Published in final edited form as:

J Vasc Interv Radiol. 2018 June ; 29(6): 850–857.e1. doi:10.1016/j.jvir.2018.01.769.

Predicting Treatment Response to Intra-arterial Therapies of Hepatocellular Carcinoma using Supervised Machine Learning—An Artificial Intelligence Concept

Aaron Abajian, M.D.¹, Nikitha Murali, BA¹, Lynn Jeanette Savic, M.D.^{1,2}, Fabian Max Laage-Gaupp, M.D.¹, Nariman Nezami, M.D.¹, James S. Duncan, PhD³, Todd Schlachter, MD¹, MingDe Lin, PhD⁴, Jean-François Geschwind, MD⁵, and Julius Chapiro, MD¹

¹Yale School of Medicine, Department of Radiology and Biomedical Imaging, 333 Cedar Street, New Haven, CT 06520, USA

²Universitätsmedizin Charité Berlin, Department of Diagnostic and Interventional Radiology, Augustenburger Platz 1, 13353 Berlin, Germany

³Yale School of Engineering & Applied Science, Department of Biomedical Engineering, 300 Cedar Street, New Haven, CT 06519, USA

⁴Philips Research North America, 2 Canal Park, 3rd Floor, Cambridge, MA 02141, USA

⁵PreScience Labs, Westport, CT, USA

Abstract

Purpose—To use magnetic resonance (MR) imaging and clinical patient data to create an artificial intelligence (AI) framework for the prediction of therapeutic outcomes of trans-arterial chemoembolization (TACE) by applying machine learning (ML) techniques.

Methods—This study included 36 patients with hepatocellular carcinoma (HCC) treated with TACE. The cohort (age 62 ± 8.9 years [mean $\pm$ standard deviation]) contained 31 males; 13 white;

Corresponding Author: Julius Chapiro, MD, Department of Radiology and Biomedical Imaging, Yale University School of Medicine, 300 Cedar Street, TAC N312A, New Haven, CT 06520, julius.chapiro@yale.edu, Fax: (203) 785-3024.

Disclosures:

A.A. No relevant relationships to disclose. *N.M.* No relevant relationships to disclose. *L.J.S.* reports grants from the National Institutes of Health (NIH/NCI R01CA206180). Activities not related to the present article: reports grants from Leopoldina Postdoctoral Fellowship, has received grants from Rolf W. Guenther Foundation of Radiological Sciences. *F.M.L.G.* No relevant relationships to disclose. *N.N.* No relevant relationships to disclose. *J.S.D.* reports grants from the National Institutes of Health (NIH/NCI R01CA206180) and Philips Healthcare.

T.S. No relevant relationships to disclose. *M.L.* reports grants from the National Institutes of Health (NIH/NCI R01CA206180) and Philips Healthcare. *J.F.G.* reports grants from the National Institutes of Health (NIH/NCI R01CA206180) and Philips Healthcare. Activities not related to the present article: has received grants from the BTG, Boston Scientific, Guerbet Healthcare; has received personal fees from Guerbet Healthcare, BTG, Threshold Pharmaceuticals, Boston Scientific, and Terumo; is a consultant for Prescience Labs. *J.C.* reports grants from the National Institutes of Health (NIH/NCI R01CA206180) and Philips Healthcare. Activities not related to the present article: has received scholarships from the Rolf W. Guenther Foundation of Radiological Sciences and the Charité Berlin Institute of Health Clinical Scientist Program; has received grants from the German-Israeli Foundation for Scientific Research and Development.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Eastern Cooperative Oncology Group Performance Status: 0–24, 1–10, 2–2; Child-Pugh: A–31, B–4, C–1; Barcelona Clinic Liver Cancer Stage: 0–1, A–12, B–10, C–13; tumor size 5.2 ± 3.0 cm; number of tumors 2.6 ± 1.1 ; TACE: 30 conventional, 6 with drug-eluting embolic agents. MR imaging was obtained prior to and one month after TACE. Image-based tumor response to TACE was assessed using the 3D quantitative European Association for the Study of the Liver (qEASL) criteria. Clinical information, baseline imaging, and therapeutic features were used to train logistic regression (LR) and random forest (RF) models to predict patients as treatment responders or non-responders. The performance of the models was compared to treatment response as determined by qEASL response criteria using leave-one-out cross-validation.

Results—Both LR and RF models predicted TACE treatment response with an overall accuracy of 78% (sensitivity 62.5%, specificity 82.1%, positive predictive value 50.0%, negative predictive value 88.5%). The strongest predictors of treatment response included a clinical variable (presence of cirrhosis) and an imaging variable (tumor signal intensity >27.0).

Conclusions—TACE outcomes in patients with HCC may be predicted pre-procedurally by combining clinical patient data with baseline MR imaging using AI and ML techniques.

