Aus dem CharitéCentrum für Innere Medizin
mit Gastroenterologie und Nephrologie CC 13

Medizinische Klinik mit Schwerpunkt Hepatologie und Gastroenterelogie
Campus Virchow-Klinikum

DISSERTATION

AI-basierte volumetrische Analyse der Lebermetastasenlast bei Patienten mit neuroendokrinen Neoplasmen (NEN)

AI-based Volumetric Analysis of Liver Metastases Load in Neuroendocrine Neoplasm (NEN) Patients

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Siyi Xin

aus Xiangyang, Hubei, China

Datum der Promotion: 25.11.2022

Preface

Teilergebnisse der vorliegenden Arbeit wurden veröffentlicht in:

Fehrenbach U, Xin S, Hartenstein A, Auer TA, Dräger F, Froböse K, Jann H, Mogl M, Amthauer H, Geisel D, Denecke T, Wiedenmann B, Penzkofer T. Automatized Hepatic Tumor Volume Analysis of Neuroendocrine Liver Metastases by Gd-EOB MRI-A Deep-Learning Model to Support Multidisciplinary Cancer Conference Decision-Making. Cancers (Basel) / 2021 May 31.

Table of Contents

Abbreviations

Abbreviations and terms	Description
Δ_{abs}HTL	Absolute HTL change
Δ_{abs}NELM	Absolute NELM volume change
Δ_{rel}HTL	Relative HTL change
Δ_{rel}NELM	Relative NELM volume change
1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
5-HIAA	5-Hydroxyindoleacetic acid
^{68}Ga-SSAs-PET/CT	^{68}Ga-labeled somatostatin analogs-PET/CT
AAIR	Age-adjusted incidence rate
ACTH	Adrenocorticotropic hormone
AI	Artificial intelligence
BL	Baseline
CgA	Chromogranin A
CNNs	Convolutional neural networks
CR	Complete remission
CT	Computed tomography
DC	Dice coefficient
DWI	Diffusion-weighted imaging
ENETS	European Neuroendocrine Tumor Society
EUS	Endoscopic ultrasound
FN	False negatives
FP	False positives
FU	Follow-up
Gd-EOB	Gadoxetic acid
GEP	Gastroenteropancreatic tract
HBP	Hepatobiliary phase
HTL	Hepatic tumor load
IQR	Interquartile ranges

LMs	Liver metastases
MANECs	Mixed adenoneuroendocrine carcinomas
MCC	Multidisciplinary cancer conferences
MiNENs	Mixed neuroendocrine–non-neuroendocrine neoplasms
MITK	Medical Imaging Interactive Toolkit
MRI	Magnetic resonance imaging
NEC	Neuroendocrine carcinoma
NELMs	Neuroendocrine liver metastases
NENs	Neuroendocrine neoplasms
NET	Neuroendocrine tumor
nnU-Net	No-new-Net
OS	Overall Survival
PD	Progress
PET	Positron emission tomography
pNENs	Pancreatic neuroendocrine neoplasms
PR	Partial remission
RECIST	Response Evaluation Criteria in Solid Tumors
ROI	Region of interest
SD	Stability
SRS	Somatostatin receptor scintigraphy
SSTRs	Somatostatin receptors
TN	True negatives
TP	True positives
US	Ultrasound
VIP	Vasoactive intestinal polypeptide
WHO	World Health Organization

List of Tables

List of Figures

Abstract (English)

Background: Quantification of liver tumor load in patients with liver metastases from neuroendocrine neoplasms is essential for therapeutic management. However, accurate measurement of three-dimensional (3D) volumes is time-consuming and difficult to achieve. Even though the common criteria for assessing treatment response have simplified the measurement of liver metastases, the workload of following up patients with neuroendocrine liver metastases (NELMs) remains heavy for radiologists due to their increased morbidity and prolonged survival. Among the many imaging methods, gadoxetic acid (Gd-EOB)-enhanced magnetic resonance imaging (MRI) has shown the highest accuracy.

Methods: 3D-volumetric segmentation of NELM and livers were manually performed in 278 Gd-EOB MRI scans from 118 patients. Eighty percent (222 scans) of them were randomly divided into training datasets and the other 20% (56 scans) were internal validation datasets. An additional 33 patients from a different time period, who underwent Gd-EOB MRI at both baseline and 12-month follow-up examinations, were collected for external and clinical validation (n = 66). Model measurement results (NELM volume; hepatic tumor load (HTL)) and the respective absolute (Δ_{abs}NELM; Δ_{abs}HTL) and relative changes (Δ_{rel}NELM; Δ_{rel}HTL) for baseline and follow-up-imaging were used and correlated with multidisciplinary cancer conferences (MCC) decisions (treatment success/failure). Three readers manually segmented MRI images of each slice, blinded to clinical data and independently. All images were reviewed by another senior radiologist.

Results: The model's performance showed high accuracy between NELM and liver in both internal and external validation (Matthew's correlation coefficient (ϕ): 0.76/0.95, 0.80/0.96, respectively). And in internal validation dataset, the group with higher NELM volume (> 16.17 cm^3) showed higher ϕ than the group with lower NELM volume (ϕ = 0.80 vs. 0.71; p = 0.0025). In the external validation dataset, all response variables (Δ_{abs}NELM; Δ_{abs}HTL; Δ_{rel}NELM; Δ_{rel}HTL) reflected significant differences across MCC decision groups (all p < 0.001). The AI model correctly detected the response trend based on Δ_{rel}NELM and Δ_{rel}HTL in all the 33 MCC patients and showed the optimal discrimination between treatment success and failure at +56.88% and +57.73%, respectively (AUC: 1.000; P < 0.001).

Conclusions: The created AI-based segmentation model performed well in the three-dimensional quantification of NELMs and HTL in Gd-EOB-MRI. Moreover, the model showed good agreement with the evaluation of treatment response of the MCC's decision.

Abstrakt (Deutsch)

Hintergrund: Die Quantifizierung der Lebertumorlast bei Patienten mit Lebermetastasen von neuroendokrinen Neoplasien ist für die Behandlung unerlässlich. Eine genaue Messung des dreidimensionalen (3D) Volumens ist jedoch zeitaufwändig und schwer zu erreichen. Obwohl standardisierte Kriterien für die Beurteilung des Ansprechens auf die Behandlung die Messung von Lebermetastasen vereinfacht haben, bleibt die Arbeitsbelastung für Radiologen bei der Nachbeobachtung von Patienten mit neuroendokrinen Lebermetastasen (NELMs) aufgrund der höheren Fallzahlen durch erhöhte Morbidität und verlängerter Überlebenszeit hoch. Unter den zahlreichen bildgebenden Verfahren hat die Gadoxetsäure (Gd-EOB)-verstärkte Magnetresonanztomographie (MRT) die höchste Genauigkeit gezeigt.

Methoden: Manuelle 3D-Segmentierungen von NELM und Lebern wurden in 278 Gd-EOB-MRT-Scans von 118 Patienten durchgeführt. 80% (222 Scans) davon wurden nach dem Zufallsprinzip in den Trainingsdatensatz eingeteilt, die übrigen 20% (56 Scans) waren interne Validierungsdatensätze. Zur externen und klinischen Validierung (n = 66) wurden weitere 33 Patienten aus einer späteren Zeitspanne des Multidisziplinäre Krebskonferenzen (MCC) erfasst, welche sich sowohl bei der Erstuntersuchung als auch bei der Nachuntersuchung nach 12 Monaten einer Gd-EOB-MRT unterzogen hatten. Die Messergebnisse des Modells (NELM-Volumen; hepatische Tumorlast (HTL)) mit den entsprechenden absoluten (Δ_{abs}NELM; Δ_{abs}HTL) und relativen Veränderungen (Δ_{rel}NELM; Δ_{rel}HTL) bei der Erstuntersuchung und der Nachuntersuchung wurden zum Vergleich mit MCC-Entscheidungen (Behandlungserfolg/-versagen) herangezogen. Drei Leser segmentierten die MRT-Bilder jeder Schicht manuell, geblindet und unabhängig. Alle Bilder wurden von einem weiteren Radiologen überprüft.

Ergebnisse: Die Leistung des Modells zeigte sowohl bei der internen als auch bei der externen Validierung eine hohe Genauigkeit zwischen NELM und Leber (Matthew's Korrelationskoeffizient (ϕ): 0,76/0,95 bzw. 0,80/0,96). Und im internen Validierungsdatensatz zeigte die Gruppe mit höherem NELM-Volumen (> 16,17 cm3) einen höheren ϕ als die Gruppe mit geringerem NELM-Volumen (ϕ = 0,80 vs. 0,71; p = 0,0025). Im externen Validierungsdatensatz wiesen alle Antwortvariablen (Δ_{abs}NELM; Δ_{abs}HTL; Δ_{rel}NELM; Δ_{rel}HTL)

signifikante Unterschiede zwischen den MCC-Entscheidungsgruppen auf (alle $p < 0,001$). Das KI-Modell erkannte das Therapieansprechen auf der Grundlage von Δ_{rel}NELM und Δ_{rel}HTL bei allen 33 MCC-Patienten korrekt und zeigte bei +56,88% bzw. +57,73% eine optimale Unterscheidung zwischen Behandlungserfolg und -versagen (AUC: 1,000; $P < 0,001$).

Schlussfolgerungen: Das Modell zeigte eine hohe Genauigkeit bei der dreidimensionalen Quantifizierung des NELMs-Volumens und der HTL in der Gd-EOB-MRT. Darüber hinaus zeigte das Modell eine gute Übereinstimmung bei der Bewertung des Ansprechens auf die Behandlung mit der Entscheidung des Tumorboards.

1. Introduction

1.1 Neuroendocrine Neoplasms (NENs)

Neuroendocrine neoplasms (NENs) are the general term for a series of malignant tumors derived from neuroendocrine cells, with specific morphological and neuroendocrine characteristics (1). “Neuroendocrine”, as the name, represents both properties of endocrine and neuronal cells (2), therefore, NENs are observed not only in pure endocrine organs or nerve structures but also in organs with a diffuse neuroendocrine cell system (lung, bronchi, reproductive system, and digestive system) (3). However, they originate most often in the gastroenteropancreatic tract (GEP) and have become the second most common digestive cancer (4).

Compared to other malignant solid tumors, NENs generally are considered indolent tumors with benign characteristics and tend to follow a prolonged clinical course even with distant metastases at first diagnosis (4,5), which account for 40-50% of NEN patients (6).

1.1.1 Epidemiology

NENs are rare, but in the last 3 decades, the incidence of NEN has steadily increased worldwide (4,7). Studies suggested that the age-adjusted incidence rate (AAIR, per 100.000 per year) of NENs increased 6.4-fold between 1973 and 2012 in the US (8), and increased 2.36-fold (from 2.48 to 5.86) from 1994 to 2009 in Canada (9). Whereas, the AAIR of gastrointestinal NENs increased 23-fold in men and 47-fold in women from 1995 to 2006 in the UK (10).

In the European database: In Germany, the AAIR of GEP-NENs increased significantly between 1976 and 2006 from 0.44 to 2.33, resulting in a 5.28-fold increase (11); whereas the AAIRs of NENs increased from 13.3 in 1993 to 21.3 in 2010 in Norway (12), from 2.1 in 1990 to 4.9 in 2010 in the Netherlands (13), and also increased both in male and female cohorts in Sweden (14,15).

Meanwhile, there is the same trend in Asia: There is an approximately 1.2-fold increase in the number of patients of pancreatic NENs (pNENs) in Japan from 2005 to 2010 (16). In India,

compared to the diagnosed period from 2001 to 2005, the number of GEP-NEN patients who were diagnosed from 2011 to 2016 has increased more than 6.3-fold (17).

Moreover, NENs, especially those of the intestine and pancreas, commonly present with metastases at initial diagnosis (5), which are most frequently located in the liver (6). “In the largest US epidemiological database (SEER), the analysis of all cases with available information, the metastatic status was 49% in localized, 24% in regional metastases, and 27% in distant metastases” (5), while the percentages of the study in Norway were 40.4%, 17.5%, and 42.1% respectively (12). However, “the five-year survival ranges from 60% to 100% for the localized disease to 25% for the metastatic disease for pNENs” (18).

1.1.2 Prognostic factors

The prognosis of patients with neuroendocrine tumors is affected by many factors. Among those, the presence of liver metastases is the factor with the worst prognosis for NENs patients (19). Moreover, the clinical occurrence, development of liver metastases, and 5-year survival rates of neuroendocrine tumor liver metastases (NELM) patients have been impacted by the primary sites, differentiation, grading, tumor load, functionalities, and treatment strategies (5,20). However, the functionalities and treatment strategies depend to some extent on other indicators. Here we mainly discuss the indicators other than treatment.

1.1.2.1 Histologic differentiation and Grading

The gold standard for the diagnosis of NEN is the histopathological analysis from biopsy samples (21), and the differentiation and grading have been suggested as the strongest predictors of survival by ENETS guidelines (5).

“Well”- and “poorly”-differentiated are used to describe the difference between NEN cells and normal neuroendocrine cells. As long as the morphological features of NEN cells are highly similar to normal neuroendocrine cells and show a low degree of atypical cells, NEN is defined as well-differentiated and is usually called neuroendocrine tumor (NET) of the digestive system (3). Conversely, NEN cells without the typical morphology of neuroendocrine cells (e.g. severe cell atypia), are labeled as poorly differentiated and are defined as neuroendocrine carcinoma (NEC) (3).

The nucleic cell cycle-dependent marker Ki67 (MIB1) antigen is used to determine the proliferative activity of NENs. Counting the mitoses per high-power field can also be used to define the grading, but not as reliable as Ki67 (21).

The classification based on histologic differentiation and grading (proliferative activity) from the World Health Organization (WHO) provides a useful standard for clinicians. Although organ-specific classifications have been historically developed and are in use, general classification principles for all NEN types were defined in 2019 (22,23). The current classification framework for GEP-NENs is provided in Table 1 (2,23). In this classification, NETs are further graded as grade 1 to grade 3, while NECs are always considered high grade, only subclassified into small- and large-cell (23). Moreover, NEC patients have more frequent distant metastases compared to NET patients (5), and the higher the grade the worse the patient's outcome (2,8).

Mixed neuroendocrine–non-neuroendocrine neoplasms (MiNENs), which were previously called mixed adenoneuroendocrine carcinomas (MANECs), combine a neuroendocrine component and a non-neuroendocrine component (generally an adenocarcinoma), and both of them account for at least 30% of the neoplasm (24). The prognosis of MiNEN lies in between that of "pure" NEN and that of "pure" adenocarcinoma of the same origin (24).

Table 1. 2019 WHO Classification for Neuroendocrine neoplasms (NENs) of the digestive system

Terminology	Histologic differentiation	Grade	Mitotic rate (mitoses/2 mm^2) [a]	Ki-67 (%) [b]
NET G1	Well differentiated	Low	<2	≤2
NET G2		Intermediate	2-20	3-20
NET G3		High	>20	>20
NEC G3, small-cell type (SCNEC)	Poorly differentiated	High	>20	>20
NEC G3, large-cell type (SCNEC)				
MiNEN	Well or poorly differentiated (both components) [c]	NA	NA	NA

WHO, World Health Organization; NET, neuroendocrine tumor; NEC, neuroendocrine carcinoma; MiNEN, Mixed neuroendocrine-non-neuroendocrine neoplasm; NA, not available.

[a] at least 50 fields are evaluated at ×40 magnification in the area of highest mitotic density.

[b] Percentage of at least 500 tumor cells in the area of highest Ki-67 nuclear labeling (hot spot).

[c] MiNEN tumors are not pure neuroendocrine neoplasms; grading is performed separately for the NEN component and the non-neuroendocrine counterpart by WHO 2019; The overall grading of NENs is based on the highest proliferation index.

This table was created according to the original table in (23).

1.1.2.2 Primary sites

The primary site is considered as another important predictor of survival (8,25) since different primary sites have different tendencies in differentiation, grading, metastatic disease. Most midgut NENs are well-differentiated and grow slowly (20), whereas, pancreas and small intestine are more often, even at initial diagnosis, combined with liver metastases (5,6). Moreover, the liver metastases from NENs originating in the different primary sites represent a significant and frequent clinical occurrence, which negatively impacts prognosis (19).

A study of 64971 patients with NENs from 1973 to 2021 in the US showed that the median overall survival (OS) of NENs ranged from 4 months to 30 years from different sites, with the best OS at sites of the small intestine and appendix; followed by the rectum, cecum, and pancreas; colon and liver had the worst prognosis (8). Another study from Spain of 2813 patients showed that the prognosis was good for the mid-gut-NENs (including appendix), intermediate for gastric or pancreatic NENs, and poor for colon, rectum, hepatobiliary, or esophageal and those of unknown primary (25). Whereas, a study from Norway showed that the overall 5-year survival rates of NENs from rectal, appendiceal, and small intestine were higher than 60%, while from the stomach, colon, and pancreas were lower than 60%, among which pancreatic NENs (pNENs) were the worst (26).

NENs from different primary tumor locations present also with different hormonal symptoms and syndromes. Depending on their hormonal hypersecretion, they may present for example with a carcinoid syndrome, mostly seen in the case of a small intestinal NEN, which is often characterized by cardiac and gastrointestinal symptoms (diarrhea, flushing); whereas pNENs are

in the majority of cases hormonally silent (27) and therefore called nonfunctional in contrast to above given functional NEN. Functional pNEN present 10-40% of all pNENs and can produce insulin, gastrin, vasoactive intestinal polypeptide (VIP), glucagon, somatostatin, adrenocorticotropic hormone (ACTH), serotonin, and so forth, giving rise to the respective clinical syndromes, respectively, such as Whipple triad syndrome (Insulinomas), Zollinger-Ellison syndrome (Gastrinomas), Verner-Morrison syndrome (VIPomas), hyperglycemia (Glucagonomas), steatorrhea, achlorhydria, diabetes mellitus, and cholelithiasis (Somatostatinomas), ACTHomas, and carcinoid syndrome, etc (4,28). Additionally, functional pNENs have a better long-term prognosis compared to nonfunctional pNENs (27).

Moreover, the primary sites are associated with histology and grading, which were also affecting the prognosis (8,26).

1.1.2.3 Biochemical Markers

Biochemical markers have important roles in the diagnosis and follow-up of NEN patients, and several new biomarkers, e.g. circulating DNA, mRNA, or circulating tumor cells are investigated but still need more clinical trials to be evaluated (29). Those which have been used in the clinic for many years can be roughly divided into two categories: hormones or amines secreted by NEN cells have been used in functional NENs, while the “general markers” have been particularly used in non-functional NENs (29,30). However, their sensitivity, specificity, and possible effects of the co-existent disease(s) and/or drugs limit their accuracy in diagnosis, but they are considered as a method of auxiliary diagnosis and follow-up (31). The three most commonly used biochemical markers for the gut and pancreas are listed below.

Chromogranin A (CgA), an acidic protein widely existing in neuroendocrine cells, is the most important general biochemical marker by far and can be used for prognosis and follow-up since its concentrations generally correlate with the tumor cell type, histological differentiation, and tumor burden of NENs (32–35). The sensitivity and specificity are generally between 60 and 90% (29,32,36) and may increase to 100% in metastatic NENs (37–39) or in combination with other diagnostic methods (40–42). Of note, it was also reported that there were some common

conditions – decreased renal function, treatment with PPIs, or treatment for hypertension – can slightly increase the levels of CgA and lead to false-positives (31,43).

Urinary 5-hydroxyindoleacetic acid (5-HIAA), the main metabolite of serotonin, is mainly produced by midgut NENs and causes diarrhea and flushing in the form of carcinoid syndrome. Therefore, the 24-hour urinary 5-HIAA collection is recommended to monitor NENs with carcinoid -syndrome (31,32) rather than serum serotonin which showed unreliable for diagnosis and in follow-up because of large individual variation (30). Although 5-HIAA has a high specificity (90–100%) in the carcinoid syndrome, the reported low sensitivity (35–68%) limits its clinical use. Moreover, its level is not directly related to the severity of symptoms but is associated with an increased incidence and severity of carcinoid heart disease (31,32,44,45). Additionally, the level of 5-HIAA interferes also by some normal foods, drinks, and drugs thereby leading to false-positive findings (31).

Insulinomas, which are the largest fractions in well-differentiated pNETs, secreted increasing insulin and cause hypoglycemia. The 72-hour fast test is the current gold standard for the diagnosis of insulinoma (30) with near 100% sensitivity and specificity by several studies (46–48). However, endogenous hyperinsulinism and other causes of hypoglycemia, such as exogenous insulin, oral hypoglycemic drugs, or autoimmunity, should also be differentiated (32).

1.1.2.4 Hepatic tumor load (HTL)

“The presentation of metastatic disease represents the most important prognostic factor after tumor grading” (5). As the most common metastasis organ, the tumor burden or the number of metastases in the liver represent additional prognostic parameters (5), which are essential in the intervention of liver metastases (49).

Hepatic tumor load (HTL), which is usually expressed as the ratio of liver tumor volume to total liver volume, is used to describe the extent of liver involvement with metastases (50). The European Neuroendocrine Tumor Society (ENETS) suggested HTL as an important negative prognostic factor and crucial information regarding tumor resectability in patients with neuroendocrine liver metastases (NELMs) for follow-up (5,6). “Macroscopically, depending on the therapeutic approach, three different patterns of liver metastases have to be differentiated”

(51). (A) “Simple pattern, whereby liver metastases occur only in one liver lobe or are confined to two neighboring liver segments. They can be found in 20–25% of the cases and can be removed by a standard anatomical resection. (B) Complex pattern, liver metastases occur predominantly in one lobe and with smaller satellites in the other lobe (10–15%). They can still be handled surgically, including ablative approaches. (C) Diffuse pattern, multifocal liver metastases are present in around 2/3 of the cases and should not be treated surgically” (5) (**Figure 1**). However, to date, no standardized quantitative method exists for fast and reliable measurement of HTL of all three metastatic types, so far (50).

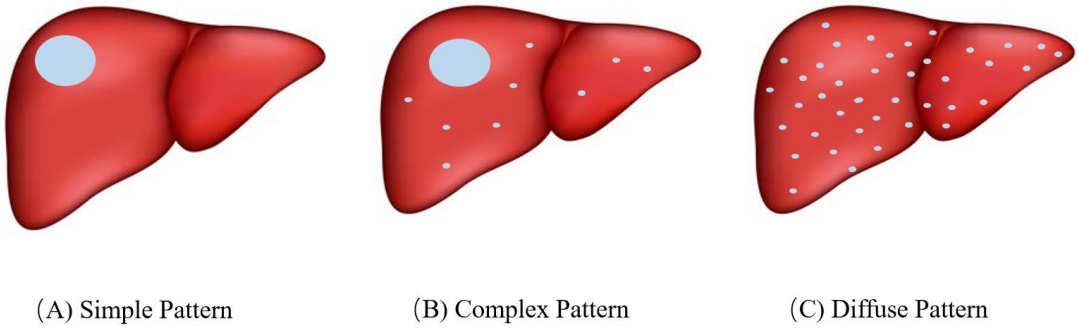


Figure 1. Three types of NEN-liver metastases

1.1.3 Radiological Diagnosis of NELMs

As the most common method of tracking treatment response and prompting key information about the tumor itself, the need for radiological diagnosis is becoming stronger, especially in the follow-up of NELM patients. The choice of imaging methods varies depending on the purpose and the patient’s tumor status (52).

1.1.3.1 MRI

Magnetic resonance imaging (MRI) plays an important role in the diagnosis of focal and diffuse liver disease and has become a cornerstone in evaluating neuroendocrine liver metastases (NELMs) from diagnosis to follow-up (53,54). The high tissue contrast makes it the most sensitive technique for detecting NELMs in contrast to, ultrasound (US), computed tomography (CT), or somatostatin receptor scintigraphy (SRS), and lead to its use in first-line imaging (49,55,56). However, the sole application of non-contrast MRI often leads to missing small metastatic lesions. The use of MRI contrast agents strongly improves the detection of small hepatic metastases detection due to the higher tissue contrast of healthy vs. metastatic tissue (57).

