Title

Convolutional neural network-automated hepatobiliary phase adequacy evaluation may optimize examination time

Authors

Cunha, Guilherme Moura
Hasenstab, Kyle A
Higaki, Atsushi
et al.

Publication Date

2020-03-01

DOI

10.1016/j.ejrad.2020.108837

Peer reviewed

HHS Public Access

Author manuscript

Eur J Radiol. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:

Eur J Radiol. 2020 March ; 124: 108837. doi:10.1016/j.ejrad.2020.108837.

Convolutional neural network-automated hepatobiliary phase adequacy evaluation may optimize examination time

Guilherme Moura Cunha[a,1,2], **Kyle A. Hasenstab**[a,b,c,1,2], **Atsushi Higaki**[a,2], **Kang Wang**[a,b,2], **Timo Delgado**[a,2], **Ryan L. Brunsing**[d,2], **Alexandra Schlein**[a,2], **Armin Schwartzman**[a,2], **Albert Hsiao**[b,2], **Claude B Sirlin**[a,2], **Katie J. Fowler**[a,2]

[a]Liver Imaging Group, Department of Radiology, University of California San Diego, La Jolla, CA, United States

[b]AiDA Laboratory, Department of Radiology, University of California San Diego, La Jolla, CA, United States

[c]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, United States

[d]Radiology, Stanford University, Palo Alto, CA, United States

Abstract

Purpose—To develop and evaluate the performance of a fully-automated convolutional neural network (CNN)-based algorithm to evaluate hepatobiliary phase (HBP) adequacy of gadoxetate disodium (EOB)-enhanced MRI. Secondarily, we explored the potential of the proposed CNN algorithm to reduce examination length by applying it to EOB-MRI examinations.

Corresponding author at: 1(858) 246 2220, Liver Imaging Group, Department of Radiology, University of California San Diego, La Jolla, CA, 0888, USA. gcunha@health.ucsd.edu (G.M. Cunha).

[1]These authors contributed equally to this work.

[2]Physical address: Altman Clinical Translational Research Institute, 9452 Medical Center Drive, Lower Level 501, La Jolla, CA 92037.

Declaration of Competing Interest

All authors or institutions involved in this work have no conflicts of interest or industry support to disclose with regard to the current manuscript. This includes financial or personal relationships that inappropriately influence his or her actions within 3 years of the work beginning submitted.

Guarantor

All authors take public responsibility for the content of this work. The data is also available by request to Dr Guilherme Moura Cunha, MD, the scientific guarantor of this publication.

Statistics and biometry

One of the authors (KH) has significant statistical expertise.

Informed consent

Written informed consent was waived by the Institutional Review Board.

Ethical approval

Institutional Review Board approval was obtained.

Study subjects or cohorts overlap

Does not apply.

Methodology

Retrospective, cross sectional observational study performed at one institution

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.ejrad.2020.108837.

Methods—We retrospectively identified EOB-enhanced MRI-HBP series from examinations performed 2011 2018 (internal and external datasets). Our algorithm, comprising a liver segmentation and classification CNN, produces an adequacy score. Two abdominal radiologists independently classified series as adequate or suboptimal. The consensus determination of HBP adequacy was used as ground truth for CNN model training and validation. Reader agreement was evaluated with Cohen s kappa. Performance of the algorithm was assessed by receiver operating characteristics (ROC) analysis and computation of the area under the ROC curve (AUC). Potential examination duration reduction was evaluated descriptively.

Results—1408 HBP series from 484 patients were included. Reader kappa agreement was 0.67 (internal dataset) and 0.80 (external dataset). AUCs were 0.97 (0.96–0.99) for internal and 0.95 (0.92 96) for external and were not significantly different from each other (p = 0.24). 48 % (50/105) examinations could have been shorter by applying the algorithm.

Conclusion—A proposed CNN-based algorithm achieves higher than 95 % AUC for classifying HBP images as adequate versus suboptimal. The application of this algorithm could potentially shorten examination time and aid radiologists in recognizing technically suboptimal images, avoiding diagnostic pitfalls.

Keywords

Liver; Magnetic resonance imaging; Gd-EOB-DTPA

1. Introduction

Gadolinium ethoxybenzyl diethylenetriaminepentaacetic acid (Gd-EOB-DTPA)-enhanced MRI (EOB-MRI) provides high sensitivity and specificity for detection and characterization of liver lesions [1 3]. Most malignant lesions appear dark compared to the hyperintense background liver during the hepatobiliary phase (HBP), increasing sensitivity. During the HBP, the liver peaks in enhancement/intensity while intrahepatic blood vessels become hypointense due to clearance of contrast from the vascular space. Peak HBP enhancement can occur between 10 60 min following injection, depending on patient characteristics [4]. Hepatocellular enhancement may be impaired or delayed by liver dysfunction, reduced number of hepatocytes (i.e. replacement of liver parenchyma by fibrosis), severe cholestasis, among other factors [5,6]. Conversely, with normal or near normal hepatocellular function, peak enhancement may occur as early as 10 min. For convenience, in clinical practice, HBP images are commonly acquired around 20 min following contrast injection [4,7 9]. However, routine 20-minute HBP delay may be inefficient and unnecessary since adequate HBP enhancement can occur earlier in a substantial proportion of patients.

Adequate HBP imaging is necessary to achieve the added value of EOB-MRI and is defined as liver parenchyma unequivocally brighter than hepatic vasculature and spleen. When these criteria are not met, the Liver Imaging Reporting and Data System (LI-RADS) classifies the HBP as suboptimal [10]. Determination of adequate versus suboptimal HBP is important clinically as assessment of features may be unreliable with suboptimal HBP [11]. Automated determination of HBP adequacy could help to improve not only diagnostic interpretation but also workflow efficiency. In patients with adequate HBP enhancement prior to 20 min, the

examination could be terminated earlier, thereby improving throughput and patient comfort [4,7,12].

However, real-time assessment of HBP adequacy by a radiologist at the scanner suite to optimize imaging protocol is neither feasible nor time efficient. Ideally this process could be automated. Convolutional neural networks (CNNs) have shown enormous potential for automating radiological processes [13,14]. A CNN capable of assessing HBP adequacy in real-time could allow termination of the exam by the technologist when adequate HBP imaging is achieved tailoring the delay to the liver hepatocellular uptake. Additionally, it may aid radiologists in recognizing technically suboptimal images, avoiding diagnostic pitfalls.

