszéki AD, Csongrády B, **Stollmayer R**, Budai BK, Folhoffer A, Kalina I, Győri G, Bérczi V, Maurovich-Horvat P., Hagymási K, Kaposi PN. Evaluation of Artificial Intelligence-Calculated Hepatorenal Index for Diagnosing Mild and Moderate Hepatic Steatosis in Non-Alcoholic Fatty Liver Disease. Med Lith. 2023;59(3). **(IF: 2.6, 2022)**

2. Rónaszéki AD, Budai BK, Csongrády B, **Stollmayer R**, Hagymási K, Werling K, Fodor T, Folhoffer A, Kalina I, Győri G, Maurovich-Horvat P, Kaposi PN. Tissue attenuation imaging and tissue scatter imaging for quantitative ultrasound evaluation of hepatic steatosis. Medicine (Baltimore). 2022;101(33):e29708. **(IF: 1.6, 2022)**

3. Rónaszéki AD, Dudás I, Zsély B, Budai BK, **Stollmayer R**, Hahn O, Csongrády B, Park B, Maurovich-Horvat P, Győri G, Kaposi PN. Microvascular flow imaging to differentiate focal hepatic lesions: the spoke-wheel pattern as a specific sign of focal nodular hyperplasia. Ultrasonography (Seoul, Korea). 2023;42(1):172-181. **(IF:2.4, 2022)**

4. Fazekas S, Budai BK, **Stollmayer R**, Kaposi PN, Bérczi V. Artificial intelligence and neural networks in radiology – Basics that all radiology residents should know. IMAGING. 2022;14(2):73-81. **(IF: 0.4, 2022)**

5. Budai BK, **Stollmayer R**, Rónaszéki AD, Körmendy B, Zsombor Z, Palotás L, Fejér B, Szendrői A, Székely E., Maurovich-Horvat P, Kaposi PN. Radiomics analysis of contrast-enhanced CT scans can distinguish between clear cell and non-clear cell renal cell carcinoma in different imaging protocols. Front Med. 2022;9:974485. **(IF: 3.9, 2022)**

6. Kaposi PN, Zsombor Z, Rónaszéki AD, Budai BK, Csongrády B, **Stollmayer R**, Kalina I, Győri G, Bérczi V, Werling K, Maurovich-Horvat P, Folhoffer A, Hagymási K. The Calculation and Evaluation of an Ultrasound-Estimated Fat Fraction in Non-Alcoholic Fatty Liver Disease and Metabolic-Associated Fatty Liver Disease. Diagnostics. 2023; 13(21):3353. **(IF: 3.6, 2022)**

7. Zsombor Z, Zsély B, Rónaszéki AD, **Stollmayer R**, Budai BK, Palotás L, Bérczi V, Kalina I, Maurovich Horvat P, Kaposi PN. Comparison of Vendor-Independent Software Tools for Liver Proton Density Fat Fraction Estimation at 1.5 T. Diagnostics. 2024; 14(11):1138. **(IF: 3.0, 2023)**

Cumulative impact factor of the candidate's publications related to the thesis: **11.374**

Total cumulative impact factor of the candidate's publications: **28.074**

DOI:10.14753/SE.2025.3142

10. Acknowledgements

I would like to thank all the members of our research group and my family for their unending support, without which the current PhD thesis could not have been born.

First and foremost, I would like to thank my supervisor, Kaposi Novák Pál, who taught and guided me through my adventure into medical imaging, deep learning, and programming. As an advisor and clinical researcher, he set an exceptional example for all of us, with his quick thinking, thorough analytical skills, and deep understanding of this complex and difficult interdisciplinary research field.

I am grateful to Professor Pál Maurovich-Horvat and Professor Viktor Bérczi for their constant support and encouragement.

I would like to express my sincere appreciation to all those collaborators, who supported my work with their expert guidance, especially Budai Bettina Katalin and Aladár Rónaszéki, to whom I could always turn to with my problems and headaches, no matter how small or insignificant they were.

I would like to thank all our TDK students who always brought an exciting, international atmosphere to our environment.

Lastly, I would like to thank my significant other, without whom I could not have overcome the difficulties of this, to me, very difficult period.