Gadoxetic acid-enhanced MRI (Gd-EOB MRI), one of the hepatocyte-specific contrast agents, is recommended for detecting hepatic lesions based on its biodistribution characteristics among all the contrast agents (58). Gd-EOB can only be taken up by normal hepatocytes, thus a difference in intensity between normal and cancerous liver tissue can be observed (59,60). Although only a few studies on its role in detecting NELMs exist, all studies demonstrated excellent performance due to the additionally available hepatobiliary phase (HBP) (53,58,61–63). In patients with normal liver function, the absorption of Gd-EOB in the liver parenchyma reaches its peak 20 minutes after injection, corresponding to the HBP, also visualizing the biliary system at this phase (61,64–71). The pattern of NELMs in HBP is not different from other malignant hepatic tumors, with a strong contrast between the hyperintense liver and non-enhancing metastases (60,72–74). However, several studies showed that HBP not only aids in better lesion detection compared to other MRI sequences in NENs (58,61,68,75–77) but also enables a precise lesion segmentation due to a clear delineation between tumor and liver (61). Therefore, Gd-EOB MRI is considered routinely for diagnosis and follow-up of NELMs patients (56,59,78).

On the other hand, diffusion-weighted imaging (DWI) has been also suggested as a good sequence for detecting NELMs without the disadvantages of intravenous contrast agents (79,80). In a recent study, the combination of DWI and HBP has been suggested as the most sensitive imaging method for NELMs (58).

1.1.3.2 CT

CT has been the standard technique and plays a key role in the diagnosis of NELMs, with the mean sensitivity of 84% and the mean specificity of 92% (49). Besides, CT is the preferred test when chest lesions are considered (49). It can reflect the location of the tumor, invasion of the area, surrounding tissues and organs, and plays an important role in determining the surgical plan and staging (81).

1.1.3.3 Scintigraphy, SPECT and PET(/CT)

Positron emission tomography (PET) is a highly sensitive method for nuclear medicine diagnosis. Positron emission tomography-computed tomography (PET-CT) performs PET and CT almost simultaneously and can fuse the information of these two complementary technologies: CT provides accurate anatomic information, while PET maps normal and abnormal tissue functions. PET-CT can be more accurate and faster to identify the type and exact location of NENs (82). In

general, PET-CT is used in NEN patients with unknown primary tumor sites or for staging to examine systemic spread.

Neuroendocrine cells express somatostatin receptors (SSTRs) on their cell surface, especially in GEP-NETs with a high incidence and density, which make them turn into unique and specific molecular targets for nuclear imaging (49,83,84). Octreotide is a synthetic somatostatin-like molecule with a much longer half-life than that of somatostatin, therefore, Octreoscan™ (^{111}In-pentetreotide) has been widely used as the first nuclide label in somatostatin receptors scintigraphy (SRS) for NENs since the 1980s (83,85). In recent years, ^{68}Ga-labeled somatostatin analogs-PET/CT (^{68}Ga-SSAs-PET/CT, i.e. ^{68}Ga-DOTA-TOC/TATE/NOC-PET/CT) have gradually been used instead of ^{111}In-labeled SRS and become the preferred modality for SSTR imaging due to their excellent biodistribution, pharmacokinetics, high sensitivity, low side effects, and better imaging quality (4,86–88). SRS still plays an important role in the search and localization of primary lesions, the detection of distant metastases, which helps to accurately determine the stage and biological behavior of NENs and guide the development of individualized treatment plans for NEN patients. From the guideline of ENETS, the mean sensitivities of ^{68}Ga-SSAs-PET/CT for well-differentiated NET (G1-G2) detection was between 88 and 93%, and the specificities ranged between 88 and 95% (49). Notably, taking into account the physiological uptake during image analysis can avoid false positives. Before SSTR imaging, long-acting somatostatin should be paused for 3 to 4 weeks, and short-acting somatostatin analogs should be used for 24 to 48 hours (83).

For high-grade NENs (normally G3), ^{18}Fluorine-labeled fluorodeoxyglucose (^{18}F-FDG) is the most commonly used positron-nuclide-labeled drug in clinical practice (49,89), since it reflects the glucose metabolism of tumors (90), which has been shown to be elevated in high-grade NENs (91). The high uptake of ^{18}F-FDG often indicates a poor prognosis so that ^{18}F-FDG positive NET G2 correlates to poor biological outcomes. However, the sensitivity of ^{18}F-FDG-PET/CT for NET detection has been reported with the ranges from 37 to 72% (49).

The demonstration of sufficient somatostatin receptor expression is a prerequisite of peptide receptor radionuclide therapy (PRRT) (49), which is considered an important treatment option.

Thereby, SRS positivity can be used to stratify the patients for eligibility for PRRT or monitor the therapeutic effect after PRRT.

1.1.3.4 Ultrasound, Endoscopy, and Endoscopic ultrasound

Ultrasound (US), known as an operator-sensitive modality, provides diagnostic information for limited parts of the abdomen, such as liver and pancreatic lesions (49,52). Although it has the advantages of being a radiation-free, real-time examination, and can guide biopsy, it is greatly restricted in the evaluation of efficacy and judgment of disease progression, which makes it only an auxiliary method in combination with other imaging techniques for the initial diagnosis of NELMs but is not suited for follow-up (92,93). Especially in obese patients, US is unreliable due to difficulties in performing the examination (52).

The widespread use of endoscopy has led to a significant increase in the detection of rectal, gastric, and duodenal NENs (11,94). The detection of small bowel NETs may be made by capsule endoscopy (52). However, endoscopy can only observe luminal lesions, which helps detect the primary site of the NENs but has no role in staging and follow-up of liver metastases.

Endoscopic ultrasound (EUS), a combination of endoscopic technology and ultrasound imaging, has the advantages of both, and is recommended as the most sensitive method for identifying pNEN (95–97). It also can be used for biopsy guidance to achieve histopathological diagnosis (98). However, EUS has the shortcomings of both US and endoscopy, making it unsuitable for the follow-up of patients with distant metastases.

1.1.4 Response Evaluation Criteria in Solid Tumors (RECIST)

The Response Evaluation Criteria in Solid Tumors (RECIST) are a tumor response assessment method, which mainly assesses the change in tumor size as the primary endpoint. It is an important feature for the clinical evaluation of the efficacy of NEN therapeutics. It divides response to treatment into four categories: complete remission (CR), partial remission (PR), stable disease (SD), and progressive disease (PD) (99). However, in the course of clinical practice, this standard also has some shortcomings: ① It is not suitable for diffuse NELMs (accounting for 60-70% (100)); ② In case of an irregular shape of the lesion measurement can be inaccurate; ③ Since only two lesions per organ are evaluated, in case of a higher number of metastases, RECIST can

be inaccurate due to mixed response. In addition, RECIST neglects HTL and its prognostic value. Therefore, RECIST in NELM patients can inaccurately lead to an over- or undertreatment of at least some NEN patients (101).

1.2 Clinical application of artificial intelligence (AI)

Recently, artificial intelligence (AI) has developed rapidly in various fields. In the field of healthcare, AI technology, including its deep learning subtype, is becoming a major component of many applications. This extends to drug discovery, remote patient monitoring, medical diagnosis and imaging, risk management, wearable devices, virtual assistants, and hospital management (102).

Radiologists usually assess medical images visually and report findings to diagnose or monitor diseases. This assessment is usually subjective. Meanwhile, with the increasing demand for imaging, radiologists are required to make more decisions in a shorter time, which inevitably increases the error rate (102). However, the judgment of treatment response relies upon the radiologists' decision. Under these circumstances, AI has been integrated into the clinical workflow as a tool to assist doctors and has shown a reasonable performance in image-based diagnoses (103–107). Moreover, AI is better at recognizing complex patterns in image data, and can automatically provide quantitative assessments, which can be more accurate and repeatable for radiological assessments (102).

Many studies have already shown that AI demonstrated excellent performance in segmentation, texture analysis, and classification in a variety of tumors such as brain tumors, prostate tumors, and breast cancer (108–112). Only a few studies have reported the application of AI in the pathology of NENs (113–117).

The main challenge in radiological practice evaluating NELM is that the traditional indicators used to estimate tumor burden and determine treatment response, including the longest tumor diameter measured by established RECIST standards, are oversimplifying. Although several semiautomatic techniques to achieve 3D lesion segmentation has been proposed (110–112), this field is still in its infancy and is only being used in small clinical trials since the accurate segmentation of lesions remains time-intensive. A general application suitable for daily clinical practice has not yet been established.

Like other image recognition approaches, the task of AI in medical imaging is to create models, which are deduced from training data, tuned to produce accurate predictions by an optimization

algorithm, to be able to deliver correct predictions for new, unseen data (118). Generally, the process can be divided into the following five steps: (1) Obtaining the target image data set and labeling information; (2) Image preprocessing; (3) Training the model on labeled data (4) Validating the model (5) Evaluation of the model on a test set. Segmentation is a basic but challenging task for such models.

For the first step, the most common form and considered the 'gold standard' in clinical practice and research is supervised learning (119). That is, the images trained by AI have corresponding classification labels. Although unsupervised learning has been widely used in other fields, to our best knowledge, no study has investigated the fully automatic segmentation of hepatic tumor because of the following reasons: (i) the anatomy of the liver is complex and (ii) there are many types of treatments for patients with liver tumors, and the manifestations of tumor lesions are variable.

Recently proposed deep learning architectures for segmentation include fully convolutional networks, U-net, and variants based on these model architectures (102).

Convolutional neural networks (CNNs), which are far more ubiquitous in image analysis now, were first used in medical applications in 1995 (120) but had their major success hand writing recognition in 1998 (121). After these promising results, the CNNs were not commonly used until 2012, in which year Krizhevsky A. et al. used CNNs to win the ImageNet challenge by a large margin (122). After that, not only did algorithmic developments lead to more efficient training, but advances were also made in terms of computing power enabling the widespread use of the technology. CNNs have now become the basis for many algorithms outperforming other techniques by a large margin.

1.2.1 U-Net:

The successful training of CNNs requires enormous, annotated training data, which is beyond reach in many biomedical tasks. To facilitate the training process, Ronneberger O. et al. (123) presented a network, called U-net due to its shape (shown in Figure 2), which has been specifically designed for the segmentation of medical images (102).

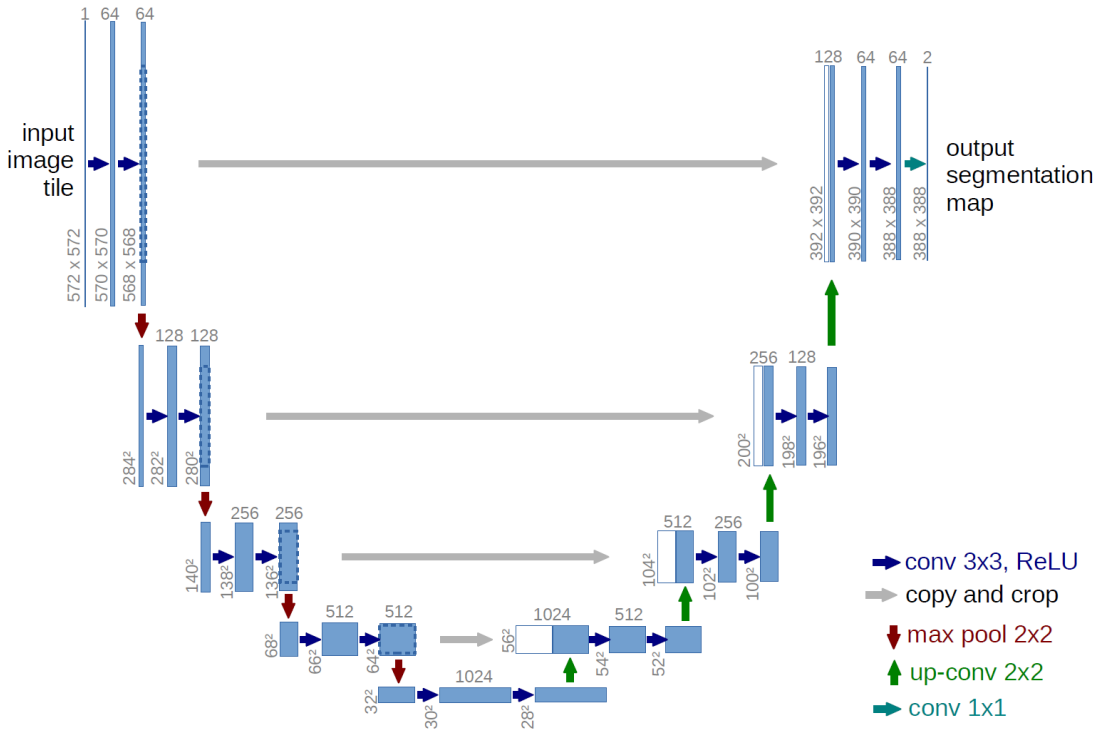


Figure 2. The structure of a U-Net (example for 32*32 pixels in the lowest resolution) (123).

The U-net architecture involves two interconnected pathways: The contracting path (left side) is used to extract image features, capture context, and compress the image into feature maps composed of features, which followed the typical architecture of a CNN; Expanding path (right side) is used for precise localization and decode the extracted feature into a segmented prediction image with the same size as the original image. Every step in the expansive path includes an up-sampling of the feature map and a concatenation cropping feature map from the contracting path, which mirrors the loss of border pixels in every convolution and gains the higher resolution feature maps (123). Remarkably, the different feature fusion method of a U-Net is very different from other common segmentation networks.

Data augmentation, especially by applying elastic deformations, is critical to train the network with few training samples but can also be used to improve training accuracy in larger networks.

It “augments” the number of available images by various distortions of the original images, such as rotation, skewing, noise filtering, or other geometric or voxel value transformations. Thereby a larger dataset, still representing the original ground truth can be created. In many applications, instead of using batches as input images, U-Net uses tiles of the images to reduce the

computational burden and to maximize usage of the GPU memory, and minimize the time of training (123).

1.2.2 nnU-Net (no new-UNet)

Although the U-Net architecture was swiftly considered the gold standard in medical image segmentation since it was presented in 2015, there was still a non-negligible issue that needed to be improved – the U-Net lacks the adaptability to novel problems – the overall performance will be affected by the choices of each step (including exact architecture, pre-processing, training, and inference) making the tuning of hyperparameters a tedious undertaking.

To save solve this issue, Isensee F. et al. (124) proposed the no-new-Net (nnU-Net) in 2018, which is a model based on U-Net. They simplified the redundant structure of U-Net and focused on improving its performance and generalization by self-adaptation to the dataset at hand. The resulting nnU-Net architecture achieved good results in the Medical Segmentation Decathlon challenge in 2018.

Moreover, the architecture of nnU-Net provided a way to solve 3D medical image segmentation tasks. The nnU-Net architecture consists of two-dimensional (2D) and 3D U-Nets which are then used in a U-Net cascade (124). In detail, during the nnU-Net training multiple low- and high-resolution 2D and 3D U-Nets are generated, due to limitations in GPU memory with patches of the imaging as input images (architecture shown in Figure 3A) (125). During the inference process, the segmentation of the different U-Net models is concatenated and refined by a second 3D U-Net.

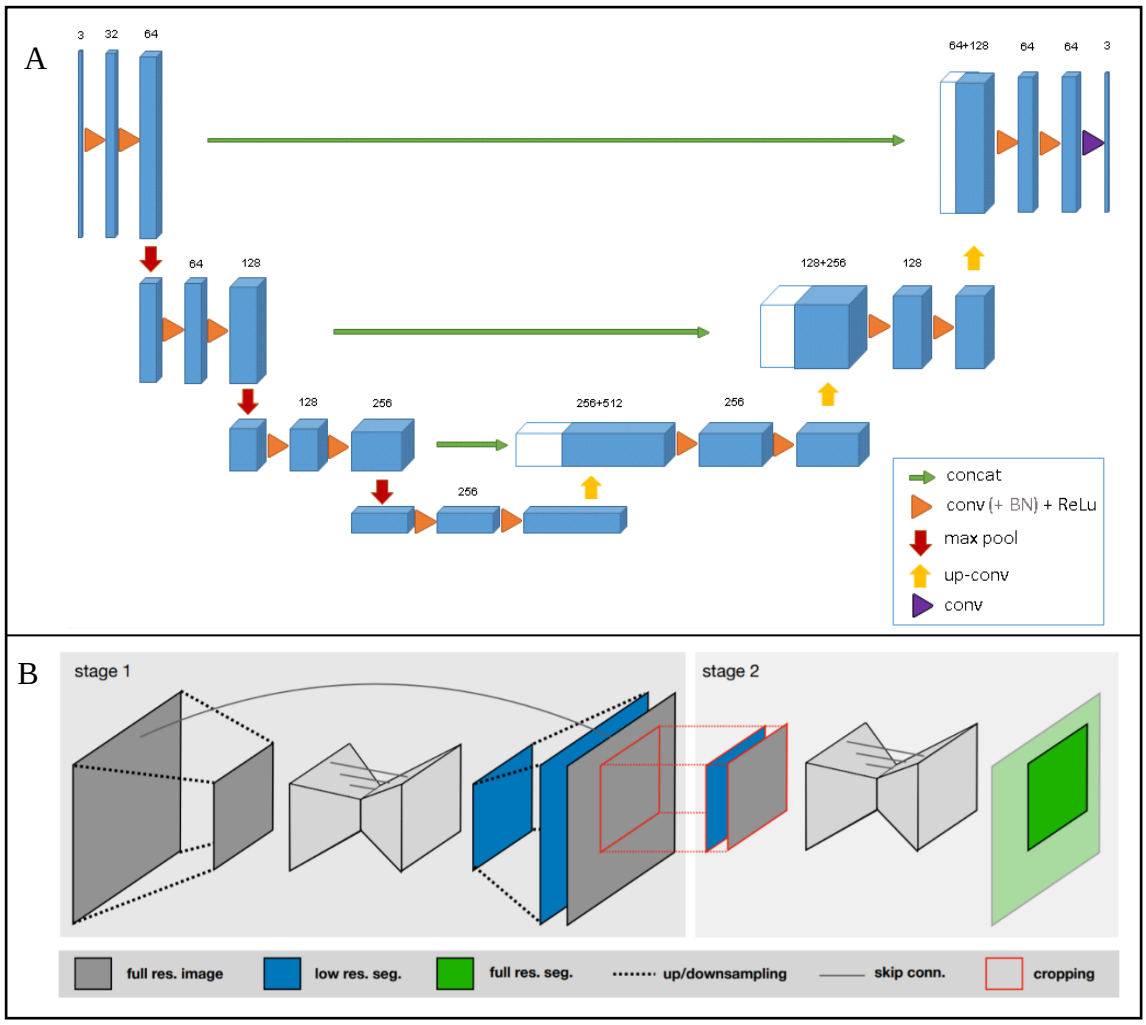


Figure 3. The 3D U-Net architecture (A) (125) and U-Net cascade (B) (124).

1.2.3 Assessment of the algorithm's performance

During the training process of a deep learning network, stopping at the right moment is crucial to avoid overfitting (training too long) and generalization (training too short) of the algorithm.

In most segmentation tasks, the category of interest represents data points with special properties, which is termed the "positive" class, while another category acts as a "background" category, which includes everything else and is called the "negative" class (126). The results of prediction can be divided into 4 categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) based on the comparison of actual and predicted conditions (126,127). (Shown in Table 2). Within this category, TP and TN are the correct classifications of positive and negative conditions, and conversely, FP and FN represent misclassification of positive and negative conditions, respectively. To evaluate the performance of the AI model in predictions,

several statistical measures are employed according to the goal of the study and the four kinds of predicated results.

Table 2. Results of prediction.

		Predicted condition	
		Positive	Negative
Actual condition	Positive	TP	FN
	Negative	FP	TN

TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

1.2.3.1 Dice coefficient

The Dice coefficient (DC), also called the F1 score, was first proposed by L. R. Dice in 1945 (128) and is now used to evaluate AI algorithms that perform segmentation on medical images by comparing the prediction with the ground truth (129–132). The DC is an ensemble similarity measurement function, defined as follows, and ranges in [0, 1], while 0 indicates that the prediction fails to match the ground truth completely, and conversely, 1 indicates that the prediction has perfectly matched the ground truth (128–132).

$$DC\ (A, B) = \frac{2\,|A \cap B|}{|A| + |B|}$$

where A is the ground truth, B is the segmentation result (prediction), A ∩ B is the intersection of A and B, which presented the TP (shown in Figure 4).

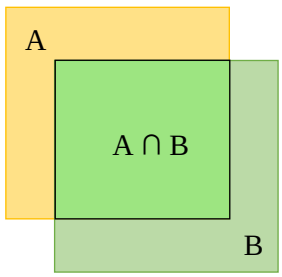


Figure 4. The performance of AI. A: Ground Truth; B Prediction; A ∩ B: True positives.

However, DC only compares the part that is predicted to be positive (B) with the truly positive part (A), which presents |TP+FP| and |TP+FN| respectively. Therefore, the equation of DC can be written as below, which does not consider the part of TN (133).

$$\text{DC (A, B)} = \frac{2\,|A \cap B|}{|A| + |B|} = \frac{2\,TP}{2TP+FN+FP}$$

In some medical segmentation tasks, it is common that the target organ occupies only a small region of the scan, which makes the predictions strongly biased towards the background and results in a skewed prediction (134).

1.2.3.2 Matthews correlation coefficient

Recently, the Matthews correlation coefficient (ϕ) was suggested as a more reliable statistical measurement compared to DC since it considers the 4 categories (TP, TN, FP, FN) together, with more robust predictions on imbalanced datasets (135). The Matthews correlation coefficient will get a high score only when all 4 predicted categories are highly accurate. In this setting, the Matthews correlation coefficient is suggested to be the more robust metric in binary classification compared to the F1 score (135–137). The Matthews correlation coefficient ranges from -1 to 1, with -1, 0, and 1 indicating a complete misprediction, random prediction, and perfect prediction, respectively (135). The larger the Matthews correlation coefficient, the better the prediction.

$$\text{Matthews correlation coefficient } (\phi) = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

(worst value: –1; best value: +1)

1.2.4 Application status of AI in imaging of liver metastases

Although MRI was suggested as the most sensitive technique for detecting liver metastases, a review in 2017 indicated that CT was the most common medical imaging modality used for AI analysis (138), because AI model training on MRI is inherently more difficult due to the only semiquantitative nature or MRI images and other factors related to the image acquisition and underlying pathology (139). Thus, in comparison with CT images, automatic liver tumor segmentation in MRI is more challenging. However, the combination of the clear tumor-liver interface on the HBP sequence of Gd-EOB-MR and the excellent performance of nnU-Net in multi-organ segmentation with a fully automatic manner seems to be the solution to these problems.

1.3 Aims of this study

1.3.1 Establishing an AI model

The first goal of the study was to establish an AI model with high accuracy and high robustness, which would allow the automatic delivery of the absolute 3D volume of the HTL.

1.3.2 Application of the AI model

Based on the obtained 3D-AI-model, the 3D volume and HTL of NELMs are automatically segmented and relevant response indicators are calculated for comparison with the MCC decisions, which is as part of our ENETS Center of Excellence.

2. Materials and methods

The main data flow is shown in the flowchart below (Figure 5), and the details are reflected in the following subsections.