The purpose of this study was to develop and evaluate the performance of a fully-automated algorithm, comprising a liver segmentation CNN and an HBP classification CNN, to evaluate HBP adequacy in patients undergoing EOB-MRI. Secondarily, we explored the potential impact of the proposed CNN algorithm to reduce examination duration by applying the algorithm to multiphasic diagnostic EOB-MRI examinations.

2. Materials and methods

This HIPAA-compliant retrospective dual-center study was approved by both institutional review boards with waived requirement for written informed consent.

2.1. Internal imaging data

We selected HBP series from patients with chronic liver disease who underwent liver EOB-MRI at our tertiary care institution for hepatocellular carcinoma (HCC) screening or diagnosis, using an abbreviated EOB-MRI (EOB-AMRI) protocol [15] or a multiphasic EOB-MRI protocol, respectively, from January 2011 to February 2018, on 1.5 T or 3 T (GE Medical Systems, WI, USA). For EOB-AMRIs, Gd-EOB-DTPA (0.025 mmol/kg) was injected using a peripheral butterfly IV. One HBP acquisition was routinely obtained about 20 min or later; additional HBP acquisitions were obtained 5 10 min later in 172 exams. For the multiphasic studies, the same dose of contrast was administered intravenously at a rate of 1 ml/second followed by a saline bolus. Post-contrast acquisitions included one to six acquisitions in the arterial phase, one in the portal venous phase (60 80 seconds), one to four in the transitional phase (2 5 min), and one or more HBP acquisitions every 5 min from 15 min. Inclusion criterion was: patients with at least one HBP image (10 min or later following contrast injection). Imaging protocols and parameters are listed in Table 1. In both protocols, among patients with more than one HBP series, acquisition intervals were approximately every 5 min. No exclusion criteria were applied. Patients and corresponding series were randomly partitioned into training and validation sets using a 70/30 split. Patients included in the training set were not included in the validation set.

2.2. External imaging data

HBP series were extracted from diagnostic examinations of consecutive patients undergoing EOB-MRI at 1.5 T and 3 T (Siemens Healthcare, Erlangen, Germany) at an outpatient imaging center from a different country *[name redacted during submission]* from September

to November 2018 (Fig. 1). Patients had various clinical indications for liver MRI, including patients without known chronic liver disease, for focal lesion characterization. A predosed syringe with 10 ml (0.25 mmol/ml) of Gd-EOB-DTPA-based contrast was infused by peripheral IV at a rate of 1 ml/second. Imaging parameters are listed in Table 1. Inclusion criterion was the same as for the internal dataset and no exclusion criteria were applied. Patient information, series numbers and acquisition time information were withheld due to anonymization during image transfer.

2.3. Image analysis and reference standard by expert readers

Series were prepared for research analysis by removing field strength and acquisition delay information. Series were randomized by patient and acquisition time so that series from the same patient were not reviewed consecutively. Two abdominal imaging fellowship-trained radiologists [initials withheld for review] with 5 and 10 years of experience in liver imaging independently classified each HBP series as A) adequate or B) suboptimal. Images were classified as adequate if the liver was unequivocally more hyperintense than hepatic blood vessels according to LI-RADS [10] and suboptimal if this criterion was not met. Discordant classifications were adjudicated in consensus. Consensus determination of HBP adequacy was used as ground truth for CNN model training and validation. Additionally, each reader classified each series as having or not having image degradation related to imaging or motion artifacts severe enough to obscure liver contours or internal anatomical structures.

2.4. Algorithm development and training

The algorithm comprised two components (Fig. 2). First, HBP series are propagated through a 2D liver segmentation CNN with U-net architecture to produce masks containing liver intensities [16]. The 10 slices containing the largest liver mask areas are then propagated through an HBP-CNN classification network, which produces a single continuous adequacy score between 0 (absolute certainty of suboptimal HBP) and 1 (absolute certainty of adequate HBP). Algorithm and training details are described in Supplementary Materials.

2.5. Algorithm validation

Algorithm validation was performed on both internal and external datasets. Individual CNN adequacy scores were compared against the reference standard and performance was evaluated. Model predictions were used to investigate model accuracy, sensitivity (of suboptimal HBP), and specificity on the validation datasets by selecting the threshold that enforced a 95 % sensitivity on the internal training dataset (selected to minimize false negatives, i.e., suboptimal series falsely classified as adequate) using consensus reader assessment of HBP as the ground truth.

2.6. Exploring potential impact of algorithm to shorten examinations

As an exploratory aim, we assessed the potential of the algorithm to shorten examinations by retrospectively applying it to all examinations containing at least two HBP acquisitions in the internal validation dataset. CNN-generated HBP adequacy scores using the threshold that achieved 95 % sensitivity for suboptimal HBP on the training dataset were calculated. For each examination, the potential reduction in examination duration was calculated by the time

between the first series classified as adequate and the last acquired series within the same examination. For example, an examination with series acquired at 15, 20 and 27 min after contrast administration with suboptimal, adequate, and adequate classifications, respectively, could potentially shorten the examination by 7 min.

2.7. Identifying how the CNN determines HBP classification

Rectified saliency maps were used to visually identify features most influential for HBP classification [17]. Saliency maps are heatmap visualizations that place higher values on image regions with greater contribution to CNN prediction.

2.8. Statistical analysis

Statistical analyses were performed by a biostatistician (initials redacted) using R-v3.4.0 software. Reader agreement for HBP classification and presence of image degradation was assessed using descriptive statistics and Cohen s kappa with the following levels of agreement: none to weak (0.59), moderate (0.6–0.79), strong (0.8) [18]. Performance of the HBP-CNN was evaluated using area under the ROC curve (AUC). Accuracies, sensitivities, and specificities were calculated for the validation datasets (internal and external) using the threshold that enforced 95 % sensitivity for suboptimal HBP on the training dataset. AUCs for internal and external validation datasets were compared using bootstrap. Influence of patient characteristics, presence of image degradation as determined by at least one reader, and imaging acquisition parameters on classification error was assessed for the internal dataset using two multivariate logistic random effects models: 1) adequate and 2) suboptimal classes, separately. Mixed effects models containing significant characteristics were determined by backward elimination. 95 % confidence intervals (CIs) were analytically calculated as appropriate. The potential to reduce examination duration was assessed descriptively.

3. Results

3.1. Internal validation cohort

Training and validation demographics are provided in Table 2. 406 patients were selected for the internal dataset. 178 patients had multiple examinations, leading to 729 examinations and 1201 individual HBP series. 826 HBP series from 284 patients were used for CNN training and remaining 375 series from 122 patients for CNN validation (Fig. 1).