Introduction

Trans-arterial chemoembolization (TACE) is primary well-established therapy used to treat patients with unresectable hepatocellular carcinoma (HCC) (1, 2). Radiologic response criteria are used to quantify TACE efficacy based on post-treatment contrast-enhanced Magnetic Resonance (MR) or Computed Tomography (CT) imaging. Conventional criteria used to assess radiologic response after TACE are measured on two-dimensional axial slices, based on diameter changes in target tumors or visually estimated changes in image enhancement (3–6). The quantitative European Association for the Study of the Liver (qEASL) response criteria measures the degree of change in three-dimensional enhancing tumor volume and has been demonstrated to be superior to all other response criteria both in reproducibility and ability to predict overall survival sooner (7).

While these improvements in radiologic response measures have advanced assessment of patients after treatment, it remains clinically challenging to predict which patients will respond to TACE prior to treatment; no single pre-treatment clinical or imaging feature is predictive of response. An accurate method for predicting a patient's likelihood of response could reduce unnecessary interventions, lower healthcare costs, and minimize patient harm. It is worthwhile, therefore, to investigate how pre-treatment patient characteristics influence treatment efficacy as measured by post-treatment response.

The challenge of applying pre-treatment imaging and clinical traits to predict post-treatment response can be solved using machine learning (ML), an application of artificial intelligence (AI) that self-improves by learning from data (8). Predicting response to treatment can be conceptualized as a classification problem, in which an ML model sorts patients into categories of treatment responders or non-responders using information gathered prior to treatment. Classification is accomplished through supervised ML, a technique that requires outcome-labeled training data. Trained models can be applied to new cases they have not previously encountered (9). For example, patient baseline imaging and clinical data,

treatment characteristics, and treatment outcomes can be applied from a retrospective patient cohort to teach a model to learn the relationships between these variables and treatment outcomes. The model predicts treatment response in new patients pre-procedurally, provided that planned treatment characteristics are specified. The purpose of this study was to use MR imaging and clinical patient data to create an AI framework for the prediction of therapeutic outcomes of TACE by applying ML techniques.

Methods

Patient Cohort

This was a Health Insurance Portability and Accountability Act compliant, single institution, institutional review board approved retrospective study. Informed consent was waived. A cohort of 36 patients with HCC treated with conventional TACE utilizing ethiodized oil or TACE with drug-eluting embolic agents from 2003–2015 was selected for analysis; only TACE-naïve patients were included in the study. The patient cohort consisted of individuals who had non-infiltrative tumors with well-delineated capsule and other classic diagnostic imaging features of HCC according to the Liver Imaging and Reporting System. Only patients with follow-up imaging within 30 days of TACE as well as complete clinical and imaging data were included. The patients were not consecutive chronologically. In patients where multiple tumors were treated in the initial TACE procedure, all target tumors were included in the qEASL response analysis. Patients with HCC treated with concomitant Sorafenib received it continuously, starting one week before the initial TACE and continued in 6-week cycles without interruptions. Clinical and demographic information and treatment characteristics used to train the learning models are reported in Table 1.

TACE Technique

After a multidisciplinary tumor board identified TACE as an appropriate treatment for each patient, TACE was performed by a single interventional radiologist with 20 years of experience. Under the guidance of intra-procedural imaging, selective or super-selective embolization was performed with a solution containing 50 mg of doxorubicin and 10 mg of mitomycin C in a 1:1 mixture with ethiodized oil (Lipiodol; Laboratoire Guerbet, Aulnay-sous-Bois, France). Microspheres with a diameter of 300–500 µm were used to embolize more proximal vessels (Embosphere; Merit Medical Systems, South Jordan, Utah). In patients receiving TACE with drug-eluting embolic agents, LC Beads (BTG, Surrey, UK) with a diameter of 100–300 µm were used with 100 mg of doxorubicin hydrochloride (25 mg/mL). Doxorubicin-eluting embolic agents were mixed with an equal volume of nonionic contrast material (Oxilan, 300 mg of iodine/mL; Guerbet, Bloomington, Indiana, USA) before intra-arterial administration. The TACE endpoint was significant flow reduction while avoiding stasis; contrast column in the feeding vessel was aimed to be cleared with 5 heart beats. Selectivity of embolization was achieved in all patients. No adverse events were recorded in the retrospective dataset. The cohort had a 30-day procedure-related mortality of 0% and no technical peri-procedural complications were observed (arterial dissections, bleed, or infections). No grade 3/4 toxicities were observed in this patient cohort. Nausea and right upper quadrant pain was the most frequently encountered toxicity (Grade 1/2 with 40% and 37%, respectively).

MR Image Analysis using qEASL

The workflow of data pre-processing is provided in Figure 1. Patients underwent pre-treatment (within 1 week prior to TACE) and follow-up (1–3 months after TACE treatment) contrast-enhanced multiphasic T1-weighted MR imaging with a 1.5-T MR unit (Magnetom Avanto; Siemens, Erlangen, Germany) using a phased-array torso coil. The protocol included breath-hold unenhanced and contrast-enhanced (0.1 mmol/kg intravenous gadopentetate dimeglumine (Magnevist; Bayer, Wayne, NJ)) T1-weighted 3D fat-suppressed spoiled gradient-echo imaging in the hepatic arterial phase (20 seconds after contrast administration), portal venous phase (70 seconds after contrast administration), and delayed phase (3 minutes after contrast administration).