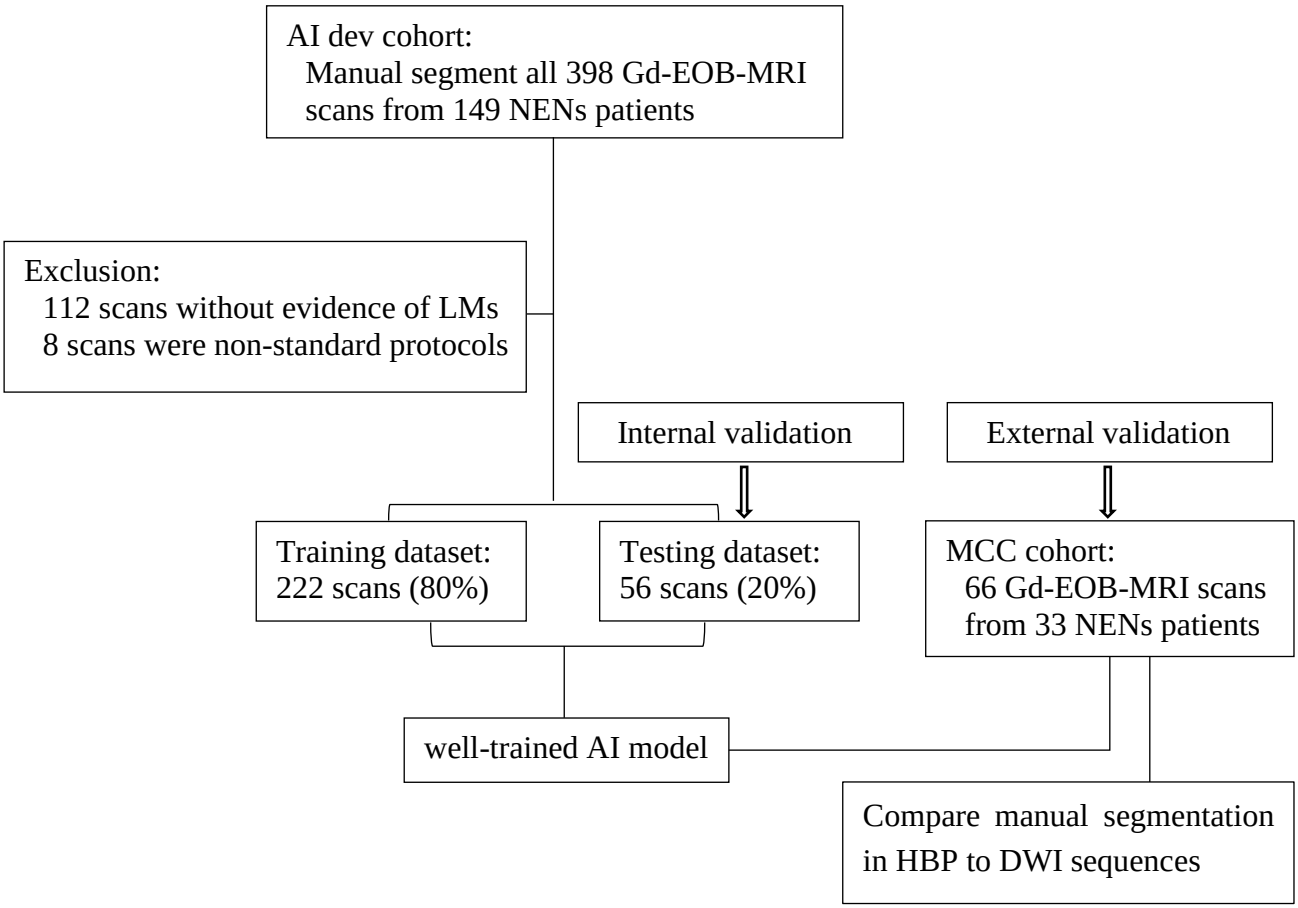


Figure 5. Data flowchart of the study.

2.1 Part 1: Establish a high-precision AI model

As shown in Figure 5, all 398 MRI scans of 149 patients with NENs who underwent Gd-EOB (Primovist, Bayer, Berlin, Germany) enhanced MRI (1.5 T, Siemens Aera) between January 2015 and August 2018 at the radiology department of Charité Campus Virchow-Klinikum were retrospectively identified from our radiology database as the AI development (AI dev) cohort, regardless of whether any therapeutic interventions had been done. Hepatobiliary phase images were acquired at minute 20 after injection of 0.025 mmol/kg (0.1 ml/kg) body weight of Gd-EOB at a rate of 2 ml/s, using the care bolus technique.

2.1.1 Manual segmentation on AI dev cohort

The HBP sequences of all MRI scans were anonymized, and participants were blinded throughout the manual segmentation process and could only identify NELMs based on radiological experience (combined with T2, DWI sequences). All segmentations were refined by a radiologist with >5 years of experience in abdominal MRI.

The task of segmentation is usually defined as identifying the contours of the object of interest in medical images. In our project, the borders of the metastases were defined between the contrast-enhancing liver tissue and the non-enhancing metastases. The Medical Imaging Interactive Toolkit (MITK) was used during segmentation and volumetric 3D segmentation was performed of the liver and the liver metastases in the HBP. The segmentation was performed manually using the polygon region of interest (ROI) built into MITK.

2.1.2.1 Manual segmentation for the liver in HBP sequences

The first step in the segmentation task is liver segmentation. It is very important to divide the contours of the liver as perfectly as possible, slide-by-slide. Of note, vessels and biliary ducts should be excluded where feasible. An example of the different slices of the liver and the corresponding manual segmentation results are shown in Figure 6. In some slices, there is overlap between the liver and adjacent organs making it difficult to distinguish the boundaries (mediastinum in C, kidney in G). Or there are non-hepatic structures within the liver (hepatic ligament, portal vein, inferior vena cava in E) that need to be excluded from the target area.

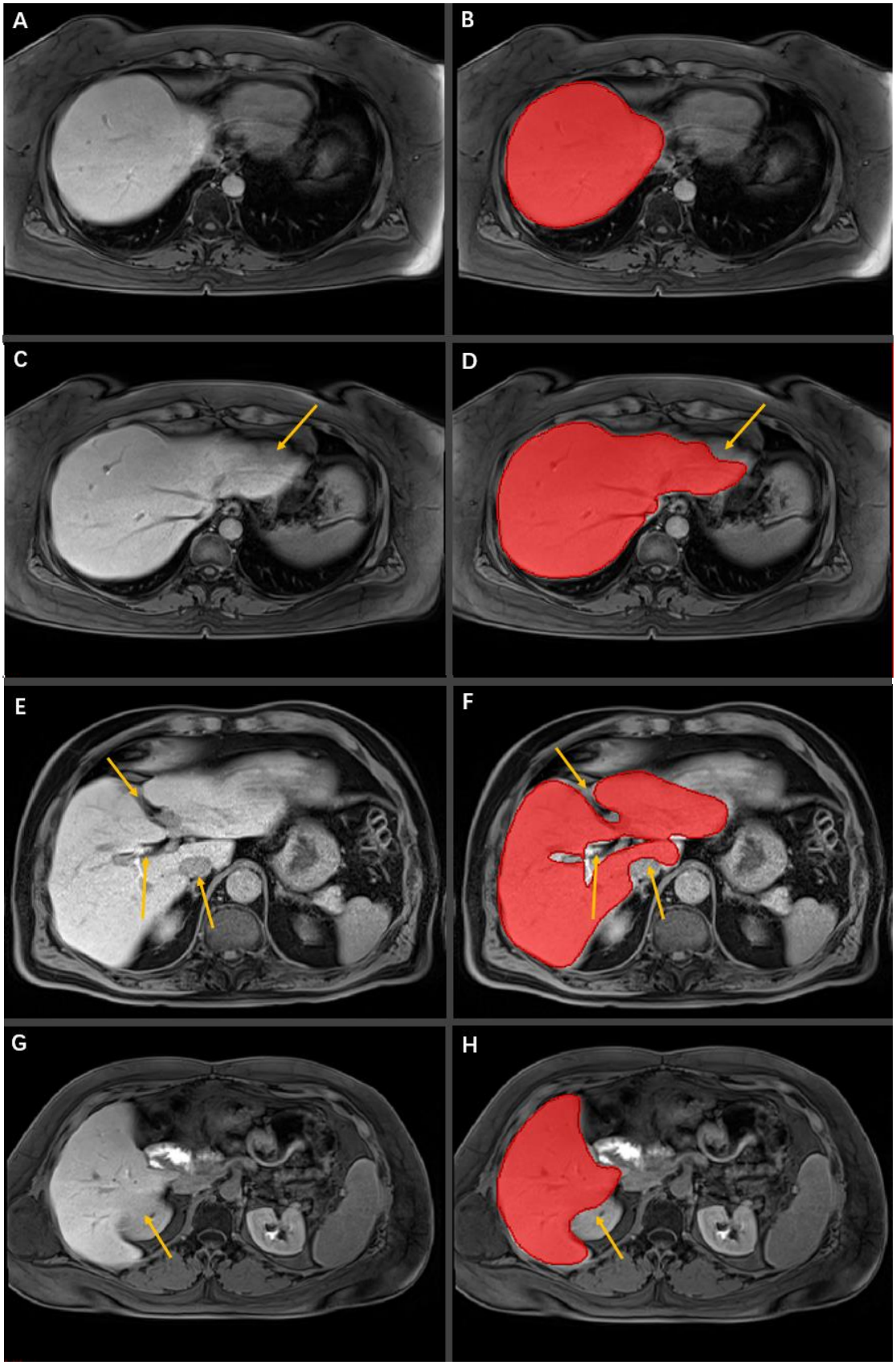

A
B
C
D
E
F
G
H

Figure 6. Manual segmentation of the liver (red areas in B, D, F, H) on different slices of the HBP sequence 20 min post Gd-EOB-MRI (A, C, E, G). Yellow arrows indicate the mediastinum (C, D), hepatic ligament, portal vein, inferior vena cava (E, F), and kidney (G, H), respectively.

2.1.2.2 Manual segmentation for NELMs in HBP sequences

The second step in the segmentation task is NELMs segmentation. The boundary between the NELMs and normal liver tissue on the HBP sequence is shown as a contrast between hypointense and hyperintense images. Notably, the hepatic vessels are also hypodense in the HBP sequence. In cases of diffuse lesions, this is a complex and time-consuming process that requires absolute patience to differentiate small hepatic tumor lesions from hepatic vessels. Additionally, the distribution patterns of NELM were assessed as follows: singular (n=1), multiple (n≤10), and diffuse (n>10) and according to their localization: unilobar or bilobar.

Manual segmentation of a patient with diffuse liver metastases (A, B, C) and a patient after right hemihepatectomy (D, E, F) is shown in Figure 7. In Figure 7A, the yellow and white arrows point to two adjacent low-density circular hypointense areas, for which it is difficult to determine whether it is a metastasis or a vessel from this slice alone. So, it is important to look dynamically at multiple consecutive slices: NELMs appear relatively fixed in position, whereas hepatic vessels show a sliding migration on successive slices. The difference between NELM (yellow arrow) and hepatic vessels (white arrow) is depicted in Figure 7B. An overlap image of the liver segmentation with the NELM segmentation is shown in Figure 7C.

The yellow and white arrows in the right hemihepatectomy case (Figure 7D) point to NELMs and hepatic vessels respectively. The green area in Figure 7E is the result of the segmentation of the two NELMs, and the overlapping image of the liver segmentation with the NELM segmentation is shown in Figure 7F. From this slice alone, the distribution pattern of NELMs is multiple (2 metastases), and unilobar (left).

In addition to dynamic observation, it is also possible to discriminate NELMs from hepatic vessels simultaneously with the help of the T2-weighted phase and/or DWI phases. The NELMs are highlighted in the T2 and DWI phases can thus be distinguished from the vessels.

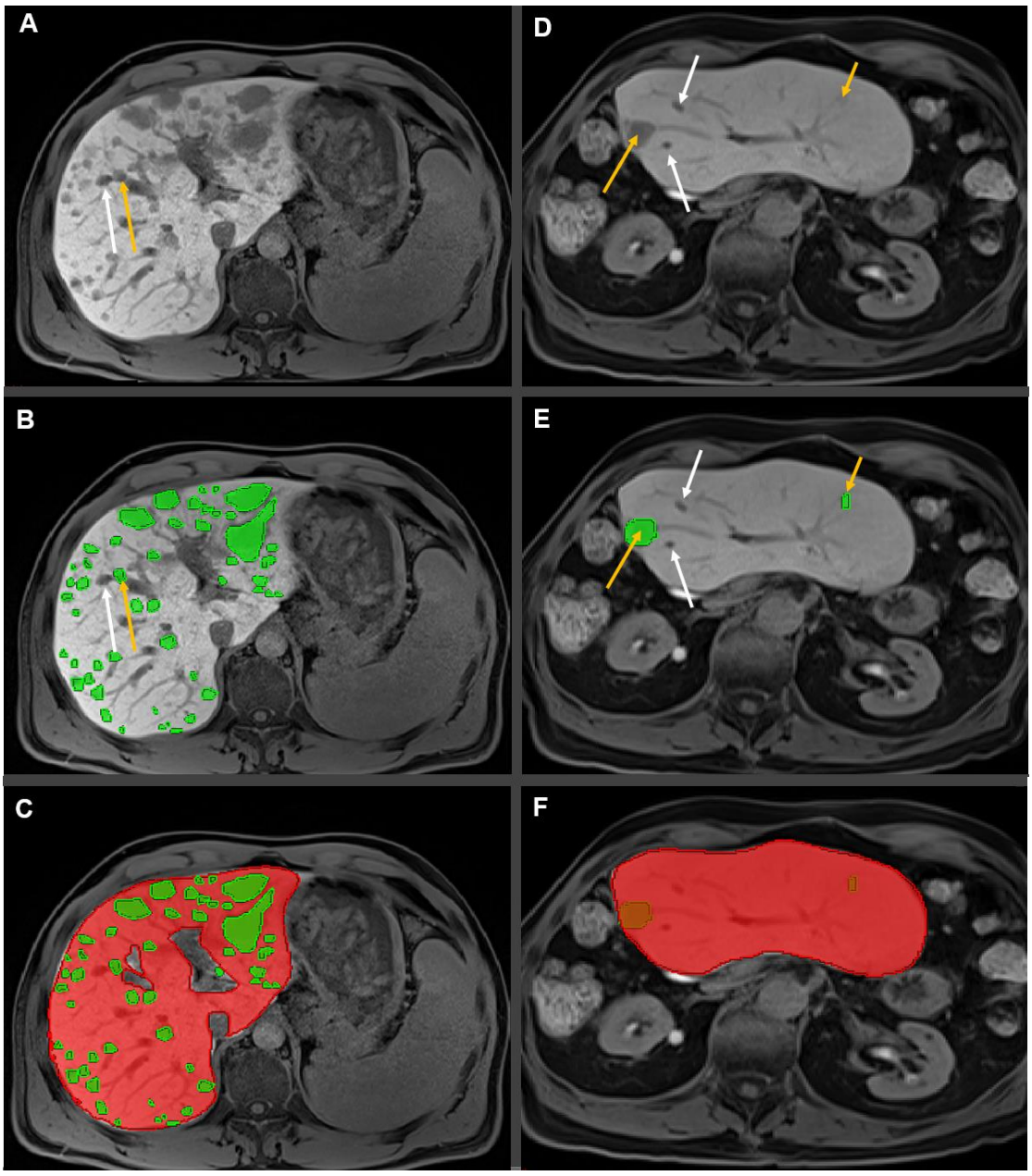


Figure 7. Examples of manual segmentation of NELMs (green areas in B, E) and the liver (red areas in C, F) on different slices of the HBP sequence on Gd-EOB-MRI (A, D). (A, B, C) diffuse pattern NELMs case; (D, E, F) post right hemihepatectomy case. Yellow arrows indicate the NELMs, whereas, white arrows indicate hepatic vessels (A, B, D, E). The HBP sequences were acquired 20 min after manual bolus injection of Gd-EOB-DTPA.

2.1.2 Model training

After checking all segmentation results and the corresponding clinical data, 120 out of 398 scans were excluded for missing evidence of NELM (n=112) [diagnosis of no NELM or CR after

treatment for NELM] or due to non-standard scan protocols (n=8). 222 out of the remaining 278 (80%) MRI scans were randomly chosen for the model training. The nnU-Net deep learning framework was used as the basic algorithm, which is available open-source on Github (https://github.com/MIC-DKFZ/nnUNet).

2.1.3.1 Preprocessing

Image preprocessing aims to standardize different types of imaging, in preparation for model training. According to the nnU-Net framework, the general preprocessing of 3D medical image segmentation can be divided into four steps, which are data format conversion, cropping, resampling resamples, and normalization (124,140).

Data format conversion: The data was converted from Dicom images to NIFTI/NRRD format for easier file handling and improved file loading using the dcm2niix conversion tool (141).

Cropping: Image cropping is to crop a 3D medical image to its non-zero area. The specific method is to find the smallest 3D bounding box in the image. The value outside the bounding box area is 0, and the image is cropped using this bounding box. Compared with before cropping, the cropped image has no effect on the final segmentation result, but it can reduce the image size, avoid useless calculations, and improve calculation efficiency. In our case, this step was omitted, as the final segmentation task should be able to work with full abdominal field-of-view imaging.

Resampling: The purpose of resampling is to solve the problem of inconsistency in the actual space size represented by a single voxel in different images in some 3D medical image data sets. Because the CNNs only operate in the voxel space, the size information in the actual physical space will be ignored. To avoid this difference, it is necessary to resize different image data in the voxel space to ensure that the actual physical space represented by each voxel is consistent in different image data. As our imaging was acquired using a standardized imaging protocol, resampling was not necessary in our case.

Normalization: The purpose of normalization is to make the gray value of each image in the training set have the same distribution and is usually applied on a per-patient basis (z-score normalization).

2.1.3.2 Training Procedure

The AI model was trained end-to-end on the dataset of liver scans from Gd-EOB MRI. Five-fold cross-validation was used in the training process. The AI's segmentation performance in liver and NELMs was assessed using the Matthews correlation coefficient.

2.1.3 Model testing (Internal validation)

The remaining 56 scans (20%) were used as the internal validation to test the model's accuracy. The original HBP sequences of Gd-EOB MRIs were fed into the trained AI model and segmentations and volumes of ROI (liver and NELMs) were automatically determined. Matthews correlation coefficient was then used to compare the results of the AI segmentation and the manual segmentation.

2.2 Part 2: Application of the AI model

2.2.1 MCC cohort (External validation):

For the external validation, additional 33 patients diagnosed with NELMs were collected from our MCC between January 2019 and January 2020 which had received Gd-EOB MRI (1.5T or 3.0T) both at baseline (BL) and follow-up (FU, within 6 months), for a total of 66 MRIs as the MCC cohort.

2.2.1.1 Manual segmentation for the liver and NELMs in HBP sequences

All HBP sequences of the 66 MRIs in the MCC cohort were manually segmented, using the same methodology as for the AI dev cohort.

2.2.1.2 Manual segmentation of liver and NELMs compared with AI in HBP sequences

The model was applied to the external MCC cohort and the volumes of NELM and the liver at BL scan and FU scan were calculated and absolute and relative change in NELM volume and HTL were compared to the MCC decisions. Based on the images provided, MCC decisions were categorized as therapy success (stable disease (SD) or partial regression (PR)) or therapeutic failure (progressive disease (PD)).

Matthews correlation coefficients were used to compare the results of AI and manual segmentation. Two independent sample rank-sum tests were used to compare the change in tumor volume (both absolute and relative changes in NELM volume and HTL) between the treatment success and failure groups. ROC curves were used to explore the optimal threshold for disease progression.

2.2.2 Comparison of HBP sequences with DWI sequences in the MCC cohort

Additionally available free-breathing fat-suppressed single-echo planar DWI using b-factors of 50, 400, and 800 s/mm^2 were available in the liver MRI protocol. Manual segmentation of DWI images with b=400 s/mm^2 was performed in the MCC cohort as an additional comparative study with the HBP sequence.

2.2.2.1 Manual segmentation for the liver and NELMs in DWI sequences

The liver is depicted hypointense on DWI sequences, while the NELMs are moderately hyperintense. The segmentation process was the same as for the HBP sequence, i.e., the contours of the liver and NELMs were distinguished as accurately as possible. Figure 8 shows an example of manual segmentation of the liver and NELMs on a DWI sequence.

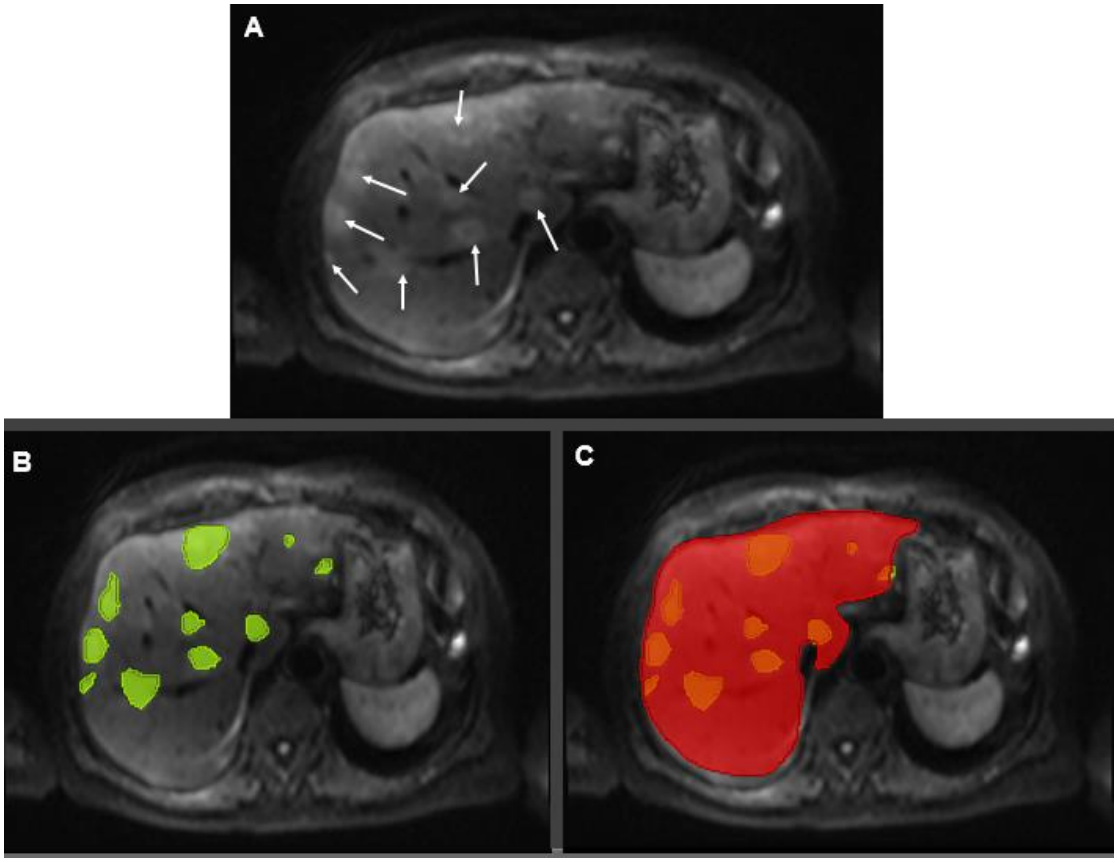


Figure 8. Example of manual segmentation of the liver and NELMs on a DWI sequence. (A) Original DWI image, the white arrows indicate the NELMs (not fully listed); (B) manual segmentation of the NELMs (green); (C) the overlapping of the segmentation of the liver (red) and the NELMs (green). Manual segmentation was performed on the DWI images with b=400 s/mm^2.

2.2.2.1 Comparison of Manual segmentation for the liver and NELMs in DWI sequences with HBP sequences

NELM and liver volumes, HTL, absolute and relevant changes in NELM volumes, and HTL in manual segmentations were calculated and correlated between the DWI and HBP sequences.

2.3 Statistics:

Statistical analysis was performed using SPSS Statistics (IBM, Version 25). Kolmogorov-Smirnov testing was applied to determine the normal distribution and in case of a non-normal distribution of the data, nonparametric testing was performed.

Median and interquartile range (IQR) was used for descriptive data. Segmentations and their relative differences were calculated using the formula: $\frac{\text{model's volume – radiologists' volume}}{\text{radiologists' volume}}$. Matthew's correlation coefficients (ϕ) were used for accuracy estimation. MCC decisions were compared to the AI-based segmentation in the following manner: HTL was calculated using the following formula: $\frac{\text{NELM volume}}{\text{Liver volume – NELM volume}} \times 100\%$. Absolute volume changes (Δ_{abs}) were calculated by the difference: $\text{Volume}_{FU} - \text{Volume}_{BL}$. Relative volume changes (Δ_{rel}) were calculated by the formula: $\frac{\text{Volume}_{FU} - \text{Volume}_{BL}}{\text{Volume}_{BL}} \times 100\%$. Statistical tests were performed using the Mann-Whitney U test (for continuous data), χ^2 test or Fisher's Exact Test (for numerical data), and sign test (for related samples). Correlation analysis of continuous variables was performed using Spearman's rank test and the corresponding correlation coefficients (rs) were calculated. ROC curve analysis was performed and the optimal cut-off value for disease progression was calculated by Youden indexes.