3.2. External validation cohort

For external validation, 207 3D T1-weighted HBP series from 78 patients were included.

3.3. Reader scores for HBP adequacy

Readers classified as suboptimal 312/1201 and 216/1201 series for the internal dataset and 43/207 and 34/207 series, for the external dataset, respectively. Of the total, 89 % (1063/1201) and 94 % (194/207) of series were classified the same by both readers for the internal and external datasets. Following consensus, 299/1201 and 40/207 series were classified as suboptimal for the internal and external datasets. Cohen s kappa between reader

adequacy classification for the internal and external datasets were 0.67 ($p < 0.001$) and 0.80 ($p < 0.001$). Readers classified 223/1201 (19 %) and 268/1201 (22 %) series as having image degradation in the internal dataset and 45/207 (22 %) and 51/207 (25 %) in the external dataset, respectively. Reader agreement for presence of image degradation was 0.56 ($p < 0.001$) and 0.30 ($p < 0.001$), respectively.

3.4. Algorithm performance

Performance metrics of the algorithm for assessment of HBP adequacy in the internal and external validation datasets are summarized in Table 3. ROC curves are shown in Fig. 3. AUCs for the internal and external datasets were 0.973 (95 % CI 0.960–0.986) and 0.952 (95 % CI 0.919–0.985), respectively. The internal dataset AUC was not significantly different from the external dataset AUC ($p = 0.24$). The threshold enforcing 95 % sensitivity for suboptimal HBP on the internal training dataset was 0.87. Using this threshold, accuracies, sensitivities and specificities were [82.7 % (310/375), 100 % (114/114) and 75.1 % (196/261)] for the internal validation dataset and [93.2 % (193/207), 85.0 % (34/40) and 94.2 % (159/167)] for the external validation dataset. Performance of the individual segmentation component of the algorithm was not in the scope of this current work and has been described elsewhere [16].

3.5. Influence of Patient/imaging characteristics on algorithm performance

Influences of patient and image characteristics on algorithm classification error are summarized in Table 4. Following backward elimination, presence of image degradation judged by at least one radiologist ($p < 0.001$) and larger frequency matrix size ($p = 0.003$) were the only independent characteristics that significantly affected adequate classification accuracy. Field strength ($p = 0.002$), slice thickness ($p < 0.001$), and frequency matrix size ($p < 0.001$) were the only independent characteristics that significantly affected suboptimal classification accuracy.

3.6. Potential of algorithm to shorten examinations

105 examinations from 76 patients in the internal validation dataset had at least two HBP acquisitions. The 0.87 HBP score threshold from the sensitivity analysis was used for classification. Of the 105 examinations, 50 (48 %) could have been shortened by real-time application of the CNN algorithm, including 7 examinations that could have been shortened by >10 min. A cumulative total of about 220 min of examination time could have been eliminated. Example applications of the HBP-CNN to three separate examinations are shown in Fig. 4.

3.7. Saliency map analysis

Saliency maps for two HBP series correctly classified as adequate and suboptimal are shown in Fig. 5. The CNN focused on the edges of vessels for adequate predictions. Predictions for the suboptimal class had little or no activation across the liver, suggesting suboptimal HBP classification is determined by unclear vessel edges in the image.

4. Discussion

In this study, we developed and evaluated the performance of a fully-automated CNN-based algorithm to evaluate HBP adequacy. The model was validated on 582 individual HBP series (375 internal series and 207 external series) from 200 unique patients. AUCs were 0.973 and 0.952 for the internal and external datasets, respectively. We additionally explored the potential of the CNN algorithm to shorten examinations. Using a threshold enforcing high sensitivity for suboptimal HBP, we found 48 % of patients in the internal cohort achieved adequate HBP at least one acquisition earlier than the last HBP series acquired, and a total of ~220 min of examination time may have been unnecessary. In clinical practice, EOB-MRI protocols often insert additional sequences between portal venous phase acquisitions and the standard 20 min delay, which may include additional breath-hold T1-weighted images [19]. If implemented at a console level, our algorithm could allow for a reduction of examination time by sending immediate feedback on HBP adequacy to the technologist. This would have reduced examination time in up to half of all patients in our cohort. Additionally, HBP adequacy classification could be informative to radiologists in the reading room when applying diagnostic criteria specific to this imaging phase.

CNNs have been applied to quickly and automatically classify medical images aiming to optimize radiology workflow [13,20 24]. Lee has proposed the use of a CNN framework for MRI protocol optimization [25]. Although Lee s work was based on clinical data and not image classification, it demonstrates the potential of CNNs to improve radiologic decisions. Esses et al. proposed a CNN approach to assess image quality by automatically classifying T2-weighted liver MR images as diagnostic and nondiagnostic [14]. However, the CNN had a high false positive rate. Most false-positive classifications by the CNN were retrospectively attributed to artifacts outside the liver; while these artifacts caused the CNN to deem images nondiagnostic, the artifacts were appropriately ignored by the radiologists. Our algorithm is less susceptible to such errors because it incorporates a segmentation CNN that constrains the classification task to the liver area.

The algorithm was robust against sex, age, BMI, liver etiology, field-of-view, and phase matrix size when classifying series as adequate or suboptimal. Images degraded by artifacts did not significantly affect accuracy of the algorithm for classifying series as suboptimal series but reduced accuracy of the algorithm for classifying series as adequate. Based on visual assessment of the saliency maps, we speculate that on degraded images the conspicuity and contrast of boundaries is reduced, causing a liver with adequate uptake to resemble a liver with suboptimal uptake. However, clinically, the misclassification of adequate series with motion artifacts as suboptimal may be acceptable since it may lead to repeat acquisitions potentially less impacted by artifact. Additional factors that significantly affected suboptimal classification included field strength (1.5 T) possibly reflecting lower SNR and larger slice thickness with volume averaging in the z-direction. Larger frequency matrix size was associated with lower accuracy for both adequate and suboptimal classification. Although it is unclear why larger frequency matrix size reduces accuracy, AUCs stratified across significant characteristics all exceeded 0.95, indicating their effect on prediction accuracy is likely not clinically impactful.

