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

Ph.D. 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

1. Introduction

Magnetic resonance imaging (MRI) with hepatocyte-specific contrast agents (HSCs) has significantly enhanced the non-invasive diagnosis of focal liver lesions (FLLs) due to its superior soft tissue contrast and safety profile. Despite these advancements, the interpretation of MRI scans remains complex and time-consuming, often requiring the expertise of seasoned radiologists. This thesis explores the application of deep learning (DL) methods, particularly convolutional neural networks (CNNs), to improve the diagnostic accuracy and efficiency of MRI in identifying and characterizing FLLs.

Artificial intelligence (AI) and DL have transformed various fields, including medical imaging. CNNs, a type of DL model, have shown exceptional performance in image classification tasks, making them ideal for analyzing MRI scans. By training these models on a large dataset of annotated MRI images, we aim to develop tools that can automatically differentiate between types of FLLs and identify specific radiological features.

2. Objectives

The primary aim of this thesis is to demonstrate the feasibility of DL-based analysis of FLLs using HSC-enhanced MRI. The specific objectives include:

1. To categorize FLLs into diagnostic classes approximating the opinion of a board-certified radiologist.
2. To evaluate the sufficiency of a single axial MRI slice versus three-dimensional MRI images for accurate FLL categorization.

3. To quantify the accuracy of DL models in identifying radiological features of FLLs and compare their performance with that of radiologists.

3. Methods

3.1. Differentiation of tumor types

A retrospective single-center study was conducted, collecting multi-phasic MRI studies of patients with FNH, HCC, or MET. These studies were performed using gadoxetate disodium, an HSC, and were sourced from the institutional picture archiving and communication system. The need for written patient consent was waived by the Institutional Research Ethics Committee due to the retrospective nature of the study. The MRI scans included T2-weighted (T2w), hepatic arterial phase (HAP), portal venous phase (PVP), and hepatobiliary phase (HBP) images. Lesions included in the study were either histologically confirmed or exhibited typical radiological characteristics of the given lesion type based on the opinion of an abdominal radiologist. Patients younger than 18 years at the time of imaging were excluded from the study. The demographic data of patients, imaging properties per lesion class, and details of metastatic lesion origins were analyzed (Table 1).

MRI scans were anonymized by removing personal identifiers such as social security numbers, birth dates, and sex. The scans were exported as DICOM files and then processed using 3D Slicer, an open-source medical image computing software. The images were resampled and spatially aligned using BSpline as a non-rigid registration method. Lesions were annotated with cubic regions of interest (ROIs) and cropped from the aligned MRI volumes. The cropped volumes were converted to NIfTI format and combined into four-dimensional (4D) input data for each lesion. These datasets were then randomly sorted into training, validation, and test datasets, with specific data augmentation techniques applied to enhance the

training samples. The training and validation datasets contained three axial slices per lesion, while the test dataset contained one slice per lesion.

Concatenated files underwent various transformations during the training process, including scaling image pixel intensity and applying data augmentation techniques such as random rotation and zooming. The images were resized to a 64 × 64 resolution and converted to tensors for model input. DenseNet264, a 2D convolutional neural network, was trained using these tensors. For the 3D DenseNet264 network, NIfTI voxels were resampled to an isovolumetric shape and rescaled, then resized to a 64 × 64 × 64 spatial resolution. These were used as multi-channel inputs for the 3D CNN. Data augmentation techniques and transformations were applied to the training samples to improve generalization. Both networks were trained for 70 epochs, and the highest-performing model weights were selected based on the area under the receiver operating characteristic curve (AUC).

3.2. Identification of radiological features

The second retrospective study included 99 patients who underwent abdominal MRI with gadoxetate disodium between September 2017 and August 2021. The Institutional Research Ethics Committee waived the need for written consent for the retrospective analysis. The final study cohort included 131 scans from 99 patients, diagnosed with 105 FNHs, 121 HCCs, 121 METs, and 32 other lesions. The MRI scans were anonymized, resampled, and coregistered to the HBP scan, with misalignments manually corrected when necessary. Lesions were marked and evaluated by an experienced radiologist and a radiology resident. The dataset was split into training, validation, and test sets for DL model training.

MRI scans were acquired using a Philips Ingenia 1.5 T scanner and 5–20 mL of intravenous gadoxetate disodium

contrast. The scans included T2-weighted (T2w), native T1-weighted, arterial (HAP), portal venous (PVP), equilibrium phase (VEN), and hepatobiliary phase (HBP) images. Each scan was anonymized and converted to 3D NIfTI format. The HBP scans were resampled to isotropic 1 × 1 × 1 mm voxel spacing, and all other scans were coregistered to the corresponding HBP scan. Lesions were cropped based on their largest diameter, with a 2-mm-wide zone added in each direction to account for misalignments. Lesion markers and manual corrections were performed using 3D Slicer. The datasets were then split into training, validation, and test sets for model training.

Multiple DL algorithms were trained using MONAI, an open-source framework for deep learning in healthcare imaging. DenseNet121, DenseNet169, DenseNet201, DenseNet264, and EfficientNet models were trained with various hyperparameter setups, including batch sizes, dropout rates, and weight decay. The models were trained for at least 300 epochs, and the best-performing model was defined as the one achieving the highest mean AUC on the validation dataset. Each model was modified to perform multi-label classification, with binary cross-entropy (BCE) loss used for weight adjustment during training. Occlusion sensitivity maps were generated to visualize the importance of different image regions in the decision-making process.

ROC analysis was performed on the test dataset to calculate AUC values and set cut-off thresholds for each feature. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and f1 scores were calculated for each feature. Inter-rater reliability was assessed using Cohen's Kappa, comparing the model predictions with expert annotations.

4. Results

4.1 Differentiation of tumor types

The 2D DenseNet model achieved the highest average AUC for FNH (0.9900), HCC (0.9600), and MET (0.9950) after 46 epochs. The 3D DenseNet model achieved an average AUC of 0.9433, demonstrating slightly lower performance compared to the 2D model. The evaluation metrics for both models showed high diagnostic accuracy, with the 2D model achieving better sensitivity and specificity across all tumor types. Attention maps generated from the models highlighted the most important regions for diagnosis, correlating well with radiological features.

4.1 Identification of radiological features

The EfficientNetB0 model achieved the highest validation mean AUC (0.9147) for radiological feature identification. The model showed excellent performance in detecting features such as delayed phase enhancement (AUC 0.99) and iso- or hyperintensity on the venous phase (AUC 0.98). Features like T2 hyperintensity and washout were less accurately predicted, with AUC values of 0.79 and 0.64, respectively. The model demonstrated high specificity and NPV for most features, indicating reliable exclusion of non-existent features.

5. Conclusions

DL models, particularly CNNs, show significant promise in enhancing MRI-based diagnosis of FLLs. Both 2D and 3D DenseNets are effective in differentiating between FLL types using HSC-enhanced MRI. The automatic identification of radiological features by DL models can aid clinicians in diagnosing liver conditions with high accuracy, potentially improving patient outcomes through more precise and timely diagnoses. However, further studies involving larger, multi-

institutional datasets and additional lesion classes are necessary to validate these findings and refine the models.

6. Bibliography of the candidate's publications

Publications related to the 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)**

Publications not related to the 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)**

∑IF: 28.074