3. Results

3.1 Patient cohorts

After excluding 120 MRI scans from 31 patients, the characteristics of the 118 training patients in the AI cohort (278 imaging datasets) and 33 validation patients in the MCC cohort (66 imaging datasets) are presented in Table 3. The primary sites are significantly different between the two groups (P = 0.0003). The most common primary tumor sites in the AI dev cohort were the ileum (50.8%) and pancreas (44.9%), whereas in the MCC cohort ileum and pancreas each accounted for 36.4, 27.2% of the tumors originated from other sites. SR status was significantly different between the two cohorts (P = 0.047). Whereas there was no significant difference in gender, age, Ki67, grading, NET/NEC, functionality, or presence of extrahepatic metastases between the AI and MCC cohorts.

Table 3. The characteristics of AI and MCC cohorts

	AI dev cohort	**MCC cohort**	**P-value**
Numbers of Patients	118	33	-
Numbers of Scans	278	66	-
Gender (M: F)	51:67	18:15	0.323
Age(median)	59.14 (49.58-66.43)	56.45 (48.62-67.40)	0.613
Ki67(%, median)	5.0 (2.0-10.0)	7.0 (2.5-13.0)	0.578
Primary Site			**0.0003**
Pancreas	53 (44.9%)	12 (36.4%)	
Ileum	60 (50.8%)	12 (36.4%)	
Other	5 (4.3%)	9 (27.2%)	
Grading			0.703
1	35 (29.7%)	8 (24.2%)	
2	73 (61.9%)	23 (69.7%)	
3	10 (8.4%)	2 (6.1%)	
NET: NEC	115:3	31:2	0.300
Functionality			0.675
yes	37 (31.4%)	12 (36.4%)	

no	81 (68.6%)	21 (63.6%)	
Extrahepatic metastases	83 (70.3%)	27 (81.8%)	0.268
SR			**0.047**
pos	97 (82.2%)	32 (97.0%)	
neg	21 (17.8%)	1 (3.0%)-	

Values are presented as n (%) or median (IQR). χ^2 test, Fisher's Exact Test, or Mann–Whitney U-test were used depending on data distribution. $P < 0.05$ was considered significant (bold) between AI dev and MCC cohorts. AI: artificial intelligence; MCC: multidisciplinary cancer conference; NET: neuroendocrine tumor; NEC: neuroendocrine carcinoma.

Adapted from Fehrenbach et al. (142).

3.2 Validation of the model

3.2.1 Internal validation (AI dev cohort) in HBP sequences

Fifty-six scans of 278 (20%) from the AI dev cohort were randomly chosen for internal validation to test the model. The median NELM volume was 17.25 cm^3 (IQR: 4.48–60.93 cm^3) as calculated using the nnU-Net model and 16.17 cm^3 (IQR: 4.87–58.16 cm^3) for the manual segmentation (as target volume). The difference between the manual and automatic segmentation for the NELMs was -3.7% (IQR: -24.54–+11.83%). The median segmented liver volume was 1639.9 cm^3 (IQR 1366.1–1960.7 cm^3, manual) and 1659.0 cm^3 (IQR 1404.2–1966.9 cm^3, automatic segmentation). The difference in median relative liver volume between the automatic and manual segmentation was +0.9% (IQR: -0.7–+4.2%). The automatic segmentation yielded a median ϕ of 0.76 (IQR: 0.68–0.83) in the segmentation of metastases and 0.95 (IQR: 0.95–0.96) for the entire liver (Figure 9).

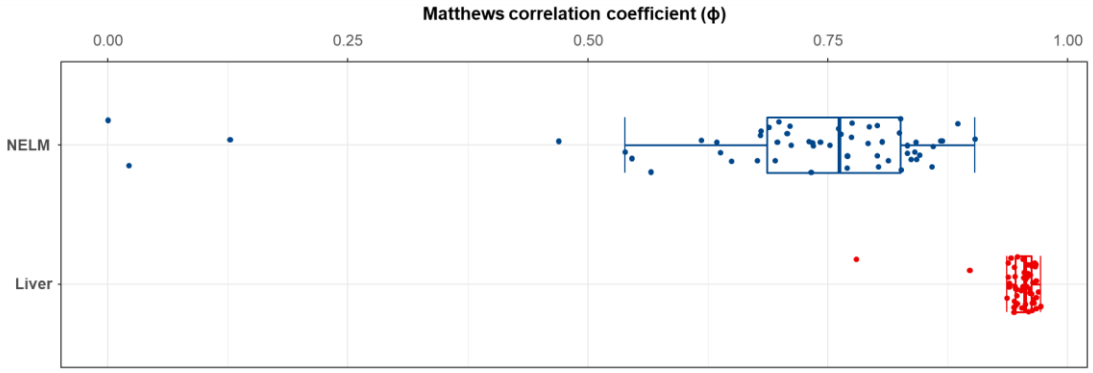


Figure 9. Distribution of Matthews correlation coefficients (ϕ) in NELM (blue) and liver (red) segmentations. Data are expressed using median and interquartile ranges (IQR).

The case-wise comparison between Matthews correlation coefficient (ϕ) and manual target volumes is presented in Figure 10. The blue dots indicate ϕ in the liver, whereas the red triangles indicate ϕ in NELMs. A weak overlap (ϕ < 0.2, orange circled) was seen in only three out of 56 patients (5.4%). A case-by-case analysis of these three cases revealed that two of the patients had very small NELM volumes (0.1 and 0.2 cm^3). The metastases showed an atypical imaging morphology in the HBP with a hyperintense aspect and were therefore omitted by the automatic segmentation. In the remaining cases, a correlation coefficient of at least 0.45 (for NELMs) and at least 0.75 (for liver volume) was observed.

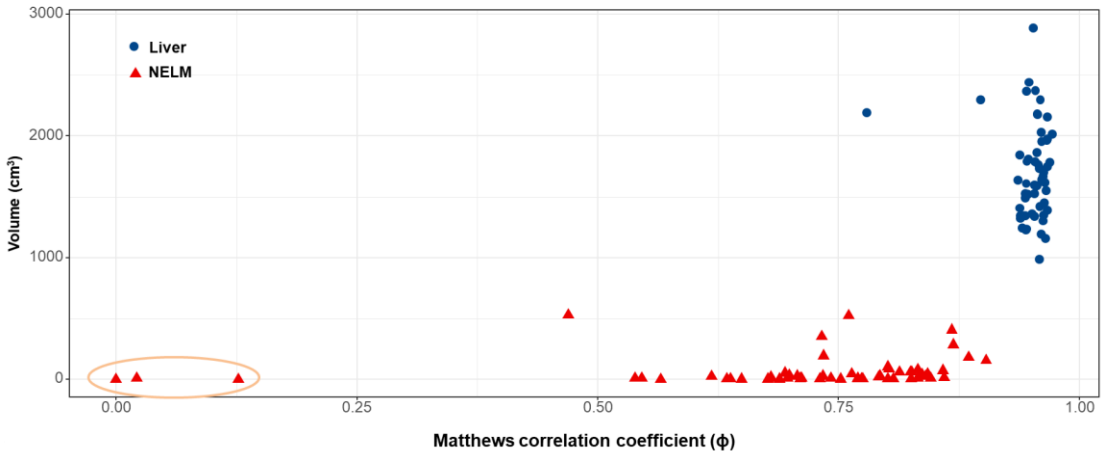


Figure 10. Distribution of Matthews correlation coefficient (ϕ, x-axis) in comparison to the target volume (cm^3, y-axis) in the liver (blue dots) and NELMs (red triangles). Orange circled three: two had very low NELM volume and one showed atypical, hyperintense signal intensities of the metastases.

When regarding high and low NELM volume patients (cut-off-point at the medial of 16.17 cm^3), there was a significantly higher ϕ in the high NELM volume cases (median ϕ: 0.80; IQR: 0.73–0.84) in comparison to the low NELM volume cases (median ϕ: 0.71, IQR: 0.64–0.78; p = 0.0025, Figure 11).

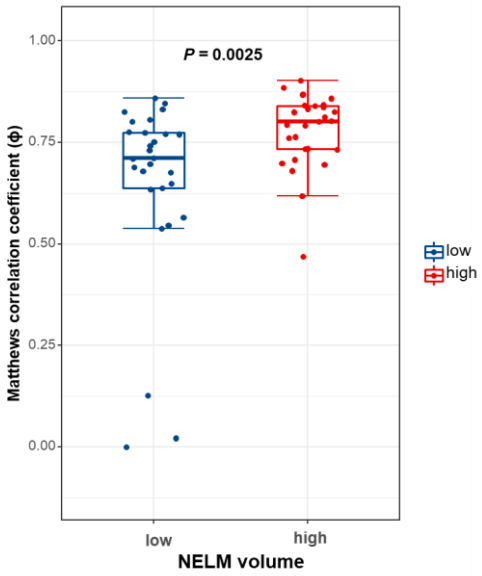


Figure 11. Comparison of Matthews correlation coefficient (ϕ) between low (blue) and high (red) NELM volume (cm^3). Data were shown in median and interquartile ranges (IQR). P-value is based on Mann–Whitney U-test. NELM: neuroendocrine liver metastases.

3.2.2 External validation (MCC cohort) in HBP sequences

Sixty-six scans of 33 patients (BL and FU) from the MCC cohort were used for external validation. In this dataset, the automatic segmentation yielded a median ϕ of 0.86 (IQR: 0.81–0.91) for NELM segmentation and of 0.96 (IQR: 0.95–0.96) for whole liver segmentation.

A comparative example of manual and AI segmentations is shown in Figure 12. In the original image (A), there are two NELMs on the right side of the liver (white arrows indicated). The manual segmentation of the liver is shown in Figure 12B, whereas the AI segmentation is shown in Figure 12C, and their overlap is shown in Figure 12D. Similarly, the segmentations of NELMs are listed in Figure 12E (manual), Figure 12F (AI's), and Figure 12G (overlap).

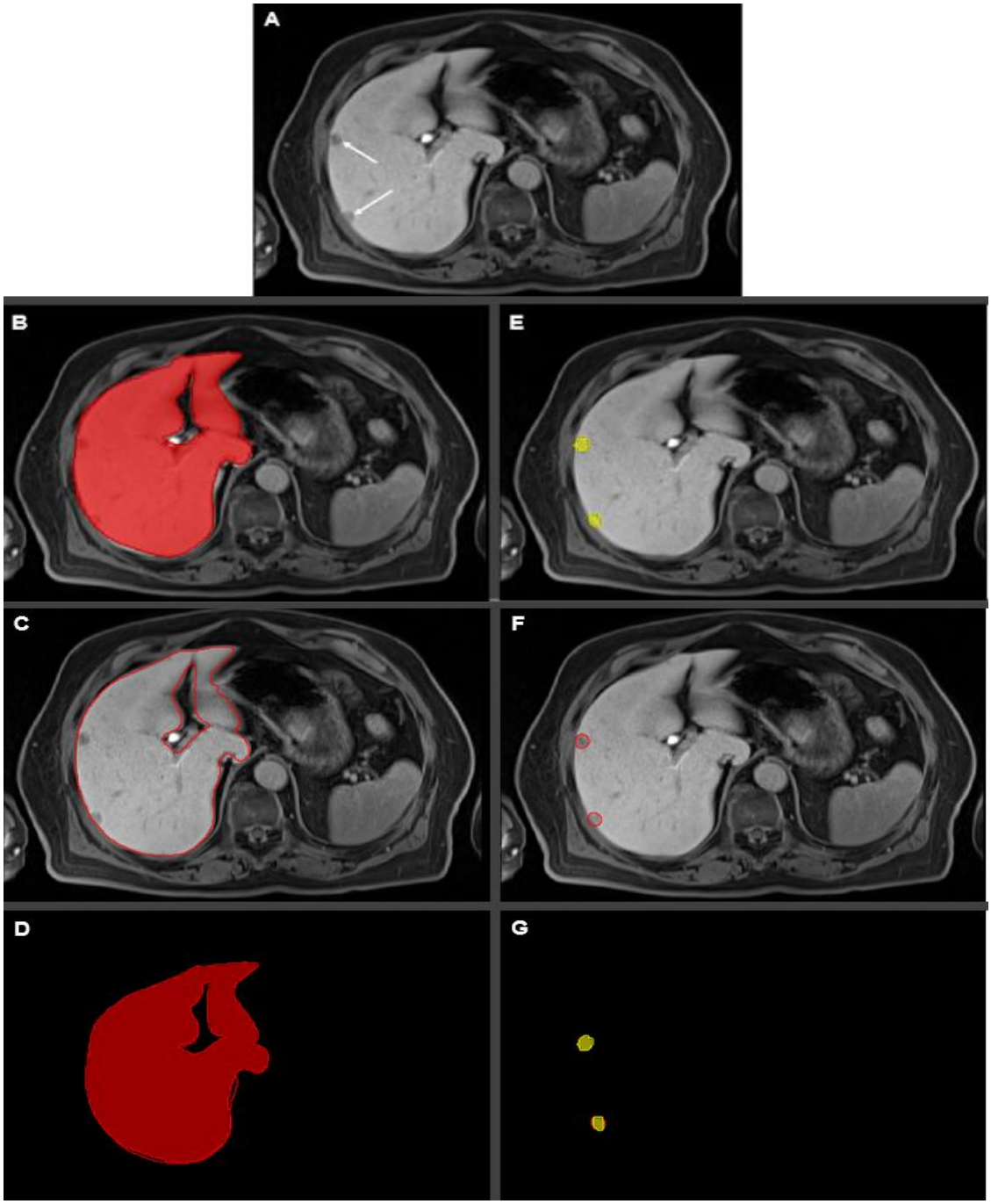


Figure 12. Comparison between manual and AI segmentations. (A) Original HBP MRI, the white arrows indicate the NELMs; (B, E) manual segmentation of the liver and NELMs; (C, F) AI's segmentation of the liver and NELMs; (D, G) overlapping the manually and AI's segmentation of the liver and NELMs.

3.3 AI-based NELM Segmentation and Correlation with MCC decision

3.3.1 Segmentation results from the AI model

The segmentation results from the model on the MCC cohort are summarized in Table 4. The median NELM volume was increased from 23.48 cm^3 (IQR: 10.45–113.17 cm^3) at the baseline to 86.93 cm^3 (IQR: 12.08–204.50 cm^3) at the follow-up, while the median liver volume was from 1582.23 cm^3 (IQR: 1336.25–2030.03 cm^3) at the baseline to 1716.75 cm^3 (IQR: 1477.12–2092.94 cm^3) at the follow-up, resulting in a median HTL of 1.57% (IQR: 0.55–7.05%) and 5.93 % (IQR: 0.99–11.74 %), respectively.

The therapeutic responses were classified by the MCC based on the RECIST 1.1 as stable disease/partial remission in 16 (48%) patients (SD: n=14; PR: n=2) and progression in 17 (PD, 52%) of the patients. In patients classified as a therapeutic success, the median NELM volume, liver volume, and HTL were 75.45 cm^3 (IQR: 12.35-141.65 cm^3), 1692.26 cm^3 (IQR: 1475.09–2061.63 cm^3), 4.41vol.-% (IQR: 0.87–7.83 vol.-%) at the baseline, and 66.78 cm^3 (IQR: 11.64–167.82 cm^3), 1725.30 cm^3 (IQR: 1471.78–2130.28 cm^3), 3.75 vol.-% (IQR: 0.75–8.88 vol.-%) at the follow-up, respectively. Whereas the median absolute NELM volume changed (Δ_{abs}NELM) +0.76 cm^3 (IQR: -18.07–39.32 cm^3), the median absolute HTL changed (Δ_{abs}HTL) -0.03 vol.-% (IQR: -1.28–0.23 vol.-%), resulting in an increase in the median relative change of the NELM volume (Δ_{rel}NELM, +3.93% (IQR -15.75–10.36%)), and a decrease in the median relative change of HTL (Δ_{rel}HTL, -3.45% (IQR: -18.11–11.15%)).

Among the patients classified as therapeutic failures, the median NELM volume, liver volume, and HTL were 19.15 cm^3 (IQR: 7.04–78.44 cm^3), 1580.35 cm^3 (IQR: 1290.13–1902.53 cm^3), 1.46 vol.-% (IQR: 0.34–5.97 vol.-%) at the baseline and were 86.93 cm^3 (IQR: 24.40–253.32 cm^3), 1716.75 cm^3 (IQR: 1451.10–2106.81 cm^3), 5.93 vol.-% (IQR: 1.47–16.78 vol.-%) at the follow-up, respectively. The median Δ_{abs}NELM was +59.70 cm^3 (IQR: 16.49–156.59 cm^3) and the median Δ_{abs}HTL was +4.94 vol.-% (IQR: 1.07–9.78 vol.-%). In therapeutic failure, median Δ_{rel}NELM was +242.68% (IQR: 124.56–463.87%) and median Δ_{rel}HTL was +204.49% (IQR: 109.39–490.19%).

Table 4. Aggregate statistics of the AI segmentation results for the MCC cohort.

Variable	Overall

	BL	FU
N	33	33
NELM (cm^3)	23.48 (10.45–113.17)	86.93 (12.08–204.50)
Liver (cm^3)	1582.23 (1336.25–2030.03)	1716.75 (1477.12–2092.94)
HTL (vol.-%)	1.57 (0.55–7.05)	5.93 (0.99–11.74)
Δ_{abs}NELM (%)	14.70 (0.76–96.35)	
Δ_{abs}HTL (%)	0.98 (-0.03–5.41)	
Δ_{rel}NELM (%)	58.51 (3.93–245.64)	
Δ_{rel}HTL (%)	64.97 (-3.44–223.31)	

	Therapy Success		Therapy Failure	
	BL	FU	BL	FU
N	16	16	17	17
NELM (cm^3)	75.45 (12.35-141.65)	66.78 (11.64–167.82)	19.15 (7.04–78.44)	86.93 (24.40–253.32)
Liver (cm^3)	1692.26 (1475.09–2061.63)	1725.30 (1471.78–2130.28)	1580.35 (1290.13–1902.53)	1716.75 (1451.10–2106.81)
HTL (vol.-%)	4.41 (0.87–7.83)	3.75 (0.75–8.88)	1.46 (0.34–5.97)	5.93 (1.47–16.78)
Δ_{abs}NELM (%)	0.76 (-18.07–39.32)		59.70 (16.49–156.59)	
Δ_{abs}HTL (%)	-0.03 (-1.28–0.23)		4.94 (1.07–9.78)	
Δ_{rel}NELM (%)	3.93 (-15.75–10.36)		242.68 (124.56–463.87)	
Δ_{rel}HTL (%)	-3.45 (-18.11–11.15)		204.49 (109.39–490.19)	

Values are shown as median and interquartile ranges (IQR). P-values are based on Mann–Whitney U-test. BL: baseline; FU: follow-up; NELM: neuroendocrine liver metastases; HTL: hepatic tumor load; Δ_{abs}NELM: absolute NELM volume change; Δ_{abs}HTL: absolute HTL change; Δ_{rel}NELM: relative NELM volume change; Δ_{rel}HTL: relative HTL change.
Table adapted from Fehrenbach et al. (142).

There was a significant difference in patients with stable disease/partial remission (therapeutic success) and progressive disease (therapeutic failure) in all response variables: Patients with stable disease/partial remission had significantly lower changes in median absolute NELM

volume (Δ_{abs}NELM), median absolute HTL (Δ_{abs}HTL), median relative NELM volume (Δ_{rel}NELM) and median relative HTL (Δ_{rel}HTL) compared to patients with progressive disease (all $p < 0.001$, shown in Figure 13).

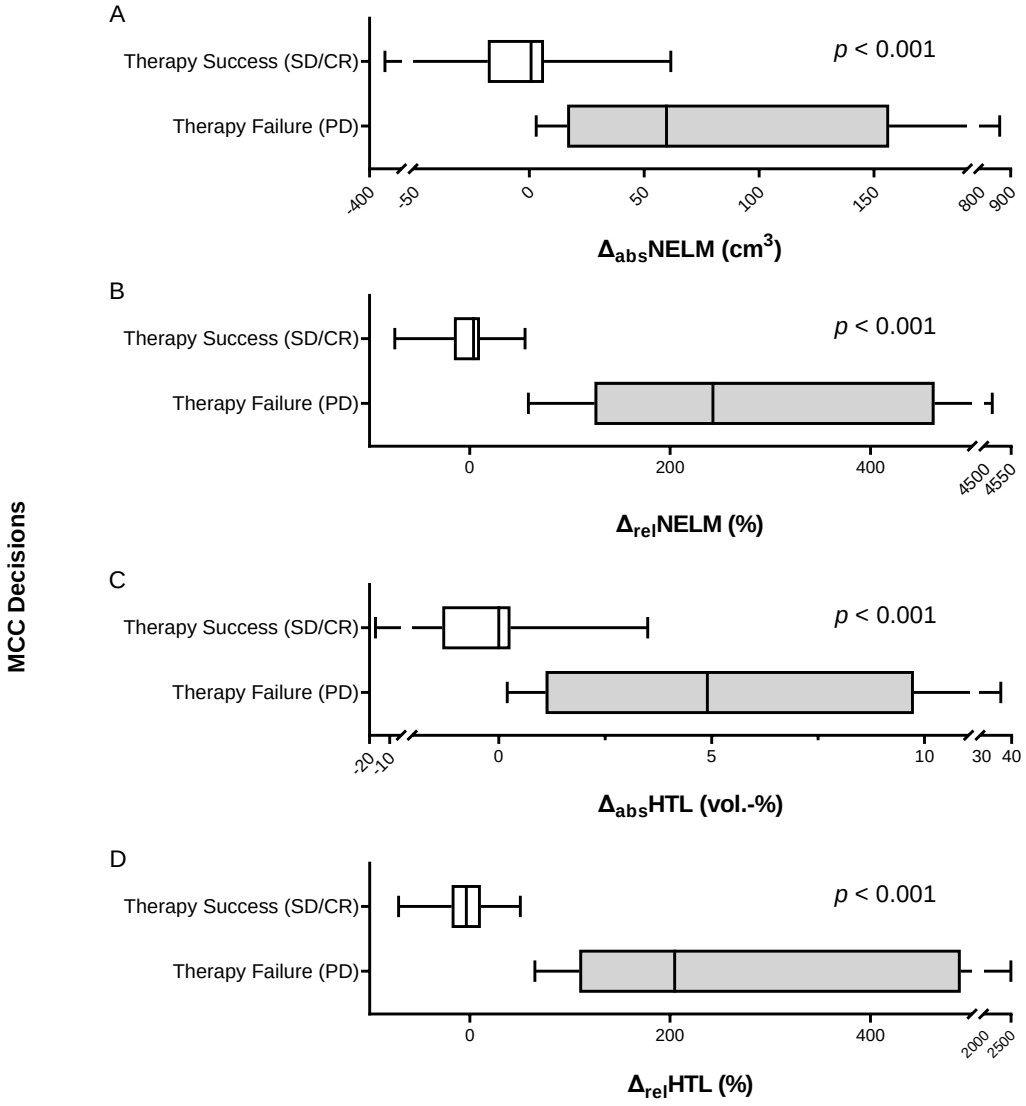


Figure 13. Correlation of MCC board decisions with a relative and absolute change of NELM and NELM proportion. Data are expressed using median and interquartile range (IQR); p-values (Mann-Whitney-U). A: Δ_{abs}NELM; B: Δ_{rel}NELM; C: Δ_{abs}HTL; D: Δ_{rel}HTL. PR: partial response; SD: stable disease; PD: progressive disease; MCC: multidisciplinary cancer conference; Δ_{abs}NELM: absolute NELM volume change; Δ_{abs}HTL: absolute HTL change; Δ_{rel}NELM: relative NELM volume change; Δ_{rel}HTL: relative HTL change.

The case-wise analysis of the 33 MCC patients is shown in Table 5. In patient NO.8 (ID-0008), for example, the output of the AI model at baseline was 1597.5 cm^3 and 137.5 cm^3 for the volumes of liver and NELMs respectively, with a calculated $HTL_{baseline}$ of 9.4%. In contrast, at follow-up, the volume of the liver and NELMs increased to 1977.1 cm^3 and 299.1 cm^3, respectively, which resulted in an increase in $HTL_{follow\text{-}up}$ to 17.8%. Thereby, we obtained an absolute volume change of 161.7 cm^3 and 8.4% for NELMs and HTL, respectively, while the relative volume change was 117.6% for NELMs and 89.3% for HTL.