Our CNN was trained and validated using radiologist consensus of HBP adequacy as a reference standard. An objective HBP adequacy score based on manual calculations of liver-to-portal vein signal ratio have been proposed by Bashir et al. also using radiologist assessment as a reference standard [26]. When applying this criteria, EOB-MRI examinations in patients without chronic liver disease could have been terminated earlier than 20 min in up to 56 % of that population [26]. Another study has shown that in up to 61 % of patients with chronic liver disease, lesion detection rates did not differ when acquiring HBP images at 10 or 20 min [7]. When searching for a diagnostic threshold most studies weigh both sensitivity and specificity, usually using the Youden index [27]. However, we exploratorily favored sensitivity over specificity since suboptimal series misclassified as adequate would lead to termination of the examination prior to adequate HBP acquisition, thereby negatively impacting clinical diagnostic accuracy. When applying this high sensitivity threshold for classification, we found up to 48 % of patients achieved adequate HBP uptake at least one acquisition earlier than the last HBP series. This is slightly lower compared to the study by Bashir et al., likely due to the exclusion of patients with chronic liver disease in that study. In contrast, our population was comprised primarily of patients with chronic liver disease and likely many with some degree of liver function impairment. Furthermore, the previously proposed method relies on manual ROI drawing, a time consuming and somewhat specialized task. We believe the implementation of a CNN-based method to evaluate HBP adequacy is advantageous as it may allow for real-time and fully automated assessment, potentially at the scanner console and during examination.

Our study has limitations. First, despite efforts to overcome reader variability, there is no objective reference standard for adequate versus suboptimal HBP. Nevertheless, reader agreement was moderate to strong for both internal and external datasets and similar to agreement described in the literature [26]. In addition, series numbers and acquisition time information in the external dataset were not available and we were unable to explore how much expendable acquisition time could have been reduced in this population. Additionally, we observed differences in sensitivity and specificity between the internal and external validation datasets. This is probably due to different rates of chronic liver disease between the internal and external cohorts. Further training and research are needed to determine which, if any, universal threshold could be applied to achieve desired sensitivity and specificity in the general population. Finally, as a current limitation and future direction of our work, the clinical benefit of the algorithm will likely be highest if allowing for the prediction of the ideal HBP delay based on information gathered from earlier acquisitions (i.e. pre-contrast, arterial or portal venous phases).

5. Conclusion

In conclusion, our proposed CNN-based algorithm achieves higher than 0.95 AUC for classifying HBP images as adequate or suboptimal. This can potentially reduce overall examination time in approximately 48 % of patients with chronic liver disease. Further prospective studies should be conducted to evaluate the feasibility and performance of applying the algorithm during clinical EOB-MRI examinations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge Diagnósticos da America SA (DASA) for the research collaboration in providing the external imaging dataset

Funding

The authors state that this work has not received any funding.

Abbreviations

Gd-EOB-DTPA	Gadolinium ethoxybenzyl diethylenetriaminepentaacetic acid
CNN	Convolutional Neural Network
HBP	Hepatobiliary Phase
LI-RADS	Liver Imaging Reporting and Data System
AUC	Area Under the ROC Curve

References

[1]. Ito T, et al., The diagnostic advantage of EOB-MR imaging over CT in the detection of liver metastasis in patients with potentially resectable pancreatic cancer, Pancreatology 17 (3) (2017) 451 456. [PubMed: 28298257]

[2]. Suh CH, et al., The diagnostic value of Gd-EOB-DTPA-MRI for the diagnosis of focal nodular hyperplasia: a systematic review and meta-analysis, Eur. Radiol 25 (4) (2015) 950 960. [PubMed: 25537979]

[3]. Costa EAC, et al., Diagnostic accuracy of preoperative gadoxetic acid enhanced 3-T MR imaging for malignant liver lesions by using Ex Vivo MR imaging matched pathologic findings as the reference standard, Radiology 276 (3) (2015) 775 786. [PubMed: 25875972]

[4]. Kessel CS, Veldhui WB, Bosch MAAJ, et al., MR liver imaging with Gd-EOB-DTPA: a delay time of 10 minutes is sufficient for lesion characterization, Eur. Radiol 22 (10) (2012) 2153 2216. [PubMed: 22645040]

[5]. Wu J-W, et al., Optimization of hepatobiliary phase delay time of Gd-EOB-DTPA-enhanced magnetic resonance imaging for identification of hepatocellular carcinoma in patients with cirrhosis of different degrees of severity, World J. Gastroenterol 24 (3) (2018) 415. [PubMed: 29391764]

[6]. Kobi M, et al., Limitations of GD-EOB-DTPA-enhanced MRI: can clinical parameters predict suboptimal hepatobiliary phase?, Clin. Radiol 72 (1) (2017) 55 62. [PubMed: 27842889]

[7]. Motosugi U, Ichikawa T, Tominaga L, et al., Delay before the hepatocyte phase of Gd-EOB-DTPA-enhanced MR imaging: is it possible to shorten the examination time?, Eur. Radiol 19 (11) (2009) 2623 2629. [PubMed: 19471935]

[8]. Tamada T, et al., Gd-EOB-DTPA-enhanced MR imaging: evaluation of hepatic enhancement effects in normal and cirrhotic livers, Eur. J. Radiol 80 (3) (2011) e311 e316. [PubMed: 21315529]

[9]. Sofue K, et al., Gd-EOB-DTPA-enhanced 3.0 T MR imaging: quantitative and qualitative comparison of hepatocyte-phase images obtained 10 min and 20 min after injection for the

detection of liver metastases from colorectal carcinoma, Eur. Radiol 21 (11) (2011) 2336. [PubMed: 21748389]

[10]. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS/CT-MRI-LI-RADS-v2018 Online, accessed in March 17, 2019.

[11]. Schwope RB, May LA, Reiter MJ, Lisanti CJ, Margolis DJ, Gadoxetic acid: pearls and pitfalls, Abdom. Imaging 40 (8 6) (2015) 2012 2029. [PubMed: 25613332]

[12]. Schreiber-Zinaman J, Rosenkrantz AB, Frequency and reasons for extra sequences in clinical abdominal MRI examinations, Abdom. Radiol 42 (1) (2017) 306 311.

[13]. Yamashita R, et al., Convolutional neural networks: an overview and application in radiology, Insights Imaging (2018) 1 19.

[14]. Js. Esses S, et al., Automated image quality evaluation of T2 weighted liver MRI utilizing deep learning architecture., J. Magn. Reson. Imaging 47 (3) (2018) 723 728. [PubMed: 28577329]

[15]. Marks RM, Ryan A, Heba ER, Tang A, Wolfson TJ, Gamst AC, Sirlin CB, Bashir MR, Diagnostic per-patient accuracy of an abbreviated hepatobiliary phase gadoxetic acid enhanced MRI for hepatocellular carcinoma surveillance, Am. J. Roentgenol 204 (3 3) (2015) 527 535. [PubMed: 25714281]

[16]. Withheld During Submission for Blinding Purposes, 2019.