Table 5. Case-by-case results of the AI segmentation results and response variables (MCC cohort).

	Baseline			**Follow-up**			**Response**				
ID	**Liver Volume (cm^3)**	**NELM Volume (cm^3)**	**HTL (vol.-%)**	**Liver Volume (cm^3)**	**NELM Volume (cm^3)**	**HTL (vol.-%)**	**MCC**	**Δ_{abs}NELM (cm^3)**	**Δ_{abs}HTL (vol.-%)**	**Δ_{rel}NELM (%)**	**Δ_{rel}HTL (%)**
0001	1649.3	100.5	6.5	1796.0	232.7	14.9	PD	132.2	8.4	131.6	129.5
0002	1582.2	69.5	4.6	1783.7	242.4	15.7	PD	172.9	11.1	248.6	242.1
0003	1516.3	23.5	1.6	1766.8	160.7	10.0	PD	137.2	8.4	584.3	536.1
0004	1476.7	2.5	0.2	1487.8	2.0	0.1	SD	-0.5	-0.0	-20.2	-20.8
0005	1304.2	9.5	0.7	1350.2	10.5	0.8	SD	0.9	0.1	9.8	6.1
0006	1247.6	87.3	7.5	1656.0	308.8	22.9	PD	221.5	15.4	253.5	204.5
0007	1092.4	0.6	0.1	1111.6	7.0	0.6	PD	6.4	0.6	1080.0	1066.3
0008	**1597.5**	**137.5**	**9.4**	**1977.1**	**299.1**	**17.8**	**PD**	**161.7**	**8.4**	**117.6**	**89.3**
0009	1474.5	21.2	1.5	1567.0	22.1	1.4	SD	0.9	-0.0	4.2	-2.0
0010	1579.1	3.3	0.2	1610.6	6.3	0.4	PD	3.0	0.2	91.9	88.5
0011	2067.7	148.0	7.7	2167.6	124.1	6.1	SD	-23.9	-1.6	-16.1	-21.2
0012	1959.7	46.5	2.4	1689.6	11.8	0.7	PR	-34.8	-1.7	-74.7	-71.1
0013	1695.2	111.4	7.0	1817.3	173.0	10.5	SD	61.6	3.5	55.3	49.5
0014	1332.6	19.2	1.5	3216.9	883.6	37.9	PD	864.5	36.4	4513.6	2497.2
0015	2209.1	9.4	0.4	2236.5	47.0	2.2	PD	37.6	1.7	398.2	400.5
0016	2326.4	13.0	0.6	2421.6	12.4	0.5	SD	-0.6	-0.1	-4.5	-8.3
0017	969.3	12.1	1.3	972.4	11.6	1.2	SD	-0.6	-0.1	-4.5	-4.9

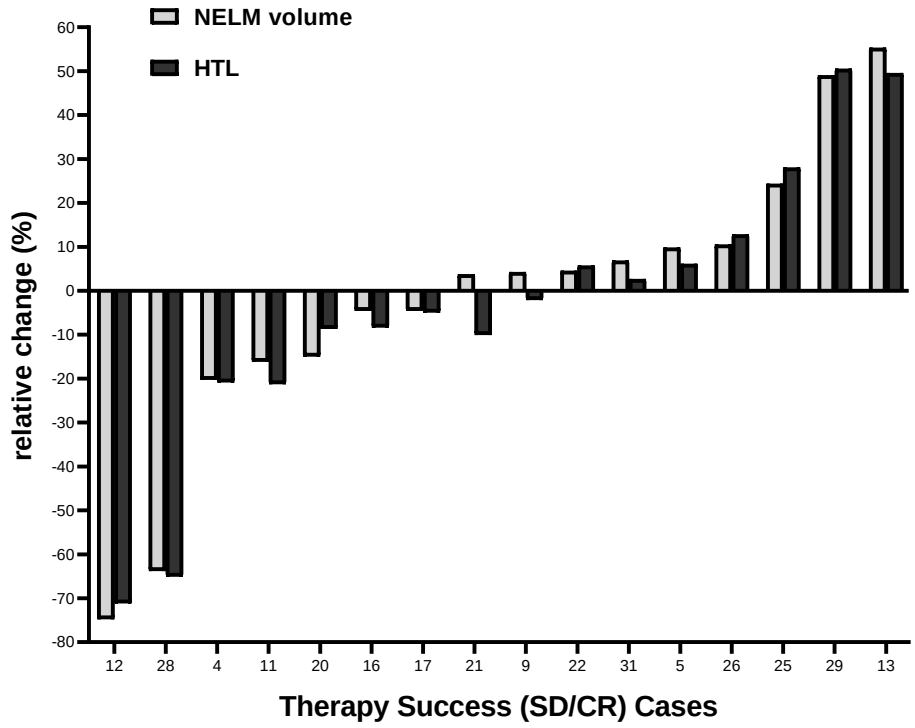
0018	2523.2	17.7	0.7	2364.0	43.3	1.9	PD	25.5	1.2	144.2	163.6
0019	1580.4	54.8	3.6	1552.8	86.9	5.9	PD	32.1	2.3	58.5	65.0
0020	1703.4	244.9	16.8	1572.5	209.2	15.3	SD	-35.7	-1.5	-14.9	-8.6
0021	1575.0	114.9	7.9	1801.2	119.2	7.1	SD	4.2	-0.8	3.7	-10.0
0022	1082.3	14.0	1.3	1071.6	14.6	1.4	SD	0.6	0.1	4.5	5.7
0023	2016.6	133.1	7.1	2303.4	264.3	13.0	PD	131.1	5.9	98.5	83.4
0024	1788.5	4.7	0.3	1639.5	22.9	1.4	PD	18.3	1.2	393.3	444.3
0025	2043.5	122.5	6.4	2018.2	152.3	8.2	SD	29.8	1.8	24.3	28.0
0026	2416.2	180.8	8.1	2389.9	199.9	9.1	SD	19.1	1.0	10.5	12.8
0027	1339.9	11.3	0.9	1297.2	71.0	5.8	PD	59.7	4.9	529.6	582.1
0028	2653.8	549.0	26.1	2386.0	199.4	9.1	PR	-349.6	-17.0	-63.7	-65.0
0029	1477.6	7.2	0.5	1466.5	10.8	0.7	SD	3.5	0.3	49.0	50.5
0030	2054.2	11.2	0.6	1716.8	25.9	1.5	PD	14.7	1.0	131.5	179.8
0031	1689.3	104.4	6.6	1761.0	111.5	6.8	SD	7.1	0.2	6.8	2.6
0032	1207.2	62.4	5.5	1322.4	214.0	19.3	PD	151.5	13.9	242.7	253.9
0033	1146.0	2.1	0.2	1349.4	6.7	0.5	PD	4.6	0.3	222.3	174.6

PR: partial response; SD: stable disease; PD: progressive disease; MCC: multidisciplinary cancer conference; Δ_{abs}NELM: absolute NELM volume change; Δ_{abs}HTL: absolute HTL change; Δ_{rel}NELM: relative NELM volume change; Δ_{rel}HTL: relative HTL change.
Table adapted from Fehrenbach et al. (142).

The case-wise illustration of the relative volume change variables (Figure 14) showed that the model correctly detected the response trend based on Δ_{rel}NELM and Δ_{rel}HTL in all 33 MCC patients, and especially detected the increased NELM volume of all 17 patients with therapeutic failure and the decreased NELM volume of all 2 patients with PR.

In these therapeutic failure patients, the increase in Δ_{abs}NELM ranged from +3.0 cm^3 to +864.5 cm^3 and Δ_{abs}HTL ranged from +0.2 vol.-% to +36.4 vol.-%. The increase in Δ_{rel}NELM ranged from +58.5% to +4513.6% and in Δ_{rel}HTL from +65.0% to +2497.2%. In patients with therapeutic success, the Δ_{abs}NELM ranged from −349.6 cm^3 to −34.8 cm^3 (in PR) and −35.7 cm^3 to +61.6 cm^3 (in SD) and the Δ_{abs}HTL from −17.0 vol.-% to −1.7 vol.-% (in PR) and −1.6 vol.-% to +3.5 vol.-% (in SD). The Δ_{rel}NELM varied from −74.7% to −63.7% (in PR) and from −20.2% to 55.3%

(in SD), and the Δ_{rel}HTL varied from −71.1% to −65.0% (in PR) and from −21.2% to +50.5% (in SD) (data shown in Table 5 and Figure 14).



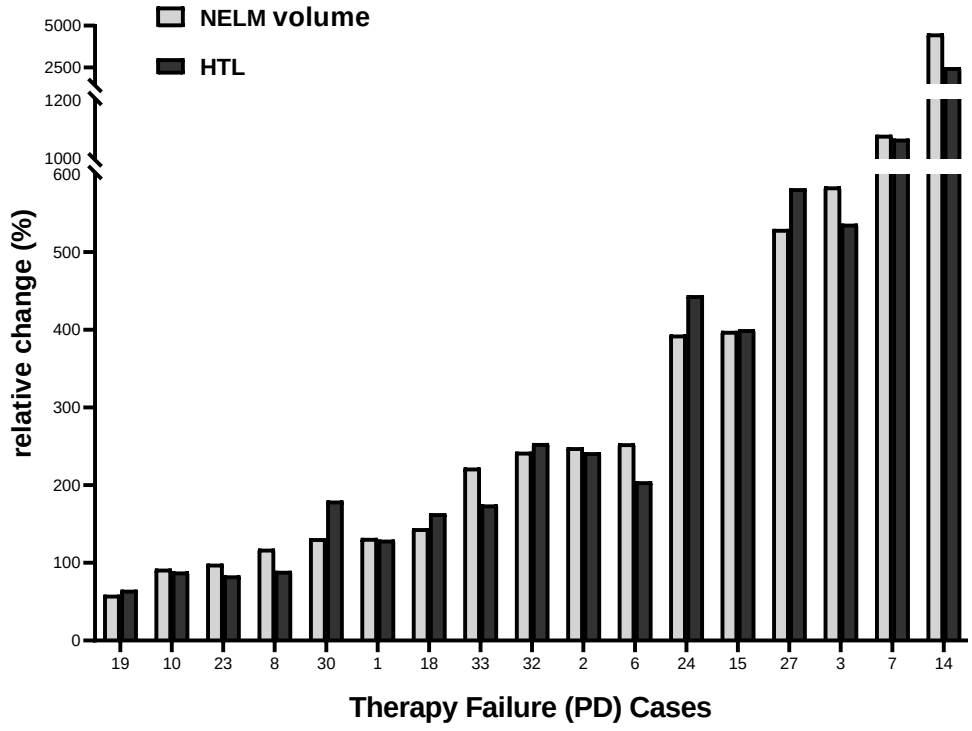


Figure 14. Case-by-case tabulation of relative volume changes (Δ_{rel}NELM and Δ_{rel}HTL) from baseline to follow-up (MCC cohort). The top bar shows the distribution of patients with treatment success and the bottom bar shows the distribution of patients with treatment failure. The grey bars indicate Δ_{rel}NELM and the black bars indicate Δ_{rel}HTL. PR: partial response; SD: stable disease; PD: progressive disease.

3.3.2 Calculating a cut-off value for progressive disease

In order to estimate the optimal cut-off values for progressive disease, ROC curve analysis based on MCC decisions were performed and the Youden indexes were calculated. The relative changes (Δ_{rel}NELM and Δ_{rel}HTL) yielded a receiver-AUC of 1.000 ($p<0.001$) (green dotted line and blue-green dotted line for Δ_{rel}NELM and Δ_{rel}HTL, respectively). The absolute volumetric measures yielded an AUC of 0.908 for Δ_{abs}NELM (blue dotted line) and of 0.925 for Δ_{abs}HTL (red dotted line) ($p<0.001$). When calculating the optimal cut-off point for Δ_{rel}NELM using the Youden index (1.000; 100% sensitivity and 100% specificity), the optimal cut-off-point was +56.88%. For Δ_{rel}HTL, the optimal cut-off was +57.73% (shown in Figure 15).

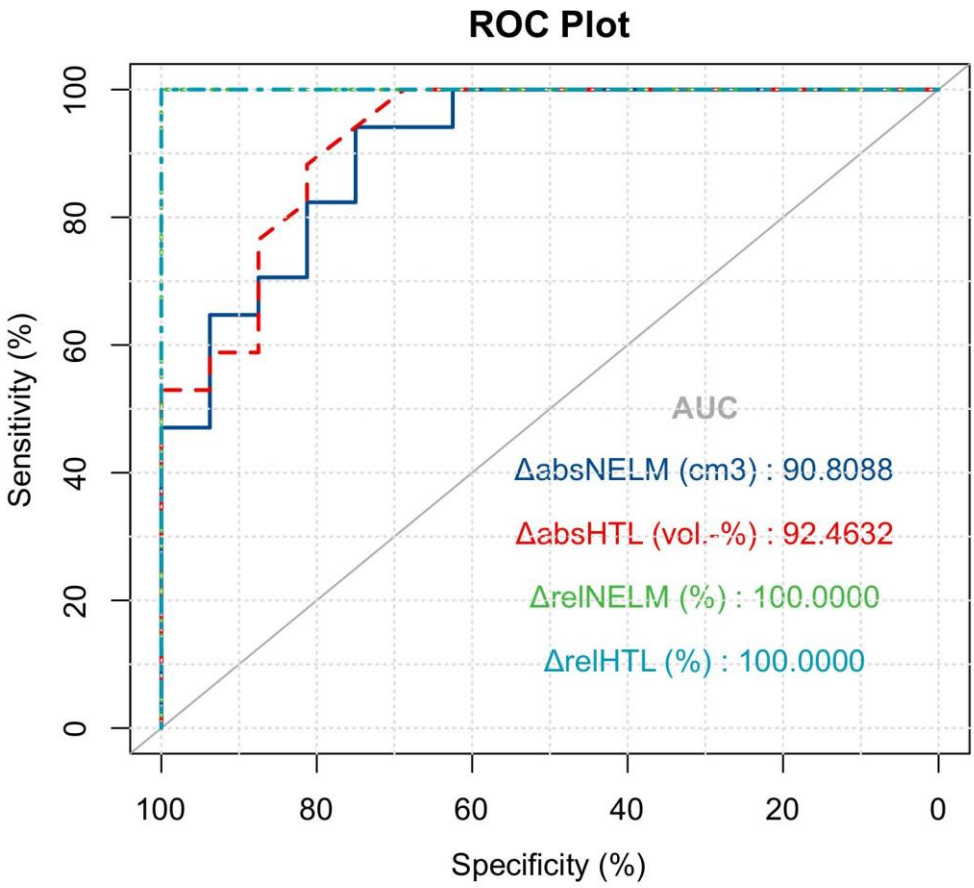


Figure 15. Receiver-Operator-Curveof the absolute and relative change in NELM and %NELM compared to the MCC decisions. AUC Δ_{abs}NELM: 0.908 (blue dotted line); AUC Δ_{abs}HTL: 0.925 (red dotted line); AUC Δ_{rel}NELM: 1.000 (green dotted line); AUC Δ_{rel}HTL: 1.000 (blue-green dotted line); $p < 0.001$.

3.4 3D and 2D visualization of the segmentation results from the AI model

3.4.1 3D visualization of the segmentation

The 3D segmentation of the liver and NELMs from the AI model are exemplarily illustrated in Figure 16 (taking patients ID-0008 and ID-00011 as examples).

In patient ID-00011, identified as having an SD response based on MCC, the absolute NELM volume and HTL changed from 148.02 cm^3 and 7.71 vol.-% to 124.12 cm^3 and 6.07 vol.-%, respectively, resulting in a Δ_{rel}NELM of -16.14% and a Δ_{rel}HTL of -21.23% (Figure 16 A, B).

In contrast, as an example of treatment failure, patient ID-0008 showed an increase in both absolute NELM volume and HTL from 137.48 cm^3 and 9.42 vol.-% at the baseline to 299.14 cm^3 and 17.83 vol.-% at the follow-up, respectively, resulting in a +117.58% increase in Δ_{rel}NELM and +89.32% increase in Δ_{rel}HTL (Figure 16 C, D).

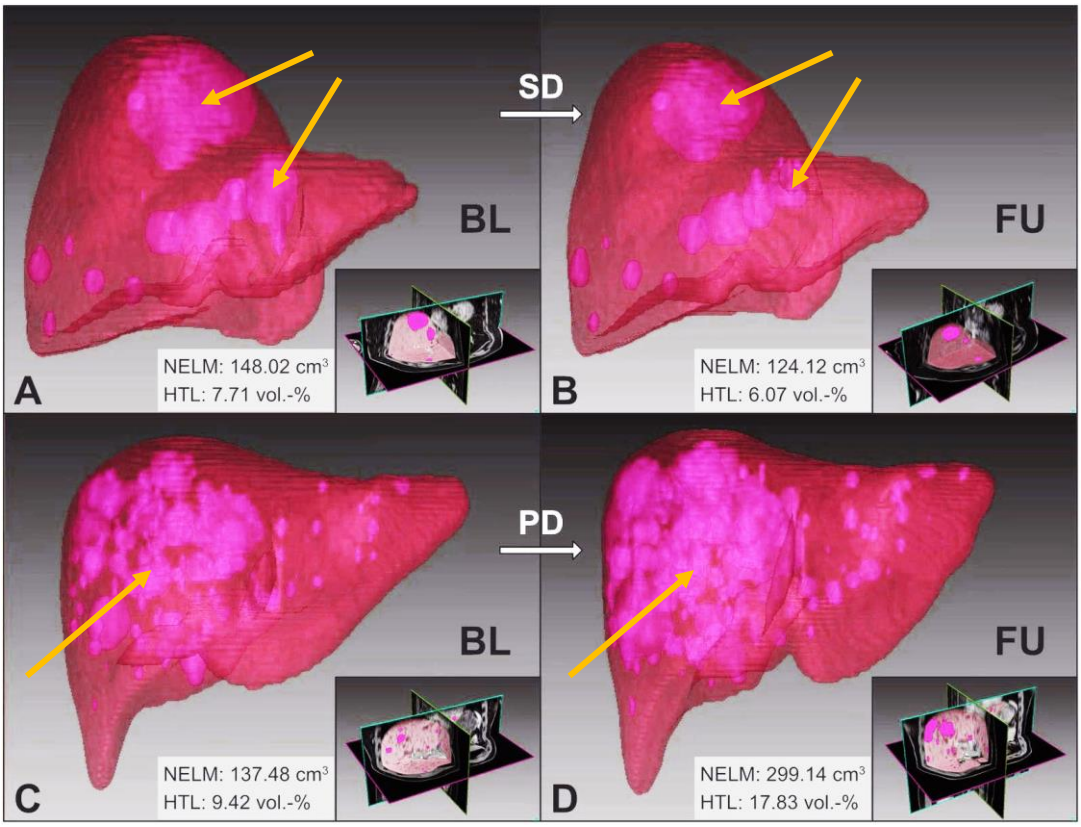


Figure 16. nnU-Net 3D segmentation of liver and NELMs. Top: Exemplary case of therapy success with stable NELMs in baseline (A) and follow-up-imaging (B); Bottom: Exemplary case of therapy failure with progressive NELMs from baseline (C) to follow-up-imaging (D); Yellow arrows indicated NELMs. SD: stable disease; PD: progressive disease; BL: baseline; FU: follow-up; NELM: neuroendocrine liver metastases; HTL: hepatic tumor load.

3.4.2 2D visualization of segmentations

The internal 2D versions of the segmentation results of these two cases are shown in Figure 17. Three levels (30%, 50%, 80%) of MR scans from the baseline (Figure 17 a, b, c, d, e, f, g) and follow-up (Figure 17 A, B, C, D, E, F, G) were chosen to present the results. The red lines showed

the segmentation results of the liver, while the green lines showed the segmentation results of the NELMs.

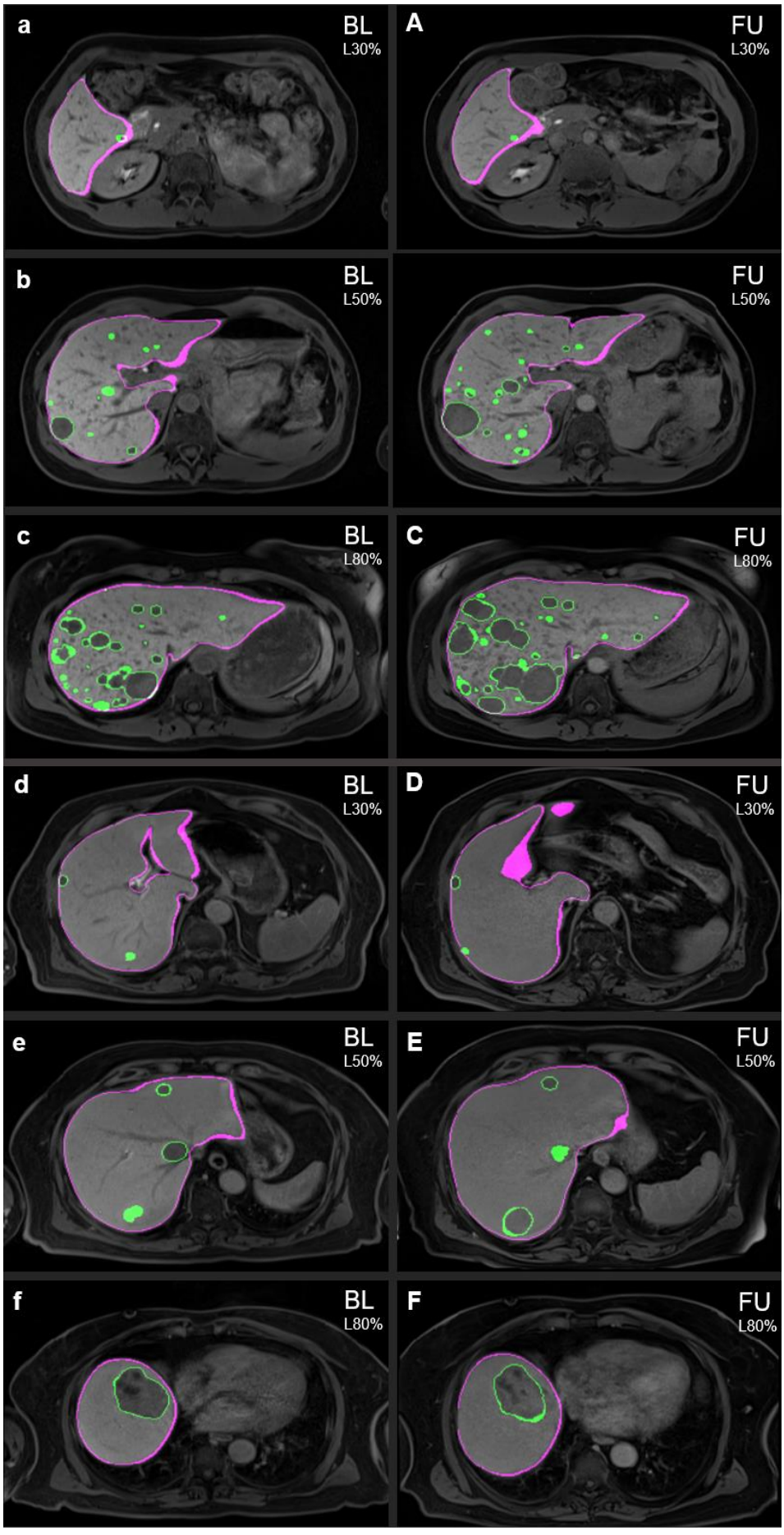

a
BL
L30%
A
FU
L30%
b
BL
L50%
FU
L50%
c
BL
L80%
C
FU
L80%
d
BL
L30%
D
FU
L30%
e
BL
L50%
E
FU
L50%
f
BL
L80%
F
FU
L80%

Figure 17. 2D views of the segmentation of the AI model for patients ID-0008 (a, b, c and A, B, C) and ID-00011 (d, e, f and D, E, F). a, b, c, d, e, f are images of the baseline and A, B, C, D, E, F are images of the follow-up. L30%, L50%, L80% are the scan slices at 30%, 50%, 80% of the MR scan respectively. The red lines showed the segmentation results of the liver, while the green lines showed the segmentation results of the NELMs. BL: baseline, FU: follow-up.

3.5 Comparison of 3D Volumetric Analysis between Hepatobiliary and Diffusion-Weighted Imaging

The results of the manual segmentation in HBP and DWI sequences are summarised in Table 6. There were no significant differences in Δ_{abs}NELM, Δ_{rel}NELM, Δ_{abs}HTL, and Δ_{rel}HTL in measurements for neither DWI nor the HBP sequence ($p = 0.072$ to 0.719). However, when comparing the manually segmented volumes for NELM, liver volume and HTL significant differences between HBP and DWI could be found (all $p < 0.01$).

When assessing the correlations, NELM (rs: 0.981; $p < 0.001$), livers (rs: 0.966; $p < 0.001$) and HTL (rs: 0.956, $p < 0.001$), as well as Δ_{abs}NELM (rs: 0.919; $p < 0.001$), Δ_{rel}NELM (rs: 0.960; $p < 0.001$), Δ_{abs}HTL (rs: 0.883, $p < 0.001$) and Δ_{rel}HTL (rs: 0.952; $p < 0.001$) between HBP and DWI sequences are highly correlated [data not shown].

Table 6. Comparison of 3D Volumetric Analysis between Hepatobiliary and Diffusion-Weighted Imaging.

Variable	HBP	DWI	Significance (p)
NELM volume (cm^3)	63.24 (12.12–174.23)	76.28 (12.61–182.48)	**0.002**
Liver volume (cm^3)	1659.28 (1387.73–2052.00)	1595.00 (1324.17–1977.54)	**< 0.001**
HTL (vol %)	4.05 (0.76–9.23)	5.45 (0.88–11.49)	**< 0.001**
Δ_{abs}NELM (cm^3)	19.57 (17.27–132.52)	30.06 (18.91–142.13)	0.072
Δ_{rel}NELM (%)	107.76 (5.28–245.04)	78.35 (11.22–221.21)	0.719
Δ_{abs}HTL (vol %)	1.20 (−0.01–8.87)	1.25 (0.10–10.47)	0.151
Δ_{rel}HTL (%)	111.36 (−0.36–254.49)	67.76 (4.20–198.88)	0.151

Values are displayed as median and interquartile ranges. P-values are based on the sign test. HBP: hepatobiliary phase; DWI: diffusion-weighted imaging; NELM: neuroendocrine liver metastasis; HTL: hepatic tumor load; Δ_{abs}NELM: absolute NELM volume change; Δ_{abs}HTL: absolute HTL change; Δ_{rel}NELM: relative NELM volume change; Δ_{rel}HTL: relative HTL change.

Adapted from Fehrenbach et al. (142)

The example in Figure 18 shows the comparison of manual segmentation in HBP and DWI sequences. Figures 18A, B, C and D, E, F show the original image, the NELMs segmentation, and the overlapping images of the NELMs and the liver segmentation in one of the same slices in the HBP sequence and the DWI sequence, respectively. The segmentation of the two sequences was performed completely independently of each other. The white arrows in Figure 18B point out two tiny NELMs found in HBP but not observed in the same slice of DWI; the white arrows in Figure 18C, F point out the difference in liver segmentation results between the two different sequences.

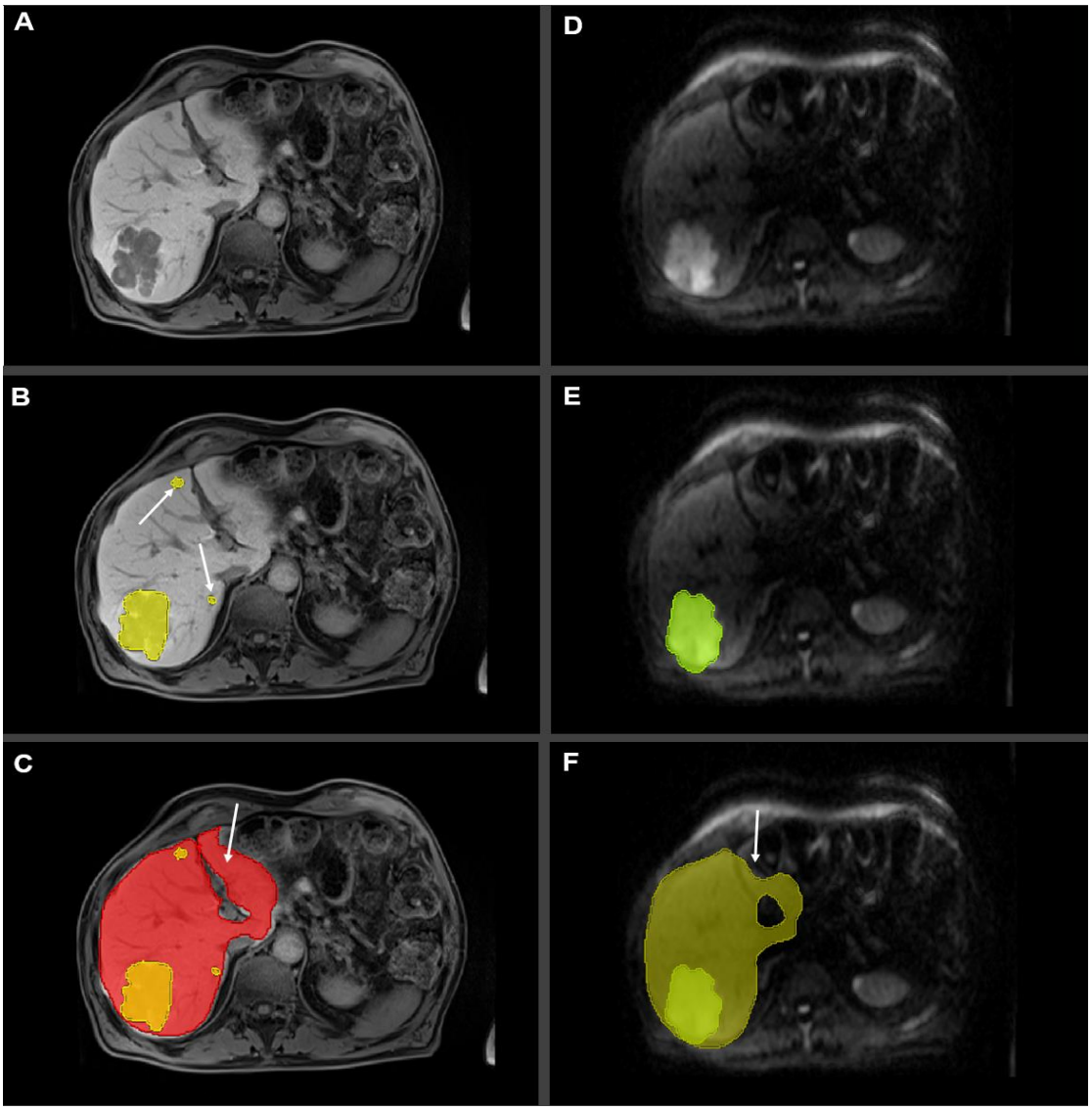


Figure 18. Comparison of manual segmentations between HBP and DWI sequences. (A, B, C) original image, liver segmentation, and NELMs segmentations in HBP sequences, respectively; (D, E, F) original image, liver segmentation, and NELMs segmentations in DWI sequences, respectively; the white arrows indicate the differences in manual segmentation results between HBP and DWI sequences: (B) two tiny NELMs were found in HBP but not observed in DWI; (C, F) the contour segmentation of the liver is different. Manual segmentation was performed on the DWI images with b=400 s/mm^2.

4. Discussion

Previous epidemiological studies have reported an increase in the incidence of NENs worldwide (4,7–17). Meanwhile, the clinical course of NEN patients continues to be prolonged with advances in treatment strategies. These have led to an explosive increase in the imaging workload of patients with NELMs. The importance of imaging technology is not only present in the diagnosis, but also in the follow-up, especially in the response to the treatment. As the response evaluation guidelines of all solid tumors, the RECIST 1.1 guidelines in its newest version published in 2009 represent the most commonly used response evaluation criteria. The diameter change of two target lesions per organ is used as the primary response variable. However, it is considered by numerous scientists to be too simplistic and insufficient to accurately assess changes in disease (143–149). More problematically, current standards are difficult to apply for the diffuse pattern NELMs, which can account for more than 60% of the cases (51).

Additionally, the accurate measurement of HTL is an important prognostic factor for effective treatment. In an ideal world, the exact HTL would be calculated to guide the treatment strategy accordingly. In daily practice, however, HTL measurement is too time-consuming and most of the time only visual assessments of HTL are used, leading to bias, and potentially affecting the patient's optimal treatment timepoint.

The development of deep learning in the field of image recognition, especially with the proposal of nnU-Net in 2018, provides a practical model for 3D segmentation that could be a solution to these problems. Therefore, the overall aim of this thesis was to train a nnU-Net for Gd-EOB MRI segmentation to obtain a high-precision AI algorithm for 3D quantification of HTL and NELMs volumes, providing a useful tool for clinical decision making.

The study consists of four parts: In the first part, we established a cohort of NELM patients who underwent Gd-EOB enhanced MRI (1.5 T) between January 2015 and August 2018 for the training of the AI model. In the second part, we selected 66 Gd-EOB enhanced MRI scans (1.5 or 3.0 T) from an additional 33 NET patients at the MCC between January 2019 and January 2020 as an external validation cohort to evaluate the robustness of the AI model. In the third section, the model’s clinical utility was evaluated. We compared the automatically measured 3D volumes of

the AI model in the MCC cohort with the response to the MCC decision and establish an optimal threshold for disease progression. In the fourth part, we compared the manual segmentation in HBP and DWI sequences.

The data in this study uses AI for segmenting the NELMs automatically, which not only provides a viable tool to replace the one-dimensional (1D) diameter change with a full 3D volume change as a treatment response criterion but could also help to establish a cut-off value for disease progression in volumetric tumor assessment.

4.1 Methodology

A recent systematic review of imaging practices with AI for PNENs concluded that none of the currently published methods was considered to have a low risk of bias in all areas (143). To minimize the risk of selective bias, we collected all MRI images from patients who were diagnosed with NENs and underwent Gd-EOB-enhanced liver MRI between January 2015 and August 2018 and blinded the staff involved in performing the manual segmentation. The reviewers were blinded to all information, including whether the patient had NELM, patient history, laboratory results, and pathological findings and treatment. To date, in contrast to some studies (144,150–152), we did not limit the number and size of NELMs nor the treatment measures, but included all Gd-EOB-enhanced liver MRI during the study time period in the manually segmented dataset, in order to train the model as realistically as possible.

4.2 Patients cohorts

We collected all patients with NENs who underwent Gd-EOB enhanced MRI (1.5 T, Siemens Aera) between January 2015 and August 2018 as the AI develop cohort, regardless of whether the NELMs had surgery or not. This allows for a more realistic picture of NELMs that may be encountered in clinical work. 120 scans were excluded due to lack of evidence of NELM or non-standard scan protocols, the rest 278 scans were divided into the training (80%) and internal testing (20%) database.

In order to train a generic AI model, both internal and external validations are required (129,132). Typically, internal validation comes from the same cohort as the training data, with both cross-validation and split-sample validation, of which the latter is considered to be superior (129). Hence,

in our study, we chose a random assignment in the ratio of 2:8, that is, 20% of the AI dev cohort as the internal validation dataset and 80% as the training dataset.

The external validation is defined as a dataset originating from the different time periods, cohorts, protocols, or clinics (129), which is ultimately used to determine the generic performance of the algorithm. However, a systematic review about the performance of AI algorithms used for the diagnostic analysis of medical images in 2019 showed that only 6% of the algorithms were externally validated (153). Thereby, we collected an additional 66 scans from MCC between January 2019 and January 2020 who received a Gd-EOB MRI (1.5T or 3.0T, different scanners and imaging protocols than in the training cohort) both at baseline and follow-up as the external validation database (MCC cohort), to avoid measuring an overfitting bias when assessing the AI performance. There were significant differences in primary sites and SR status (P=0.0003 and 0.047, respectively), from that respect the MCC cohort is not similar to the AI cohort, further adding to the robustness of the external validation. Combined with the different time points and different MRI units (1.5 T/ 3.0T versus 1.5 T), the MCC cohort was considered suitable as an external validation of the model.

4.3 Performance of the model

Matthews correlation coefficient (ϕ) was used in both internal and external validation datasets to demonstrate the performance of the model. In the internal validation dataset, ϕ were 0.76 (IQR: 0.68–0.83) in the segmentation of NELMs and 0.95 (IQR: 0.95–0.96) in liver segmentation (Figure 9). It is hard to compare the performance of our AI model with other data sets because MRI sequences and evaluation metrics are different. Rough statistics of the recently proposed automatic liver and/or liver tumor MRI segmentation algorithms show that the reported DC for liver and hepatic lesions ranged from 0.87 to 0.91 and 0.68 to 0.84, respectively (150,154–157).

In the external validation dataset of our study, the model achieved a median ϕ of 0.86 (IQR: 0.81–0.91) in the segmentation of NELM and of 0.96 (IQR: 0.95–0.96) in liver segmentation. However, there are few studies on the performance of liver and/or liver tumor MRI segmentation algorithms in external validation datasets.

In the internal validation, there are three cases (orange circled in Figure 10) in which the segmentation performance is low due to small or atypical tumor volumes. Therefore, we divided the patients of the internal validation dataset into two groups according to the median target NELM volume (16.17 cm^3): low volume and high volume. The comparison shows that the Matthews correlation coefficient was significantly higher in the high-volume group than in the low-volume group (P = 0.0025, in Figure 11).

This, however, does not necessarily suggest that patients with low NELM volumes are not suitable for this model, because low Matthews correlation coefficient scores may also be due to inaccurate target volume. During manual segmentation, we found that the accuracy of manual segmentation of small-sized tumors, especially cornified tumors which were considered to be the target lesion, was reduced.

In figure 12, which shows the comparison between our well-trained AI segmentation and manual segmentation in a patient from the MCC cohort with a small tumor volume, we can see that there are only two small tumors at this slice, and the AI segmentation (Figure 12 F) results in more accurate edges than the manual segmentation (Figure 12 E). Similarly, AI segmentation of the liver (Figure 12 C) was slightly better than manual segmentation (Figure 12 B), as demonstrated by the margins and the region of the hepatic portal vein that should be excluded. From this, we can see that our AI model performs well at least for segmentation on this small tumor volume case.

The segmentation performance of the AI model on entire liver is stable and accurate in both internal and external validation datasets (ϕ: 0.95 and 0.96 respectively).

4.4 AI-based NELM Segmentation and Correlation with MCC decision

The patients of the MCC cohort have been determined as having different treatment responses (PR, SD, PD) according to RECIST 1.1 criteria in the MCC meeting. Since the number of cases with PR was low (n = 2 cases) and the next treatment option was similar as for patients with SD; PR and SD were defined as the treatment success group (n=16 cases), while PD was defined as the treatment failure group (n=17 cases). The relative and absolute changes in 3D volume output by AI automatically reflected significant differences between the two groups: the treatment failure group showed significantly higher Δ_{abs}NELM (cm^2), Δ_{abs}HTL (%), Δ_{rel}NELM (%), and Δ_{rel}HTL

(%) than the treatment success group (all $p < 0.001$, Figure 13). This reflects the consistency of our AI model with MCC decision-making.

To compare the accuracy of 3D volume and 1D diameter for response determination, we then analyzed case-by-case the correlation of changes in tumor volume and HTL with MCC decisions. The case-wise analysis showed that the model correctly detected trends in Δ_{rel}NELM and Δ_{rel}HTL-based responses in all 33 patients (Figure 14). However, individual data trends in Δ_{abs}NELM and Δ_{abs}HTL were disrupted (Table 4), for example, the Δ_{abs}NELM and Δ_{abs}HTL of case-ID0010, which was determined as PD, were lower than for cases ID0029 and ID0021 (Δ_{abs}NELM :3 cm^2 vs. 3.5 cm^2 and 4.2 cm^2, Δ_{abs}HTL: 0.2% vs. 0.3% and -0.8%, respectively), which were determined as SD. This is explainable due to the fact that the MCC decision was based on a 20% increase in diameter, which is a relative change rather than an absolute change. Moreover, in clinical work, the need for follow-up of patients with NELMs is much more than a single volumetric measurement, that is, changes in tumor load-related indicators are more important.

Since our model perfectly classifies the treatment outcomes of patients in the MCC cohort, it demonstrates the ability of our model to upgrade the original RECIST1.1 criteria from a 1D to a 3D metric. Therefore, the cutoff values for determining disease progression need to be redefined.

ROC curves analysis based on MCC decision were performed in Δ_{rel}NELM, Δ_{rel}HTL, Δ_{abs}NELM, and Δ_{abs}HTL (Figure 15). The results showed that high AUC values were achieved for all indicators (Figure 15, AUC: Δ_{abs}NELM 0.908, Δ_{abs}HTL 0.925, Δ_{rel}NELM 1.000, Δ_{rel}HTL 1.000, all $p < 0.001$), whereas, a rough comparison with recently proposed automated 3D liver tumor MRI segmentation algorithms, the reported AUCs ranged from 0.78 to 0.94 (150,154).

Due to indicators of relative change being more meaningful in clinical practice, we searched for the optimal cut-off value in Δ_{rel}NELM and Δ_{rel}HTL to determine disease progression. The highest Youden index (1.000; 100% sensitivity and 100% specificity) was found at the cut-off value +56.88% and +57.73% in Δ_{rel}NELM and Δ_{rel}HTL, respectively. Assuming the patient has only one spherical target lesion, the Δ_{rel}NELM should increase by approximately 73% according to the RECIST 1.1 criteria (sum of the diameter of the target lesions increases by at least 20%), whereas our suggested cut-off value is 56.88%, which modified this indicator by approximately 16%.

4.5 3D and 2D versions of the segmentation results from the AI model

Our model is based on a 3D volume segmentation algorithm, and two examples of its 3D segmentation results are presented in Figure 16. We can see that the 3D volumes of NELMs are irregular, which on the other hand proves the current standard RECIST 1.1, measured only by the maximum diameter in one slice, could be insufficient. It is even more challenging when two relevant lesions are located close to each other, in that case, they are likely to merge and become difficult to distinguish, which introduced another inaccuracy when using 1D diameter to define the response.

After obtaining the 3D volume segmentation results, we still want to confirm the AI segmentation in the internal 2D plane of these two cases. In figure 17 we can see that the liver is well segmented, with larger NELMs segmented better than small NELMs. At the 50% (Figure 17 b, B) and 80% (Figure 17 c, C) levels of case ID-0008, not all the diffuse NELMs were segmented, meanwhile, at the 50% level of case ID-00011 (Figure 17 e, E), the hepatic vein area was mistakenly considered to be a NELM.

Small NELMs are indeed difficult to distinguish from a single cross-section of the hepatic blood vessel. This might be because both hepatic tumor and vessel are non-enhancing in the HBP sequence, and they are very similar in morphology. Even manually, other methods must assist in the identification, e.g., by anterior-posterior contiguous scrolling: the NELM remains fixed in position, while the blood vessels show movement trajectories on continuous levels due to their tubular structure.

However, this degree of mis-segmentation is probably acceptable because the omission of these partially small tumors has minimal impact on the overall tumor volume, and it has an even lower impact on the relevant changes. Therefore, the accuracy of our model for the classification of therapy response evaluation remains high.

4.6 Comparison of 3D Volumetric Analysis between Hepatobiliary and Diffusion-Weighted Imaging

DWI has been highly suggested for detecting NELMs for its advantages as a non-contrast imaging technique (79,80). Previous studies also suggested that the combination of DWI and HBP

sequences with Gd-EOB is more sensitive than each of them alone to detect hepatic lesions (58). To date, most studies of DWI discussed their ability to detect lesions, and only a few articles used DWI sequences as the basis for segmentation (158–162). Therefore, to explore its suitability for 3D segmentation, the DWI sequences of the MCC cohort were also manually segmented in our study to compare with the manual segmentation of HBP sequences.

Compared to the HBP in Gd-EOB-MRI the boundaries of NELMs and liver are not that clearly depicted due to the low resolution of the DWI imaging(163), whereas identification of NELM is excellent in DWI. This is slightly different from the results of another study (162), which claimed that DW-MRI was superior to non-enhanced MRI for obtaining accurate segmentation of the liver margins.

Figure 18 demonstrates the results of the manual segmentation of the HBP sequence compared to the DWI sequence on the same slice, from which it can be seen that the HBP sequence (left side) has an obviously higher resolution than the DWI sequence (right side). In addition, there are two small NELMs in figure 18B which are not discernible in the corresponding DWI sequence (Figure 18E), and the results of liver segmentation in HBP sequences were also more accurate compared to the DWI sequences (Figure 18C vs. 18Figure 18F, white arrows).

When comparing the overall results of HBP and DWI manual segmentation in the MCC cohort (Table 6), it showed that the exact volume of liver, NELM, and HTL in DWI sequences are significantly different from the HBP sequence (all $p < 0.01$). This indicates that the DWI sequence might be not suitable for calculating the exact tumor size and tumor load at only one-time point, and it leads to significant differences in results from the more accurate HBP sequence. However, the absolute/relevant volume changes of liver, NELM, HTL are not different from the HBP sequence. This suggests that DWI sequences at multiple time points can also be used to assess changes in relevant indicators to determine disease progression status. This is of interest for clinical centers where only DWI sequences are available for follow-up, or when the cost of follow-up is considered.

4.7 Strengths and Limitations of the Study

There are several strengths in our study. First of all, to date, our study is the second study of 3D volume segmentation based on GOB-MRI of NELMs (157), but the first to compare manual segmentation comparisons on HBP sequences and DWI sequences. Secondly, manual segmentation is considered to be the most difficult challenge in supervised deep learning (119), especially in heterogeneous and diffuse lesions. However, in this thesis, we manually segmented 398 scans with a considerable number of metastases, which is to date the largest number of manual segmentation studies of focal liver lesions (151,157). We also include a variety of treatment sequelae (e.g., prior liver resection, excessive pretreatment, ablative therapy, or prior intra-arterial therapy), which improves the robustness of our model in preparation for routine clinical use. Thirdly, although HTL is considered an important prognostic imaging indicator (5,6,49), in clinical practice, total HTL is not routinely quantified, and in most cases, tumor load is estimated visually. In contrast, our model not only provides more robust information on lesion distribution, but we further propose a cutoff point for determining disease progression, providing a standardized basis for 3D volume measurement. Finally, our model is not limited to the segmentation of NELM. We choose NELM as the disease model since it has a long tumor-bearing lifetime and reflects well the distinctive features of liver metastases. It is more suitable as a basic disease model than other origins of liver metastases, and the optimized model can be extended to other origins of liver metastases for segmentation tasks and calculation of liver tumor burden, as liver metastases from various primary tumors exhibit the same typical imaging features in HBP sequences (164).

Our study has some limitations as well. First of all, the ground truth for accuracy is based on manual segmentation of liver metastases (119), which could be not perfect due to the sometimes pronounced, even small foci of liver metastases. Second, 3D assessment methods need to be further evaluated in a larger clinical collective, compared directly with 1D measurements, and to assess the impact on clinical endpoints. Third, our study was single-center and retrospective which was considered to have a limited level of evidence (143), so our next step will be to consider further validation in a prospective patient dataset.

5. Conclusion

In conclusion, the AI model presented in this thesis shows the high accuracy of NELM 3D volumes and HTL measurements in HBP sequences on Gd-EOB MRI, providing a usable algorithm for fully automated 3D assessment of neoplastic lesions in the liver.

The model also provides useful information about HTL and NELM volumes and can be used to assist physicians in the assessment of treatment response of hepatic neoplastic lesions, paving the way for precise treatment.