[17]. Raghavendra Kotikalapudi contributors. keras-vis. GitHubhttps://github.com/raghakot/keras-vis2017Accessed on January 03, 2019

[18]. McHugh ML, Interrater reliability: the kappa statistic, Biochemia Medica: Biochemia Medica 22 (10 3) (2012) 276 282. [PubMed: 23092060]

[19]. Cruite I, Schroeder M, Merkle EM, Sirlin CB, Gadoxetate disodium enhanced MRI of the liver: part 2, protocol optimization and lesion appearance in the cirrhotic liver, Am. J. Roentgenol 195 (7 1) (2010) 29 41. [PubMed: 20566795]

[20]. Kuestner T, et al., A machine-learning framework for automatic reference-free quality assessment in MRI., Magn. Reson. Imaging 53 (2018) 134 147. [PubMed: 30036653]

[21]. Brown AD, Marotta TR, Using machine learning for sequence-level automated MRI protocol selection in neuroradiology, J. Am. Med. Inform. Assoc 25 (5) (2017) 568 571.

[22]. Zhang C, et al., Accelerated simultaneous multi-slice MRI using subject-specific convolutional neural networks, 2018 52nd Asilomar Conference on Signals, Systems, and Computers, IEEE, 2018.

[23]. Lv J, et al., Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study, Br. J. Radiol 91 (2018) xxxx 20170788.

[24]. Bahrami N, et al., Automated selection of myocardial inversion time with a convolutional neural network: spatial temporal ensemble myocardium inversion network (STEMI NET)., Magn. Reson. Med (2019).

[25]. Lee YH, Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks, J. Digit. Imaging 31 (10 5) (2018) 604 610. [PubMed: 29619578]

[26]. Bashir MR, Breault SR, Braun R, Do RK NRC, Reeder SB, Optimal timing and diagnostic adequacy of hepatocyte phase imaging with gadoxetate-enhanced liver MRI, Acad. Radiol 21 (6 6) (2014) 726 732. [PubMed: 24717550]

[27]. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH, Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions, Clin. Chem 54 (4 4) (2008) 729 737. [PubMed: 18258670]

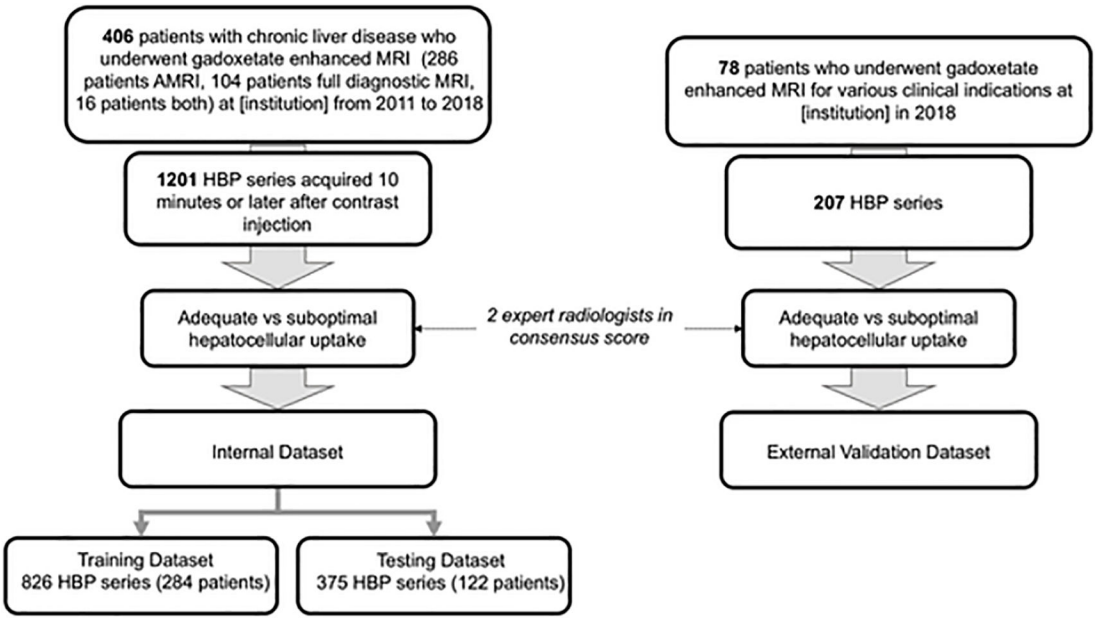

Fig. 1.

Imaging data - The internal dataset comprises HBP series of 253 patients with chronic liver disease acquired 10 min or later after contrast injection at 1.5 and 3 T (GE Medical Systems, WI, USA) at our tertiary care institution for hepatocellular carcinoma (HCC) screening or diagnosis. The external validation dataset comprises nominal HBP series of 78 patients with various indications for gadoxetate enhanced MRI, including patients without chronic liver disease for focal lesion characterization, from an outside institution from another country. Due to full exam anonymization, demographics or time delay information were not available.

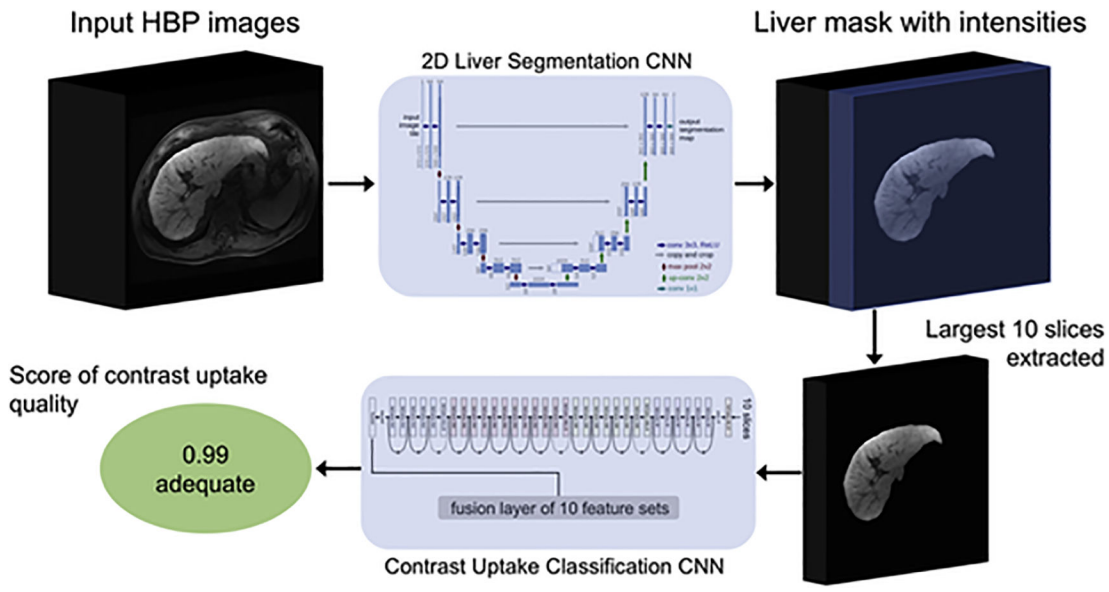

Fig. 2.