6. References

1. Rindi G, Wiedenmann B. Neuroendocrine neoplasms of the gut and pancreas: new insights. Nat Rev Endocrinol. 2012 Jan;8(1):54–64.
2. Rindi G, Wiedenmann B. Neuroendocrine neoplasia of the gastrointestinal tract revisited: towards precision medicine. Nat Rev Endocrinol. 2020 Oct;16(10):590–607.
3. Rindi G, Inzani F. Neuroendocrine neoplasm update: toward universal nomenclature. Endocr Relat Cancer. 2020 Jun 1;27(6):R211–8.
4. Cives M, Strosberg JR. Gastroenteropancreatic Neuroendocrine Tumors. CA Cancer J Clin. 2018;68(6):471–87.
5. Pavel M, Baudin E, Couvelard A, Krenning E, Öberg K, Steinmüller T, Anlauf M, Wiedenmann B, Salazar R, Barcelona Consensus Conference participants. ENETS Consensus Guidelines for the management of patients with liver and other distant metastases from neuroendocrine neoplasms of foregut, midgut, hindgut, and unknown primary. Neuroendocrinology. 2012;95(2):157–76.
6. Pavel M, O''Toole D, Costa F, Capdevila J, Gross D, Kianmanesh R, Krenning E, Knigge U, Salazar R, Pape U-F, Öberg K, Participants all other VCC. ENETS Consensus Guidelines Update for the Management of Distant Metastatic Disease of Intestinal, Pancreatic, Bronchial Neuroendocrine Neoplasms (NEN) and NEN of Unknown Primary Site. Neuroendocrinology. 2016;103(2):172–85.
7. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, Abdalla EK, Fleming JB, Vauthey J-N, Rashid A, Evans DB. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. J Clin Oncol Off J Am Soc Clin Oncol. 2008 Jun 20;26(18):3063–72.
8. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, Shih T, Yao JC. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. JAMA Oncol. 2017 Oct 1;3(10):1335–42.

9. Hallet J, Law CHL, Cukier M, Saskin R, Liu N, Singh S. Exploring the rising incidence of neuroendocrine tumors: a population-based analysis of epidemiology, metastatic presentation, and outcomes. Cancer. 2015 Feb 15;121(4):589–97.
10. Ellis L, Shale MJ, Coleman MP. Carcinoid tumors of the gastrointestinal tract: trends in incidence in England since 1971. Am J Gastroenterol. 2010 Dec;105(12):2563–9.
11. Scherübl H, Streller B, Stabenow R, Herbst H, Höpfner M, Schwertner C, Steinberg J, Eick J, Ring W, Tiwari K, Zappe SM. Clinically detected gastroenteropancreatic neuroendocrine tumors are on the rise: Epidemiological changes in Germany. World J Gastroenterol WJG. 2013 Dec 21;19(47):9012–9.
12. Boyar Cetinkaya R, Aagnes B, Thiis-Evensen E, Tretli S, Bergestuen DS, Hansen S. Trends in Incidence of Neuroendocrine Neoplasms in Norway: A Report of 16,075 Cases from 1993 through 2010. Neuroendocrinology. 2017;104(1):1–10.
13. Korse CM, Taal BG, van Velthuysen M-LF, Visser O. Incidence and survival of neuroendocrine tumours in the Netherlands according to histological grade: experience of two decades of cancer registry. Eur J Cancer Oxf Engl 1990. 2013 May;49(8):1975–83.
14. Hemminki K, Li X. Incidence trends and risk factors of carcinoid tumors: a nationwide epidemiologic study from Sweden. Cancer. 2001 Oct 15;92(8):2204–10.
15. Landerholm K, Falkmer S, Järhult J. Epidemiology of Small Bowel Carcinoids in a Defined Population. World J Surg. 2010 Jul 1;34(7):1500–5.
16. Ito T, Igarashi H, Nakamura K, Sasano H, Okusaka T, Takano K, Komoto I, Tanaka M, Imamura M, Jensen RT, Takayanagi R, Shimatsu A. Epidemiological trends of pancreatic and gastrointestinal neuroendocrine tumors in Japan: a nationwide survey analysis. J Gastroenterol. 2015 Jan;50(1):58–64.
17. Palepu J, Shrikhande SV, Bhaduri D, Shah RC, Sirohi B, Chhabra V, Dhar P, Sastry R, Sikora S. Trends in diagnosis of gastroenteropancreatic neuroendocrine tumors (GEP-NETs) in India: A report of multicenter data from a web-based registry. Indian J Gastroenterol Off J Indian Soc Gastroenterol. 2017;36(6):445–51.
18. Young K, Lawlor RT, Ragulan C, Patil Y, Mafficini A, Bersani S, Antonello D, Mansfield D, Cingarlini S, Landoni L, Pea A, Luchini C, Piredda L, Kannan N, Nyamundanda G, Morganstein D, Chau I, Wiedenmann B, Milella M, Melcher A, Cunningham D, Starling N,

Scarpa A, Sadanandam A. Immune landscape, evolution, hypoxia-mediated viral mimicry pathways and therapeutic potential in molecular subtypes of pancreatic neuroendocrine tumours. Gut. 2021 Oct;70(10):1904–13.

19. Kennedy A, Bester L, Salem R, Sharma RA, Parks RW, Ruszniewski P. Role of hepatic intra-arterial therapies in metastatic neuroendocrine tumours (NET): guidelines from the NET-Liver-Metastases Consensus Conference. HPB. 2015 Jan 1;17(1):29–37.
20. Plöckinger U, Rindi G, Arnold R, Eriksson B, Krenning EP, Herder WW de, Goede A, Caplin M, Öberg K, Reubi JC, Nilsson O, Fave GD, Ruszniewski P, Ahlman H, Wiedenmann B. Guidelines for the Diagnosis and Treatment of Neuroendocrine Gastrointestinal Tumours. Neuroendocrinology. 2004;80(6):394–424.
21. Perren A, Couvelard A, Scoazec J-Y, Costa F, Borbath I, Delle Fave G, Gorbounova V, Gross D, Grossma A, Jense RT, Kulke M, Oeberg K, Rindi G, Sorbye H, Welin S, Antibes Consensus Conference participants. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: Pathology: Diagnosis and Prognostic Stratification. Neuroendocrinology. 2017;105(3):196–200.
22. Inzani F, Rindi G. Classification of Neuroendocrine Neoplasms. In: Pacak K, Taïeb D, editors. Diagnostic and Therapeutic Nuclear Medicine for Neuroendocrine Tumors [Internet]. Cham: Springer International Publishing; 2017 [cited 2021 May 11]. p. 1–13. (Contemporary Endocrinology). Available from: https://doi.org/10.1007/978-3-319-46038-3_1
23. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, Washington KM, Carneiro F, Cree IA. The 2019 WHO classification of tumours of the digestive system. Histopathology. 2020 Jan;76(2):182–8.
24. Mestier L de, Cros J, Neuzillet C, Hentic O, Egal A, Muller N, Bouché O, Cadiot G, Ruszniewski P, Couvelard A, Hammel P. Digestive System Mixed Neuroendocrine-Non-Neuroendocrine Neoplasms. Neuroendocrinology. 2017;105(4):412–25.
25. Nuñez-Valdovinos B, Carmona-Bayonas A, Jimenez-Fonseca P, Capdevila J, Castaño-Pascual Á, Benavent M, Pi Barrio JJ, Teule A, Alonso V, Custodio A, Marazuela M, Segura Á, Beguiristain A, Llanos M, Martinez Del Prado MP, Diaz-Perez JA, Castellano D, Sevilla I, Lopez C, Alonso T, Garcia-Carbonero R. Neuroendocrine Tumor Heterogeneity Adds

Uncertainty to the World Health Organization 2010 Classification: Real-World Data from the Spanish Tumor Registry (R-GETNE). The Oncologist. 2018 Apr;23(4):422–32.

26. Hauso O, Gustafsson BI, Kidd M, Waldum HL, Drozdov I, Chan AKC, Modlin IM. Neuroendocrine tumor epidemiology. Cancer. 2008;113(10):2655–64.
27. Halfdanarson TR, Rabe KG, Rubin J, Petersen GM. Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival. Ann Oncol Off J Eur Soc Med Oncol. 2008 Oct;19(10):1727–33.
28. Öberg K. Management of functional neuroendocrine tumors of the pancreas. Gland Surg. 2018 Feb;7(1):20–7.
29. Oberg K, Couvelard A, Delle Fave G, Gross D, Grossman A, Jensen RT, Pape U-F, Perren A, Rindi G, Ruszniewski P, Scoazec J-Y, Welin S, Wiedenmann B, Ferone D, Antibes Consensus Conference participants. ENETS Consensus Guidelines for Standard of Care in Neuroendocrine Tumours: Biochemical Markers. Neuroendocrinology. 2017;105(3):201–11.
30. O'Toole D, Grossman A, Gross D, Fave GD, Barkmanova J, O'Connor J, Pape U-F, Plöckinger U. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: Biochemical Markers. Neuroendocrinology. 2009;90(2):194–202.
31. Herrera-Martínez AD, Hofland LJ, Moreno MAG, Castaño JP, Herder WW de, Feelders RA. Neuroendocrine neoplasms: current and potential diagnostic, predictive and prognostic markers. Endocr Relat Cancer. 2019 Mar 1;26(3):R157–79.
32. Hofland J, Zandee WT, de Herder WW. Role of biomarker tests for diagnosis of neuroendocrine tumours. Nat Rev Endocrinol. 2018 Nov;14(11):656–69.
33. Jensen EH, Kvols L, McLoughlin JM, Lewis JM, Alvarado MD, Yeatman T, Malafa M, Shibata D. Biomarkers predict outcomes following cytoreductive surgery for hepatic metastases from functional carcinoid tumors. Ann Surg Oncol. 2007 Feb;14(2):780–5.
34. Khan MS, Caplin ME. The use of biomarkers in neuroendocrine tumours. Frontline Gastroenterol. 2013 Jul 1;4(3):175–81.
35. Frilling A, Modlin IM, Kidd M, Russell C, Breitenstein S, Salem R, Kwekkeboom D, Lau W, Klersy C, Vilgrain V, Davidson B, Siegler M, Caplin M, Solcia E, Schilsky R, Working

Group on Neuroendocrine Liver Metastases. Recommendations for management of patients with neuroendocrine liver metastases. Lancet Oncol. 2014 Jan;15(1):e8-21.

36. Bajetta E, Ferrari L, Martinetti A, Celio L, Procopio G, Artale S, Zilembo N, Di Bartolomeo M, Seregni E, Bombardieri E. Chromogranin A, neuron specific enolase, carcinoembryonic antigen, and hydroxyindole acetic acid evaluation in patients with neuroendocrine tumors. Cancer. 1999 Sep 1;86(5):858–65.
37. Sobol RE, Memoli V, Deftos LJ. Hormone-negative, chromogranin A-positive endocrine tumors. N Engl J Med. 1989 Feb 16;320(7):444–7.
38. O'Connor DT, Pandlan MR, Carlton E, Cervenka JH, Hslao RJ. Rapid radioimmunoassay of circulating chromogranin A: in vitro stability, exploration of the neuroendocrine character of neoplasia, and assessment of the effects of organ failure. Clin Chem. 1989 Aug;35(8):1631–7.
39. Zatelli MC, Torta M, Leon A, Ambrosio MR, Gion M, Tomassetti P, De Braud F, Delle Fave G, Dogliotti L, degli Uberti EC, Italian CromaNet Working Group. Chromogranin A as a marker of neuroendocrine neoplasia: an Italian Multicenter Study. Endocr Relat Cancer. 2007 Jun;14(2):473–82.
40. Lv Y, Han X, Zhang C, Fang Y, Pu N, Ji Y, Wang D, Xuefeng X, Lou W. Combined test of serum CgA and NSE improved the power of prognosis prediction of NF-pNETs. Endocr Connect. 2018 Jan;7(1):169–78.
41. Cimitan M, Buonadonna A, Cannizzaro R, Canzonieri V, Borsatti E, Ruffo R, De Apollonia L. Somatostatin receptor scintigraphy versus chromogranin A assay in the management of patients with neuroendocrine tumors of different types: clinical role. Ann Oncol Off J Eur Soc Med Oncol. 2003 Jul;14(7):1135–41.
42. Walter T, Chardon L, Chopin-laly X, Raverot V, Caffin A-G, Chayvialle J-A, Scoazec J-Y, Lombard-Bohas C. Is the combination of chromogranin A and pancreatic polypeptide serum determinations of interest in the diagnosis and follow-up of gastro-entero-pancreatic neuroendocrine tumours? Eur J Cancer. 2012 Aug 1;48(12):1766–73.
43. Vinik AI, Silva MP, Woltering G, W. Go VL, Warner R, Caplin M. Biochemical Testing for Neuroendocrine Tumors. Pancreas. 2009 Nov;38(8):876.

44. Zuetenhorst JM, Bonfrer JMGM, Korse CM, Bakker R, van Tinteren H, Taal BG. Carcinoid heart disease: the role of urinary 5-hydroxyindoleacetic acid excretion and plasma levels of atrial natriuretic peptide, transforming growth factor-beta and fibroblast growth factor. Cancer. 2003 Apr 1;97(7):1609–15.

45. Bhattacharyya S, Toumpanakis C, Chilkunda D, Caplin ME, Davar J. Risk factors for the development and progression of carcinoid heart disease. Am J Cardiol. 2011 Apr 15;107(8):1221–6.

46. Service FJ, Natt N. The Prolonged Fast. J Clin Endocrinol Metab. 2000 Nov 1;85(11):3973–4.

47. Hirshberg B, Livi A, Bartlett DL, Libutti SK, Alexander HR, Doppman JL, Skarulis MC, Gorden P. Forty-Eight-Hour Fast: The Diagnostic Test for Insulinoma. J Clin Endocrinol Metab. 2000 Sep 1;85(9):3222–6.

48. van Bon AC, Benhadi N, Endert E, Fliers E, Wiersinga WM. Evaluation of endocrine tests. D: the prolonged fasting test for insulinoma. Neth J Med. 2009 Aug;67(7):274–8.

49. Sundin A, Arnold R, Baudin E, Cwikla JB, Eriksson B, Fanti S, Fazio N, Giammarile F, Hicks RJ, Kjaer A, Krenning E, Kwekkeboom D, Lombard-Bohas C, O'Connor JM, O'Toole D, Rockall A, Wiedenmann B, Valle JW, Vullierme M-P, Participants all other ACC. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: Radiological, Nuclear Medicine and Hybrid Imaging. Neuroendocrinology. 2017;105(3):212–44.

50. Luo Y, Ameli S, Pandey A, Khoshpouri P, Ghasabeh MA, Pandey P, Li Z, Hu D, Kamel IR. Semi-quantitative visual assessment of hepatic tumor burden can reliably predict survival in neuroendocrine liver metastases treated with transarterial chemoembolization. Eur Radiol. 2019 Nov 1;29(11):5804–12.

51. Steinmüller T, Kianmanesh R, Falconi M, Scarpa A, Taal B, Kwekkeboom DJ, Lopes JM, Perren A, Nikou G, Yao J, Delle Fave GF, O'Toole D, Frascati Consensus Conference participants. Consensus guidelines for the management of patients with liver metastases from digestive (neuro)endocrine tumors: foregut, midgut, hindgut, and unknown primary. Neuroendocrinology. 2008;87(1):47–62.

52. Sundin A, Vullierme M-P, Kaltsas G, Plöckinger U. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: Radiological Examinations. Neuroendocrinology. 2009;90(2):167–83.

53. Fehrenbach U, Kahn J, Fahlenkamp U, Baur A, Pavel M, Geisel D, Denecke T. Optimized imaging of the lower abdomen and pelvic region in hepatocyte-specific MRI: evaluation of a whole-abdomen first-pass shuttle protocol in patients with neuroendocrine neoplasms. Acta Radiol. 2019 Sep 1;60(9):1074–83.

54. Van Beers BE, Pastor CM, Hussain HK. Primovist, Eovist: what to expect? J Hepatol. 2012 Aug;57(2):421–9.

55. Dromain C, de Baere T, Lumbroso J, Caillet H, Laplanche A, Boige V, Ducreux M, Duvillard P, Elias D, Schlumberger M, Sigal R, Baudin E. Detection of liver metastases from endocrine tumors: a prospective comparison of somatostatin receptor scintigraphy, computed tomography, and magnetic resonance imaging. J Clin Oncol Off J Am Soc Clin Oncol. 2005 Jan 1;23(1):70–8.

56. Ronot M, Clift AK, Baum RP, Singh A, Kulkarni HR, Frilling A, Vilgrain V. Morphological and Functional Imaging for Detecting and Assessing the Resectability of Neuroendocrine Liver Metastases. Neuroendocrinology. 2018;106(1):74–88.

57. Neri E, Bali MA, Ba-Ssalamah A, Boraschi P, Brancatelli G, Alves FC, 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):921–31.

58. Hayoz R, Vietti-Violi N, Duran R, Knebel J-F, Ledoux J-B, Dromain C. The combination of hepatobiliary phase with Gd-EOB-DTPA and DWI is highly accurate for the detection and characterization of liver metastases from neuroendocrine tumor. Eur Radiol. 2020 Dec 1;30(12):6593–602.

59. Vreugdenburg TD, Ma N, Duncan JK, Riitano D, Cameron AL, Maddern GJ. Comparative diagnostic accuracy of hepatocyte-specific gadoxetic acid (Gd-EOB-DTPA) enhanced MR imaging and contrast enhanced CT for the detection of liver metastases: a systematic review and meta-analysis. Int J Colorectal Dis. 2016 Nov;31(11):1739–49.

60. Karaosmanoglu AD, Onur MR, Ozmen MN, Akata D, Karcaaltincaba M. Magnetic Resonance Imaging of Liver Metastasis. Semin Ultrasound CT MR. 2016 Dec;37(6):533–48.

61. Tirumani SH, Jagannathan JP, Braschi-Amirfarzan M, Qin L, Balthazar P, Ramaiya NH, Shinagare AB. Value of hepatocellular phase imaging after intravenous gadoxetate disodium for assessing hepatic metastases from gastroenteropancreatic neuroendocrine tumors: comparison with other MRI pulse sequences and with extracellular agent. Abdom Radiol. 2018 Sep 1;43(9):2329–39.

62. Luersen GF, Wei W, Tamm EP, Bhosale PR, Szklaruk J. Evaluation of Magnetic Resonance (MR) Biomarkers for Assessment of Response With Response Evaluation Criteria in Solid Tumors: Comparison of the Measurements of Neuroendocrine Tumor Liver Metastases (NETLM) With Various MR Sequences and at Multiple Phases of Contrast Administration. J Comput Assist Tomogr. 2016 Oct;40(5):717–22.

63. Armbruster M, Zech CJ, Sourbron S, Ceelen F, Auernhammer CJ, Rist C, Haug A, Singnurkar A, Reiser MF, Sommer WH. Diagnostic accuracy of dynamic gadoxetic-acid-enhanced MRI and PET/CT compared in patients with liver metastases from neuroendocrine neoplasms. J Magn Reson Imaging JMRI. 2014 Aug;40(2):457–66.

64. Seidensticker M, Seidensticker R, Mohnike K, Wybranski C, Kalinski T, Luess S, Pech M, Wust P, Ricke J. Quantitative in vivo assessment of radiation injury of the liver using Gd-EOB-DTPA enhanced MRI: tolerance dose of small liver volumes. Radiat Oncol Lond Engl. 2011 Apr 17;6:40.

65. Granata V, Fusco R, Maio F, Avallone A, Nasti G, Palaia R, Albino V, Grassi R, Izzo F, Petrillo A. Qualitative assessment of EOB-GD-DTPA and Gd-BT-DO3A MR contrast studies in HCC patients and colorectal liver metastases. Infect Agent Cancer [Internet]. 2019 Nov 27 [cited 2020 May 21];14. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6882051/

66. Lee Y, Kim SY, Lee SS, Park SH, Kim KW, Byun JH, Lee M. Pitfalls in Gd-EOB-DTPA–Enhanced Liver Magnetic Resonance Imaging With an Emphasis on Nontumorous Lesions. Clin Liver Dis. 2018 Sep 7;12(2):50–9.

67. Inchingolo R, Faletti R, Grazioli L, Tricarico E, Gatti M, Pecorelli A, Ippolito D. MR with Gd-EOB-DTPA in assessment of liver nodules in cirrhotic patients. World J Hepatol. 2018 Jul 27;10(7):462–73.

68. Zech CJ, Herrmann KA, Reiser MF, Schoenberg SO. MR imaging in patients with suspected liver metastases: value of liver-specific contrast agent Gd-EOB-DTPA. Magn Reson Med Sci MRMS Off J Jpn Soc Magn Reson Med. 2007;6(1):43–52.

69. Borusewicz P, Stańczyk E, Kubiak K, Spużak J, Glińska-Suchocka K, Jankowski M, Sławuta P, Kubiak-Nowak D, Podgórski P. Magnetic resonance imaging of liver tumors using gadoxetic acid (Gd-EOB-DTPA) - pilot study. BMC Vet Res [Internet]. 2019 Aug 14 [cited 2020 May 22];15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6694506/

70. Campos JT, Sirlin CB, Choi J-Y. Focal hepatic lesions in Gd-EOB-DTPA enhanced MRI: the atlas. Insights Imaging. 2012 Jun 15;3(5):451–74.

71. Hamm B, Staks T, Mühler A, Bollow M, Taupitz M, Frenzel T, Wolf KJ, Weinmann HJ, Lange L. Phase I clinical evaluation of Gd-EOB-DTPA as a hepatobiliary MR contrast agent: safety, pharmacokinetics, and MR imaging. Radiology. 1995 Jun;195(3):785–92.

72. Khosa F, Khan AN, Eisenberg RL. Hypervascular Liver Lesions on MRI. Am J Roentgenol. 2011 Aug 1;197(2):W204–20.

73. Bhayana R, Baliyan V, Kordbacheh H, Kambadakone A. Hepatobiliary phase enhancement of liver metastases on gadoxetic acid MRI: assessment of frequency and patterns. Eur Radiol. 2021 Mar 1;31(3):1359–66.

74. Danet I-M, Semelka RC, Leonardou P, Braga L, Vaidean G, Woosley JT, Kanematsu M. Spectrum of MRI Appearances of Untreated Metastases of the Liver. Am J Roentgenol. 2003 Sep 1;181(3):809–17.

75. Chen L, Zhang J, Zhang L, Bao J, Liu C, Xia Y, Huang X, Wang J. Meta-analysis of gadoxetic acid disodium (Gd-EOB-DTPA)-enhanced magnetic resonance imaging for the detection of liver metastases. PloS One. 2012;7(11):e48681.

76. Dromain C, de Baere T, Baudin E, Galline J, Ducreux M, Boige V, Duvillard P, Laplanche A, Caillet H, Lasser P, Schlumberger M, Sigal R. MR Imaging of Hepatic Metastases

Caused by Neuroendocrine Tumors: Comparing Four Techniques. Am J Roentgenol. 2003 Jan 1;180(1):121–8.

77. Morse B, Jeong D, Thomas K, Diallo D, Strosberg JR. Magnetic Resonance Imaging of Neuroendocrine Tumor Hepatic Metastases: Does Hepatobiliary Phase Imaging Improve Lesion Conspicuity and Interobserver Agreement of Lesion Measurements? Pancreas. 2017 Oct;46(9):1219–24.
78. Giesel FL, Kratochwil C, Mehndiratta A, Wulfert S, Moltz JH, Zechmann CM, Kauczor HU, Haberkorn U, Ley S. Comparison of neuroendocrine tumor detection and characterization using DOTATOC-PET in correlation with contrast enhanced CT and delayed contrast enhanced MRI. Eur J Radiol. 2012 Oct;81(10):2820–5.
79. Lestra T, Kanagaratnam L, Mulé S, Janvier A, Brixi H, Cadiot G, Dohan A, Hoeffel C. Measurement variability of liver metastases from neuroendocrine tumors on different magnetic resonance imaging sequences. Diagn Interv Imaging. 2018 Feb 1;99(2):73–81.
80. Lavelle LP, O'Neill AC, McMahon CJ, Cantwell CP, Heffernan EJ, Malone DE, Daly L, Skehan SJ. Is diffusion-weighted MRI sufficient for follow-up of neuroendocrine tumour liver metastases? Clin Radiol. 2016 Sep 1;71(9):863–8.
81. Tamm EP, Kim EE, Ng CS. Imaging of neuroendocrine tumors. Hematol Oncol Clin North Am. 2007 Jun;21(3):409–32; vii.
82. Blodgett TM, Meltzer CC, Townsend DW. PET/CT: Form and Function. Radiology. 2007 Feb 1;242(2):360–85.
83. Hofman MS, Lau WFE, Hicks RJ. Somatostatin Receptor Imaging with 68Ga DOTATATE PET/CT: Clinical Utility, Normal Patterns, Pearls, and Pitfalls in Interpretation. RadioGraphics. 2015 Mar 1;35(2):500–16.
84. Patel YC. Somatostatin and Its Receptor Family. Front Neuroendocrinol. 1999 Jul 1;20(3):157–98.
85. Krenning EP, Breeman W a. P, Kooij PPM, Lameris JS, Bakker WH, Koper JW, Ausema L, Reubi JC, Lamberts SWJ. Localisation of endocrine-related tumours with radioiodinated analogue of somatostatin. The Lancet. 1989 Feb 4;333(8632):242–4.
86. Buchmann I, Henze M, Engelbrecht S, Eisenhut M, Runz A, Schäfer M, Schilling T, Haufe S, Herrmann T, Haberkorn U. Comparison of 68Ga-DOTATOC PET and 111In-DTPAOC

(Octreoscan) SPECT in patients with neuroendocrine tumours. Eur J Nucl Med Mol Imaging. 2007 Oct 1;34(10):1617–26.