Diagram of the proposed CNN framework for contrast uptake classification. HBP images are propagated through an independently developed 2D liver segmentation CNN to produce a liver mask populated with intensities. The largest 10 liver slices of the liver mask image are then sent to a contrast uptake classification network to produce a single score for adequate or suboptimal HBP.

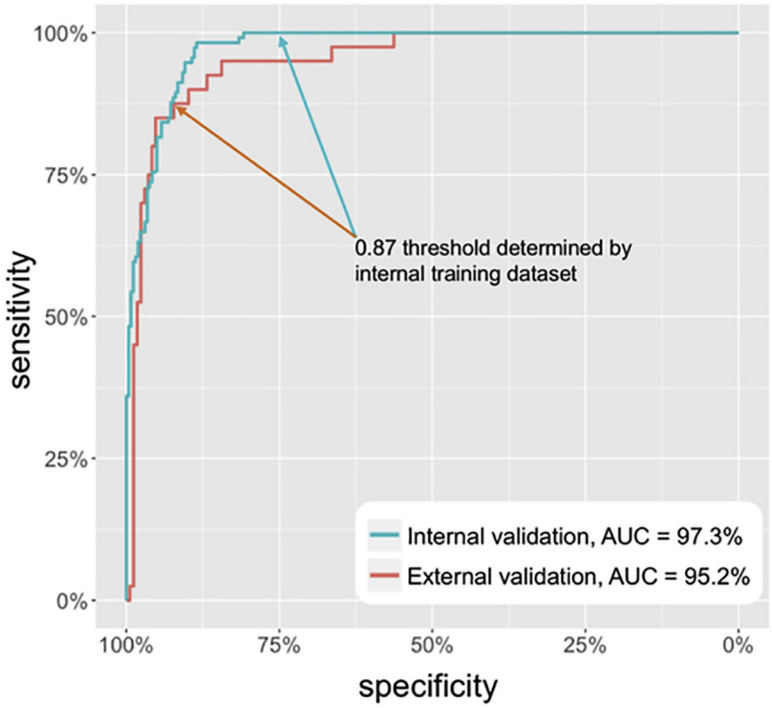

Fig. 3.

ROC curves for the internal and external validation sets. Arrows show the sensitivities and specificities of the validation datasets using a 0.87 threshold determined by enforcing a 95 % sensitivity for suboptimal HBP on the internal training dataset. Using this threshold, sensitivities and specificities were [100 % (114/114) and 75.1 % (196/261)] for the internal validation dataset and [85.0 % (34/40) and 94.2 % (159/167)] for the external validation dataset.

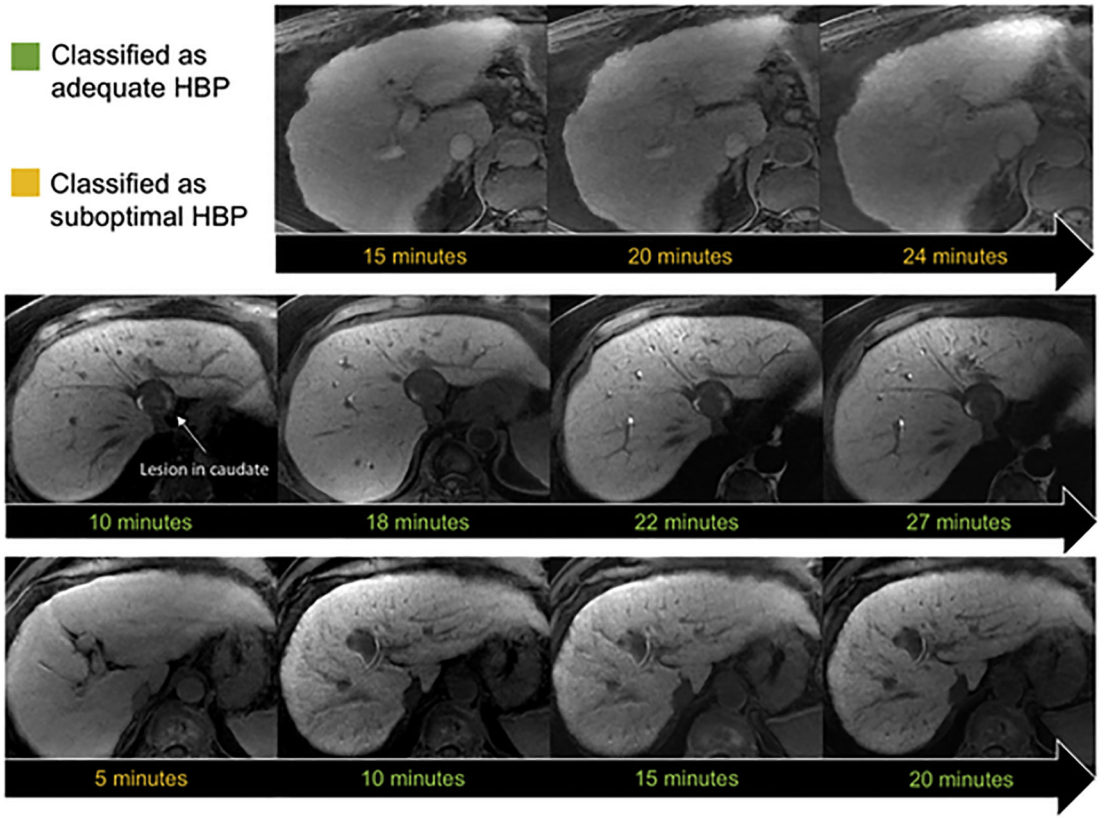

Fig. 4.

Tl-weighted post-contrast MR images. **Top row:** Patient with chronic liver disease and poor liver contrast uptake. Model correctly classifies all HBP images as suboptimal. All series were scored as suboptimal and the exam defaults to the standard HBP delay. **Middle row:** Patient with preserved liver function. Adequate HBP is identified as early as 10 min. Examination time could potentially be reduced by 17 min. **Bottom row:** As proof of concept, model was applied to all post contrast series in a dynamic diagnostic study: model correctly classified images as suboptimal HBP at 5 min and as adequate HBP at 10 min. Examination time could potentially be reduced from 20 min to 10 min, a 10-minute reduction.

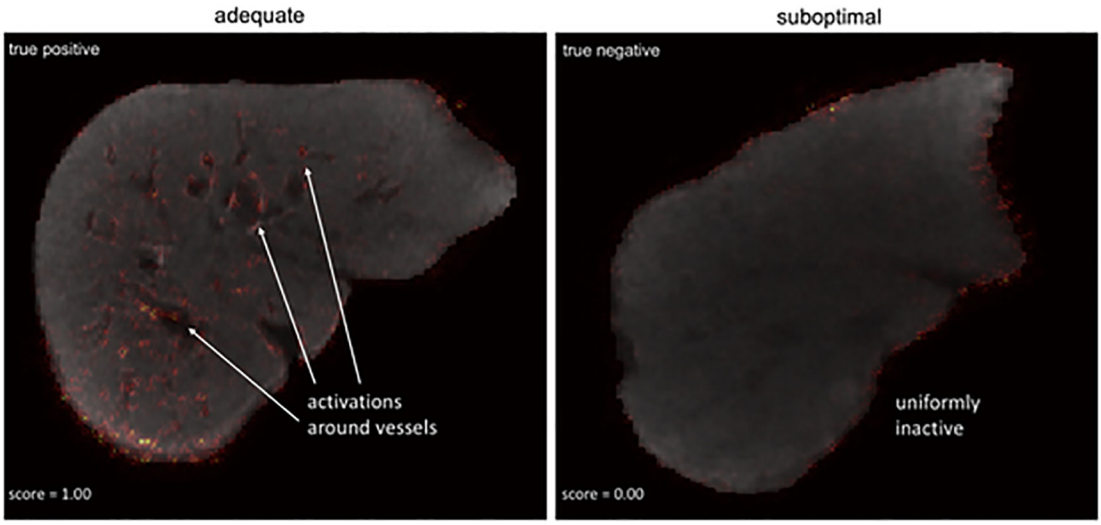

Fig. 5.

Saliency maps highlighting anatomical and structural features most influential for HBP classification. Red areas indicate regions within the liver that most influenced the CNN prediction. Saliency maps showed activation where contrast between vessels and background liver parenchyma is pronounced when classifying adequate images. There is little or no activation for classification of suboptimal images due to the poor contrast between vessels and background liver parenchyma. Interestingly, a high contrast between these structures (hyperintense liver vs hypointense liver vessels) is the visual human criteria to classify HBP images as having adequate hepatocellular contrast uptake.

Table 1

Scanner and imaging parameters for the full-diagnostic and AMRI exams of the internal dataset and the exams of the external dataset.

Characteristic	Full-Diagnostic	AMRI	External Dataset
Protocol			
T1wi Fat Pre-contrast	x		
T1wi IP/OP Pre-contrast	x		
T1wi Arterial Phase Post-contrast	x		
T1wi Portal Venous Phase Post-contrast	x		
T1wi Transitional Phase Post-contrast	x		
T1wi Hepatobiliary Phase Post-contrast*	x	x	x
T1wi Diffusion Weighted Post-contrast	x	x	
T2wi (SSFSE, HASTE)	x	x	
Total Series	408	793	207
Scanner Manufacturer and Model			
GE Discovery MR750w	53	328	/
GE Signa HDxt	355	465	/
GE Optima MR450w	/	/	3
GE Signa Excite	/	/	28
GE Signa Explorer	/	/	5
Philips Ingenia	/	/	2
Siemens Avanto	/	/	165
Siemens Espree	/	/	3
Siemens Verio	/	/	1
Field Strength			
1.5 T	127	132	206
3.0T	281	661	1
***Hepatobiliary Phase (HBP) Imaging Parameters**			
Repetition Time (ms)	3.47 ± 0.70 (2.52 6.05)	4.00 ± 0.87 (2.86 6.73)	3.76 ± 0.80 (2.63 6.56)
Echo Time (ms)	1.57 ± 0.33 (1.18 3.13)	1.70 ± 0.37 (1.24 3.13)	1.39 ± 0.35 (1.01 3.13)
Field of View (mm)	409.75 ± 30.73 (340 480)	410.59 ± 34.00 (300 500)	392.83 ± 38.07 (275 500)
Flip Angle (degrees)	14.94 ± 1.63 (6 25)	15.56 ± 2.70 (11 25)	17.82 ± 8.18 (10 30)

Characteristic	Full-Diagnostic	AMRI	External Dataset
Pixel Bandwidth (Hz)	415.45 ± 122.48 (122.07 651.06)	463.96 ± 155.26 (122.07 1302.11)	637.10 ± 232.17 (244.14 1028.00)
Pixel Spacing (mm)	0.79 ± 0.09 (0.37 1.41)	0.84 ± 0.19 (0.41 1.95)	1.10 ± 0.27 (0.63 1.56)
Slice Spacing (mm)	2.20 ± 0.36 (1.10 3.00)	2.40 ± 0.47 (2.00 4.40)	2.36 ± 0.72 (1.50 4.00)
Slice Thickness (mm)	4.37 ± 0.73 (2.20 8.00)	4.75 ± 0.83 (3.00 6.00)	2.67 ± 0.75 (1.60 5.00)
Matrix	(256 384) × (160 320)	(256 384) × (128 256)	(192 324) × (146 256)

AMRI = Abbreviated MRI.

Table 2

Demographic summary of the training and validation sets in the internal dataset.