87. Guenter R, Aweda T, Carmona Matos DM, Jang S, Whitt J, Cheng Y-Q, Liu XM, Chen H, Lapi SE, Jaskula-Sztul R. Overexpression of somatostatin receptor type 2 in neuroendocrine tumors for improved Ga68-DOTATATE imaging and treatment. Surgery. 2020 Jan;167(1):189–96.
88. Wild D, Mäcke HR, Waser B, Reubi JC, Ginj M, Rasch H, Müller-Brand J, Hofmann M. 68Ga-DOTANOC: a first compound for PET imaging with high affinity for somatostatin receptor subtypes 2 and 5. Eur J Nucl Med Mol Imaging. 2005 Jun 1;32(6):724–724.
89. Baur ADJ, Pavel M, Prasad V, Denecke T. Diagnostic imaging of pancreatic neuroendocrine neoplasms (pNEN): tumor detection, staging, prognosis, and response to treatment. Acta Radiol Stockh Swed 1987. 2016 Mar;57(3):260–70.
90. Kapoor V, McCook BM, Torok FS. An Introduction to PET-CT Imaging. RadioGraphics. 2004 Mar 1;24(2):523–43.
91. Has Simsek D, Kuyumcu S, Turkmen C, Sanlı Y, Aykan F, Unal S, Adalet I. Can complementary 68Ga-DOTATATE and 18F-FDG PET/CT establish the missing link between histopathology and therapeutic approach in gastroenteropancreatic neuroendocrine tumors? J Nucl Med Off Publ Soc Nucl Med. 2014 Nov;55(11):1811–7.
92. Jensen RT, Cadiot G, Brandi ML, de Herder WW, Kaltsas G, Komminoth P, Scoazec J-Y, Salazar R, Sauvanet A, Kianmanesh R, Barcelona Consensus Conference participants. ENETS Consensus Guidelines for the management of patients with digestive neuroendocrine neoplasms: functional pancreatic endocrine tumor syndromes. Neuroendocrinology. 2012;95(2):98–119.
93. Dörffel Y, Wermke W. Neuroendocrine Tumors: Characterization with Contrast-Enhanced Ultrasonography. Ultraschall Med - Eur J Ultrasound. 2008 Oct;29(05):506–14.
94. Ito T, Sasano H, Tanaka M, Osamura RY, Sasaki I, Kimura W, Takano K, Obara T, Ishibashi M, Nakao K, Doi R, Shimatsu A, Nishida T, Komoto I, Hirata Y, Nakamura K, Igarashi H, Jensen RT, Wiedenmann B, Imamura M. Epidemiological study of gastroenteropancreatic neuroendocrine tumors in Japan. J Gastroenterol. 2010 Feb 1;45(2):234–43.

95. Fujimori N, Osoegawa T, Lee L, Tachibana Y, Aso A, Kubo H, Kawabe K, Igarashi H, Nakamura K, Oda Y, Ito T. Efficacy of endoscopic ultrasonography and endoscopic ultrasonography-guided fine-needle aspiration for the diagnosis and grading of pancreatic neuroendocrine tumors. Scand J Gastroenterol. 2016;51(2):245–52.

96. Kann PH, Balakina E, Ivan D, Bartsch DK, Meyer S, Klose K-J, Behr T, Langer P. Natural course of small, asymptomatic neuroendocrine pancreatic tumours in multiple endocrine neoplasia type 1: an endoscopic ultrasound imaging study. Endocr Relat Cancer. 2006 Dec 1;13(4):1195–202.

97. Chang KJ. State of the art lecture: Endoscopic ultrasound (EUS) and FNA in pancreatico-biliary tumors. Endoscopy. 2006 Jun;39(S 1):56–60.

98. Pavel M, Öberg K, Falconi M, Krenning EP, Sundin A, Perren A, Berruti A. Gastroenteropancreatic neuroendocrine neoplasms: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2020 Jul 1;31(7):844–60.

99. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan 1;45(2):228–47.

100. Riihimäki M, Hemminki A, Sundquist K, Sundquist J, Hemminki K. The epidemiology of metastases in neuroendocrine tumors. Int J Cancer. 2016 Dec 15;139(12):2679–86.

101. Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, Shankar L, Bogaerts J, Chen A, Dancey J, Hayes W, Hodi FS, Hoekstra OS, Huang EP, Lin N, Liu Y, Therasse P, Wolchok JD, Seymour L. RECIST 1.1-Update and clarification: From the RECIST committee. Eur J Cancer Oxf Engl 1990. 2016 Jul;62:132–7.

102. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018 Aug;18(8):500–10.

103. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, Cai W, Kermany DS, Sun X, Chen J, He L, Zhu J, Tian P, Shao H, Zheng L, Hou R, Hewett S, Li G, Liang P, Zang X, Zhang Z, Pan L, Cai H, Ling R, Li S, Cui Y, Tang S, Ye H, Huang X, He W, Liang W, Zhang Q, Jiang J, Yu W, Gao J, Ou W, Deng Y, Hou Q, Wang B, Yao C, Liang Y, Zhang S, Duan Y, Zhang R, Gibson S, Zhang CL, Li O, Zhang ED, Karin G, Nguyen N, Wu X,

Wen C, Xu J, Xu W, Wang B, Wang W, Li J, Pizzato B, Bao C, Xiang D, He W, He S, Zhou Y, Haw W, Goldbaum M, Tremoulet A, Hsu C-N, Carter H, Zhu L, Zhang K, Xia H. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019 Mar;25(3):433–8.

104. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–44.
105. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. In: Classification in BioApps. Springer; 2018. p. 323–50.
106. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine Learning in Medical Imaging. IEEE Signal Process Mag. 2010 Jul;27(4):25–38.
107. 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 1;29(7):3338–47.
108. Tiwari P, Prasanna P, Wolansky L, Pinho M, Cohen M, Nayate AP, Gupta A, Singh G, Hatanpaa KJ, Sloan A, Rogers L, Madabhushi A. Computer-Extracted Texture Features to Distinguish Cerebral Radionecrosis from Recurrent Brain Tumors on Multiparametric MRI: A Feasibility Study. AJNR Am J Neuroradiol. 2016 Dec;37(12):2231–6.
109. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, Liu X, Chen L, Qu J, Xu S, Tang X, Mao Y, Liu Y, Li D. Deep learning–based detection and segmentation-assisted management of brain metastases. Neuro-Oncol. 2020 Apr;22(4):505–14.
110. Ha R, Mutasa S, Karcich J, Gupta N, Pascual Van Sant E, Nemer J, Sun M, Chang P, Liu MZ, Jambawalikar S. Predicting Breast Cancer Molecular Subtype with MRI Dataset Utilizing Convolutional Neural Network Algorithm. J Digit Imaging. 2019 Apr;32(2):276–82.
111. Aldoj N, Lukas S, Dewey M, Penzkofer T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. Eur Radiol. 2020 Feb 1;30(2):1243–53.
112. Ushinsky A, Bardis M, Glavis-Bloom J, Uchio E, Chantaduly C, Nguyentat M, Chow D, Chang PD, Houshyar R. A 3D-2D Hybrid U-Net Convolutional Neural Network Approach

to Prostate Organ Segmentation of Multiparametric MRI. Am J Roentgenol. 2021 Jan 1;216(1):111–6.

113. Govind D, Jen K-Y, Matsukuma K, Gao G, Olson KA, Gui D, Wilding GE, Border SP, Sarder P. Improving the accuracy of gastrointestinal neuroendocrine tumor grading with deep learning. Sci Rep. 2020 Jul 6;10(1):11064.
114. Niazi MKK, Tavolara TE, Arole V, Hartman DJ, Pantanowitz L, Gurcan MN. Identifying tumor in pancreatic neuroendocrine neoplasms from Ki67 images using transfer learning. PloS One. 2018;13(4):e0195621.
115. Gonzalez D, Dietz RL, Pantanowitz L. Feasibility of a deep learning algorithm to distinguish large cell neuroendocrine from small cell lung carcinoma in cytology specimens. Cytopathol Off J Br Soc Clin Cytol. 2020 Sep;31(5):426–31.
116. Gao X, Wang X. Deep learning for World Health Organization grades of pancreatic neuroendocrine tumors on contrast-enhanced magnetic resonance images: a preliminary study. Int J Comput Assist Radiol Surg. 2019 Nov;14(11):1981–91.
117. Luo Y, Chen X, Chen J, Song C, Shen J, Xiao H, Chen M, Li Z-P, Huang B, Feng S-T. Preoperative Prediction of Pancreatic Neuroendocrine Neoplasms Grading Based on Enhanced Computed Tomography Imaging: Validation of Deep Learning with a Convolutional Neural Network. Neuroendocrinology. 2020;110(5):338–50.
118. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. RadioGraphics. 2017 Feb 17;37(2):505–15.
119. Gotra A, Sivakumaran L, Chartrand G, Vu K-N, Vandenbroucke-Menu F, Kauffmann C, Kadoury S, Gallix B, de Guise JA, Tang A. Liver segmentation: indications, techniques and future directions. Insights Imaging. 2017 Aug;8(4):377–92.
120. Lo SB, Lou SA, Lin JS, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging. 1995;14(4):711–8.
121. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. Proc IEEE. 1998;86:2278–324.
122. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst. 2012 Jan;25(2):1097–105.

123. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv150504597 Cs [Internet]. 2015 May 18 [cited 2021 Mar 13]; Available from: http://arxiv.org/abs/1505.04597

124. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, Maier-Hein KH. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. ArXiv180910486 Cs [Internet]. 2018 Sep 27 [cited 2021 Mar 16]; Available from: http://arxiv.org/abs/1809.10486

125. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Cham: Springer International Publishing; 2016. p. 424–32. (Lecture Notes in Computer Science).

126. Baştanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol Clifton NJ. 2014;1107:105–28.

127. Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. Cancers. 2019 Aug 23;11(9).

128. Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945;26(3):297–302.

129. Park SH, Choi J, Byeon JS. Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence. Korean J Radiol. 2021 Mar;22(3):442–53.

130. Dreyer KJ, Geis JR. When Machines Think: Radiology's Next Frontier. Radiology. 2017 Dec;285(3):713–8.

131. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging. 1994;13(4):716–24.

132. Do S, Song KD, Chung JW. Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. Korean J Radiol. 2020 Jan;21(1):33–41.

133. Hand DJ, Christen P, Kirielle N. F*: an interpretable transformation of the F-measure. Mach Learn. 2021 Mar 15;1–6.

134. Nguyen Q, Lee S-W. Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network. IEEE Access. 2019 Mar 11;PP.

135. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020 Jan 2;21(1):6.

136. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Min. 2021 Feb 4;14(1):13.

137. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PloS One. 2017;12(6):e0177678.

138. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec;42:60–88.

139. Le T-N, Bao PT, Huynh HT. Liver Tumor Segmentation from MR Images Using 3D Fast Marching Algorithm and Single Hidden Layer Feedforward Neural Network. BioMed Res Int. 2016;2016:3219068.

140. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021 Feb;18(2):203–11.

141. 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 1;264:47–56.

142. Fehrenbach U, Xin S, Hartenstein A, Auer TA, Dräger F, Froböse K, Jann H, Mogl M, Amthauer H, Geisel D, Denecke T, Wiedenmann B, Penzkofer T. Automatized Hepatic Tumor Volume Analysis of Neuroendocrine Liver Metastases by Gd-EOB MRI-A Deep-Learning Model to Support Multidisciplinary Cancer Conference Decision-Making. Cancers. 2021 May 31;13(11).

143. Partouche E, Yeh R, Eche T, Rozenblum L, Carrere N, Guimbaud R, Dierickx LO, Rousseau H, Dercle L, Mokrane F-Z. Updated Trends in Imaging Practices for Pancreatic

Neuroendocrine Tumors (PNETs): A Systematic Review and Meta-Analysis to Pave the Way for Standardization in the New Era of Big Data and Artificial Intelligence. Front Oncol. 2021;11:628408.

144. Doemel LA, Chapiro J, Laage Gaupp F, Savic LJ, Kucukkaya AS, Petukhova A, Tefera J, Zeevi T, Lin M, Schlachter T, Jaffe A, Strazzabosco M, Patel T, Stein SM. Reliable prediction of survival in advanced-stage hepatocellular carcinoma treated with sorafenib: comparing 1D and 3D quantitative tumor response criteria on MRI. Eur Radiol. 2021 May 1;31(5):2737–46.

145. Aghighi M, Boe J, Rosenberg J, Von Eyben R, Gawande RS, Petit P, Sethi TK, Sharib J, Marina NM, DuBois SG, Daldrup-Link HE. Three-dimensional Radiologic Assessment of Chemotherapy Response in Ewing Sarcoma Can Be Used to Predict Clinical Outcome. Radiology. 2016 Sep;280(3):905–15.

146. Orsatti G, Morosi C, Giraudo C, Varotto A, Crimì F, Bonzini M, Minotti M, Frigo AC, Zanetti I, Chiaravalli S, Casanova M, Ferrari A, Bisogno G, Stramare R. Pediatric Rhabdomyosarcomas: Three-Dimensional Radiological Assessments after Induction Chemotherapy Predict Survival Better than One-Dimensional and Two-Dimensional Measurements. Cancers. 2020 Dec 17;12(12):E3808.

147. Schiavon G, Ruggiero A, Bekers DJ, Barry PA, Sleijfer S, Kloth J, Krestin GP, Schöffski P, Verweij J, Mathijssen RHJ. The effect of baseline morphology and its change during treatment on the accuracy of Response Evaluation Criteria in Solid Tumours in assessment of liver metastases. Eur J Cancer Oxf Engl 1990. 2014 Mar;50(5):972–80.

148. Rothe JH, Grieser C, Lehmkuhl L, Schnapauff D, Fernandez CP, Maurer MH, Mussler A, Hamm B, Denecke T, Steffen IG. Size determination and response assessment of liver metastases with computed tomography--comparison of RECIST and volumetric algorithms. Eur J Radiol. 2013 Nov;82(11):1831–9.

149. Hadjiiski L, Weizer AZ, Alva A, Caoili EM, Cohan RH, Cha K, Chan H-P. Treatment Response Assessment for Bladder Cancer on CT Based on Computerized Volume Analysis, World Health Organization Criteria, and RECIST. AJR Am J Roentgenol. 2015 Aug;205(2):348–52.

150. 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 29;8:671–83.

151. Jansen MJA, Kuijf HJ, Niekel M, Veldhuis WB, Wessels FJ, Viergever MA, Pluim JPW. Liver segmentation and metastases detection in MR images using convolutional neural networks. J Med Imaging Bellingham Wash. 2019 Oct;6(4):044003.

152. Takenaga T, Hanaoka S, Nomura Y, Nakao T, Shibata H, Miki S, Yoshikawa T, Hayashi N, Abe O. Multichannel three-dimensional fully convolutional residual network-based focal liver lesion detection and classification in Gd-EOB-DTPA-enhanced MRI. Int J Comput Assist Radiol Surg. 2021 Sep 1;16(9):1527–36.

153. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean J Radiol. 2019 Mar;20(3):405–10.

154. Trivizakis E, Manikis GC, Nikiforaki K, Drevelegas K, Constantinides M, Drevelegas A, Marias K. Extending 2-D Convolutional Neural Networks to 3-D for Advancing Deep Learning Cancer Classification With Application to MRI Liver Tumor Differentiation. IEEE J Biomed Health Inform. 2019 May;23(3):923–30.

155. Huynh HT, Le-Trong N, Bao PT, Oto A, Suzuki K. Fully automated MR liver volumetry using watershed segmentation coupled with active contouring. Int J Comput Assist Radiol Surg. 2017 Feb 1;12(2):235–43.

156. Dura E, Domingo J, Göçeri E, Martí-Bonmatí L. A method for liver segmentation in perfusion MR images using probabilistic atlases and viscous reconstruction. Pattern Anal Appl. 2018 Nov 1;21(4):1083–95.

157. Goehler A, Harry Hsu T-M, Lacson R, Gujrathi I, Hashemi R, Chlebus G, Szolovits P, Khorasani R. Three-Dimensional Neural Network to Automatically Assess Liver Tumor Burden Change on Consecutive Liver MRIs. J Am Coll Radiol JACR. 2020 Nov;17(11):1475–84.

158. Zarghampour M, Fouladi DF, Pandey A, Ghasabeh MA, Pandey P, Varzaneh FN, Khoshpouri P, Shao N, Pan L, Grimm R, Kamel IR. Utility of volumetric contrast-enhanced

and diffusion-weighted MRI in differentiating between common primary hypervascular liver tumors. J Magn Reson Imaging JMRI. 2018 Oct;48(4):1080–90.

159. Piana G, Trinquart L, Meskine N, Barrau V, Beers BV, Vilgrain V. New MR imaging criteria with a diffusion-weighted sequence for the diagnosis of hepatocellular carcinoma in chronic liver diseases. J Hepatol. 2011 Jul 1;55(1):126–32.
160. Jha AK, Kupinski MA, Rodríguez JJ, Stephen RM, Stopeck AT. Task-based evaluation of segmentation algorithms for diffusion-weighted MRI without using a gold standard. Phys Med Biol. 2012 Jul 7;57(13):4425–46.
161. Zheng R, Wang L, Wang C, Yu X, Chen W, Li Y, Li W, Yan F, Wang H, Li R. Feasibility of automatic detection of small hepatocellular carcinoma (≤2 cm) in cirrhotic liver based on pattern matching and deep learning. Phys Med Biol. 2021 Apr 16;66(8).
162. Li J, Yang Y. Clinical Study of Diffusion-Weighted Imaging in the Diagnosis of Liver Focal Lesion. J Med Syst. 2019 Jan 16;43(3):43.
163. Taouli B, Koh D-M. Diffusion-weighted MR Imaging of the Liver. Radiology. 2010 Jan 1;254(1):47–66.
164. Tsurusaki M, Sofue K, Murakami T. Current evidence for the diagnostic value of gadoxetic acid-enhanced magnetic resonance imaging for liver metastasis. Hepatol Res Off J Jpn Soc Hepatol. 2016 Aug;46(9):853–61.

7. Eidesstattliche Versicherung

„Ich, Siyi Xin, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „AI-basierte volumetrische Analyse der Lebermetastasenlast bei Patienten mit neuroendokrinen Neoplasmen (NEN)“ („AI-based Volumetric Analysis of Liver Metastases Load in Neuroendocrine Neoplasm (NEN) Patients“) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum Unterschrift

Anteilserklärung an etwaigen erfolgten Publikationen

Siyi Xin hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Fehrenbach U, Xin S, Hartenstein A, Auer TA, Dräger F, Froböse K, Jann H, Mogl M, Amthauer H, Geisel D, Denecke T, Wiedenmann B, Penzkofer T. Automatized Hepatic Tumor Volume Analysis of Neuroendocrine Liver Metastases by Gd-EOB MRI-A Deep-Learning Model to Support Multidisciplinary Cancer Conference Decision-Making. Cancers (Basel). 2021 May 31;13(11):2726. doi: 10.3390/cancers13112726. PMID: 34072865; PMCID: PMC8199286.

Beitrag im Einzelnen (bitte detailliert ausführen):

1. **Erstellung von KI-Trainingssätzen, internen Validierungssätzen und externen Validierungssätzen.**

Frau Siyi Xin war an der Segmentierung von Leber und Lebertumoren für die meisten MRTs in der AI-Entwicklungskohorte (296 von 398) und an der Segmentierung von HBP-Sequenzen für 66 Patienten in der MCC-Kohorte beteiligt. Sie wurden alle von Dr. Uli Fehrenbach überprüft.

2. **Sammlung klinischer Datensätze.**

Erfassung der klinischen Daten aller Patienten und Erstellung einer Datenbank (einschließlich Tests, Behandlung, Nachsorge und Prognose) zur Archivierung und Analyse der Patientendaten.

3. **Aufbereitung der Daten und Vorbereitung für die statistische Analyse.**

Frau Siyi Xin bereitete die Daten auf und erstellte SPSS-Tabellen (Tabelle 1) für die statistische Analyse.

4. **Verfassen der Publikation, Literaturrecherche.**

Frau Siyi Xin führte die Literaturrecherche durch und schrieb den einleitenden Teil der Publikation unter der Aufsicht von Prof. Dr. med. Bertram Wiedenmann, PD Dr. med. Tobias Penzkofer, und Dr. med. Uli Fehrenbach.

_____

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

_____ _____

Unterschrift des Doktoranden/der Doktorandin

8. Curriculum vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

9. List of Publications

Fehrenbach U, Xin S, Hartenstein A, Auer TA, Dräger F, Froböse K, Jann H, Mogl M, Amthauer H, Geisel D, Denecke T, Wiedenmann B, Penzkofer T. **Automatized Hepatic Tumor Volume Analysis of Neuroendocrine Liver Metastases by Gd-EOB MRI-A Deep-Learning Model to Support Multidisciplinary Cancer Conference Decision-Making.** Cancers (Basel). 2021 May 31;13(11):2726. doi: 10.3390/cancers13112726. PMID: 34072865; PMCID: PMC8199286.

10. Acknowledgements

I take this opportunity to express my gratitude to those who have played a part in completing this work.

I would first like to express my sincere gratitude to my wonderful supervisors, and I feel fortunate to have been co-mentored by Prof. Dr. med. Bertram Wiedenmann, PD Dr. med. Tobias Penzkofer, and Dr. med. Uli Fehrenbach. They are all experts in their respective fields of research.

Prof. Dr. med. Bertram Wiedenmann is a leading expert in the field of neuroendocrine tumors, and I have benefited from his knowledgeable and cutting-edge expertise, his patient guidance and open-mindedness over the years. I am grateful to him for his accurate insight, interesting discussions and for inspiring my scientific thinking and motivating me to become a better scientist.

PD Dr. med. Tobias Penzkofer has great achievements in the field of AI-assisted imaging. His patient and attentive guidance helped me to open up the field of deep learning, which was new and most challenging for me. His optimistic attitude and open-mindedness have been invaluable, as well as his support and advice for the completion of my thesis.

Dr. med. Uli Fehrenbach is the senior imaging expert in our Neuroendocrine tumor conference. I thank him for his patience, kindness, and incredible help at every step of the way. His careful explanation of my imaging expertise and constant encouragement during the work process gave me the confidence to cope with most of the challenges during this time.

I am particularly grateful to Dr. med. Henning Jann for his tremendous help and guidance in my work, for giving me valuable advice and suggestions, and for his great support in the collection of clinical data.

I would also like to thank my best friend, Dr. med. Xueqi Cheng, especially during the years of the novel coronavirus-19 pandemic, who was with me when I was homesick in a foreign country. Without her encouragement and support, this thesis would not have been possible.

Last but not least, I would like to express my heartfelt gratitude to my mother, Mrs. Wen Xin, for raising me on her own and giving me the utmost support in pursuing my dreams in Germany. my

mother has given her life to me and because of all the hard work she has done, I am where I am today. (I love you)