	Training Dataset			Internal Validation Dataset		
Characteristic	**# Series**	**# Exams**	**# Patients**	**# Series**	**# Exams**	**# Patients**
Overall	826	513	284	375	216	122
Gender						
F	388	238	133	150	91	49
M	438	275	151	225	125	73
Age (years) *	58.49 ± 12.50 (13 84)			59.97 ± 10.80 (33 84)		
<45 (ref)	107	65	39	35	19	11
45–64	462	283	164	229	132	73
>65	257	165	81	111	65	38
BMI (kg/m ^{2}) *	28.85 ± 5.86 (14.6 53.2)			28.83 ± 5.74 (17.30 41.70)		
<24.9 (ref)	233	141	81	88	51	34
25–30	271	167	94	158	86	44
>30	317	203	107	129	79	44
missing	5	2	2	0	0	0
Etiology						
HBV	126	74	34	49	29	14
HCV	347	213	114	206	113	65
Alcohol	94	67	49	46	25	12
NAFLD	130	83	40	32	20	13
Autoimmune	38	22	14	9	6	3
Hepatitis						
Other	91	54	33	33	23	15
Uptake						
Adequate	641	/	/	261	/	/
Suboptimal	185	/	/	114	/	/

* Mean ± standard deviation; range in parentheses.

BMI = Body Mass Index, HCV = Hepatitis C Virus, HBV = Hepatitis B Virus, NAFLD = Non-Alcoholic Liver Disease.

Table 3

CNN performance metrics for the internal and external validation datasets using a 0.87 threshold determined by enforcing a 95 % sensitivity for suboptimal HBP on the internal training dataset. Numbers in parentheses are 95 % confidence intervals.

Dataset	Sample Size (adequate / suboptimal)	AUC (%)	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
Internal Validation Dataset	261 / 114 69.6 % / 30.4 %	97.3 (96.0, 98.6)	0.87	82.7 (78.4, 86.4)	100 (96.8, 100)	75.1 (69.4, 80.2)
External Validation Dataset	167 / 40 80.7 % / 19.3 %	95.2 (91.9, 98.5)	0.87	93.2 (88.9, 96.3)	85.0 (70.2, 94.3)	95.2 (90.8, 97.9)

AUC = Area Under the Curve; Number in parentheses indicate confidence interval.

Table 4

Multivariate logistic random effects models to assess the influence of patient and imaging characteristics on prediction accuracy for the adequate and suboptimal classes.

			Adequate Multivariate Analysis*		Suboptimal Multivariate Analysis‡	
Characteristic	**# Series Adequate/Suboptimal**	**AUC (%)**	**Odds Ratio (95 % CI)**	**P-Value**	**Odds Ratio (95 % CI)**	**P-Value**
Gender						
F	120/30	96.2 (93.9, 98.5)	1		1	
M	141/84	98.4 (97.0, 99.7)	0.98 (0.17, 5.74)	0.99	0.35 (0.06, 2.37)	0.37
Age						
<45	29/6	98.3 (96.6, 99.9)	1		1	
45 64	145/84	96.4 (93.7, 99.2)	7.51 (0.38, 148.56)	0.21	8.64 (0.16, 420.45)	0.39
>65	87/24	97.5 (95.3, 99.6)	6.97 (0.30, 165.55)	0.26	1.73 (0.02, 123.42)	0.84
BMI (kg/m ^{2})						
<24.9	58/30	96.4 (93.7, 99.2)	1		1	
25 30	104/54	97.5 (95.3, 99.6)	0.13 (0.01, 1.09)	0.08	0.8 (0.13, 4.89)	0.84
>30	99/30	98.3 (96.6, 99.9)	0.2 (0.02, 1.74)	0.17	2.17 (0.28, 15.77)	0.54
Etiology						
HCV	130/76	99.3 (98.0, 100)	1		1	
HBV	47/2	98.3 (96.0, 100)	0.15 (0.01, 1.69)	0.15	4.32 (0.06, 285.79)	0.58
Alcohol	21/25	98.3 (95.8, 100)	36.86 (1, 1356.77)	0.07	0.15 (0.03, 1.03)	0.11
NAFLD	29/3	95.3 (90.7, 100)	1.04 (0.08, 13.70)	0.98	4.84 (0.02, 2258.01)	0.66
Autoimmune Hepatitis	6/3	97.9 (94.8, 100)	10.85 (0.08, 1524.2)	0.37	14.91 (0.17, 1047.73)	0.33
Other	28/5	94.9 (90.2, 99.7)	0.06 (0.00, 0.90)	0.06	1.04 (0.06, 23.16)	0.98
Image Degradation						
No Degradation	222/49	98.4 (97.0, 99.7)	1		1	
Degradation	39/65	96.2 (93.9, 98.5)	**34.95 (13.35, 87.46)**	**<0.001**	1.19 (0.60, 2.36)	0.62
Field Strength						
1.5 T	40/37	98.4 (97.0, 99.7)	1		1	
3 T	221/77	96.2 (93.9, 98.5)	3.09 (0.60, 15.75)	0.19	**0.07 (0.02, 0.45)**	**0.002**
Field of View (mm)						
<400	67/38	96.2 (93.9, 98.5)	1		1	

			Adequate Multivariate Analysis*		Suboptimal Multivariate Analysis‡	
Characteristic	**# Series Adequate/Suboptimal**	**AUC (%)**	**Odds Ratio (95 % CI)**	**P-Value**	**Odds Ratio (95 % CI)**	**P-Value**
>=400	194/76	98.4 (97.0, 99.7)	0.45 (0.19, 1.12)	0.09	2.63 (0.98, 5.62)	0.04
Slice Thickness (mm)						
<4	100/37	97.5 (95.3, 99.6)	1		1	
4 5	103/60	98.3 (96.6, 99.9)	1.3 (0.50, 3.32)	0.60	0.79 (0.28, 3.01)	0.71
>5	58/17	96.4 (93.7, 99.2)	0.47 (0.13, 1.77)	0.27	**50.15 (3.85, 253.62)**	**<0.001**
Matrix - Frequency						
<300	88/53	98.4 (97.0, 99.7)			1	
>=300	173/61	96.2 (93.9, 98.5)	**3.13 (0.98, 9.28)**	**0.05**	**13.29 (3.72, 38.51)**	**<0.001**
Matrix - Phase						
<200	145/68	98.4 (97.0, 99.7)			1	
>=200	116/46	96.2 (93.9, 98.5)	0.68 (0.24, 2.01)	0.48	0.43 (0.15, 1.28)	0.1404

*Artifacts and matrix size (bold) significantly affected adequate classification after stepwise variable selection.

‡Field strength, slice thickness, and frequency matrix size (bold) affected suboptimal classification after stepwise variable selection.

AUC = Area Under the Curve, BMI = Body Mass Index, HCV = Hepatitis C Virus, HBV = Hepatitis B Virus, NAFLD = Non-Alcoholic Liver Disease; Numbers in parentheses are confidence intervals.