Only the pre-contrast and 20 second arterial phase images were used for image analysis. The arterial phase of the MRI was selected for semi-automated 3D tumor segmentation using qEASL software (IntelliSpace Portal version 8; Philips Healthcare, Haifa, Israel). The qEASL software has been validated in multiple studies as an accurate predictor of survival in liver malignancies with high inter-reader reproducibility (7, 10–13). The qEASL analysis was performed by a fourth-year medical student and confirmed by a second-year radiology resident, supervised by a board-certified radiologist who did not perform the TACE procedures. Each qEASL measurement included three parenchymal regions of interest (ROI) to generate an average for parenchymal intensity. These averages were confirmed by an independent implementation of a qEASL algorithm that heuristically selected the ROI. Readers did not perform TACE treatments and were blinded to survival outcomes. qEASL values are the enhancing tumor volume expressed as a percent of the total tumor volume. A positive change in qEASL after treatment reflects a decrease in viable tumor volume or decrease in tumor enhancement (13). As defined in previous HCC qEASL studies, patients with changes in qEASL $\geq 65\%$ (or reduction of enhancing tumor volume by $\geq 65\%$) after therapy as compared to baseline imaging were labeled as treatment responders. Changes in qEASL $< 65\%$ were defined as treatment non-responders (7). The labels of treatment responder or non-responder were attributed to each patient in the cohort and were used as the outcome-labeled dataset applied to train ML models to predict treatment response from pre-treatment patient data.

Feature Selection

First, the input clinical and imaging variables used to teach the two learning models (referred to as features) were identified. Based on clinical and imaging variables known to contribute to TACE treatment response, a collection of 25 clinical, laboratory, demographic, and imaging features was assembled in the retrospective patient cohort. These 25 features were sorted into binary categories (e.g. pre-TACE number of tumors > 2 or ≤ 2) to standardize the scale of each variable. A detailed explanation of how the features were sorted is provided in the Appendix. A variance threshold was applied to remove features that were all negative or all positive more than 80% of the time. A univariate chi-squared test was performed with each feature and the outcome variable of treatment response or non-response. A p -value threshold of 0.55 was selected after variance thresholding with the aim of reducing the number of features to five; any variable with a p -value > 0.55 was excluded. Features that satisfied both criteria included one feature diagnosed through clinical history

and laboratory results (presence of cirrhosis), two features derived from baseline imaging (pre-TACE tumor signal intensity > 27.0, pre-TACE number of tumors >2), and two therapeutic features (whether or not the patient received conventional TACE, and whether or not the patient was previously or simultaneously treated with sorafenib). Pre-TACE tumor signal intensity is a measure of tumor enhancement and was obtained by taking the average voxel brightness value (intensity) within a tumor volume.

These five features were used to train logistic regression (LR) and random forest (RF) models to predict patients as treatment responders or non-responders (Figure 1). For both LR and RF models, 30 different combinations of the five selected features were tested to identify the most accurate predictive model. For additional information on the parameters used for each model, refer to the Appendix.

Machine Learning Model Evaluation

Once trained, the models' accuracy at predicting treatment response compared to the qEASL categorization of treatment response was tested using leave-one-out cross-validation (LOOCV). The model was trained using 35/36 patients, and tested on the remaining patient. This training-testing process was repeated 36 times, with a different patient left-out of training each time. LOOCV was used to validate both LR and RF models.

Results

Labeled outcomes of patient cohort

qEASL applied to post-TACE MR images of 36 patients classified eight patients as responders and 28 patients as non-responders.

Feature selection

11/25 features passed the variance criterion and 14/25 features satisfied the chi-squared p -value criterion. Altogether, 5/25 features satisfied both criteria and were used for model training: one clinical feature (presence of cirrhosis, $p=0.3$), two imaging features (pre-TACE tumor signal intensity >27.0, $p=0.2$; number of tumors >2, $p=0.5$), and two therapeutic features (treatment with sorafenib, $p=0.5$; treatment with ethiodized oil, $p=0.5$) (Table 2).

Model evaluation

The models are initially biased towards responders (80% responders, 20% non-responders). The LR classifier trained on all five features achieved an overall accuracy of 72%, a sensitivity of 50.0%, and a specificity of 78.6% (4/8 responders, 22/28 non-responders). This corresponds to a positive predictive value (PPV) of 40.0% and a negative predictive value (NPV) of 84.6%.

The RF classifier trained on all five features achieved an overall accuracy of 66%, a sensitivity of 62.5%, and a specificity of 67.9%. This corresponds to a PPV of 35.7% and a NPV of 86.4% (5/8 responders, 19/28 non-responders).

The highest overall accuracy was achieved when the models were trained using the two features of pre-TACE tumor signal intensity > 27.0 and presence of cirrhosis; when trained with these two features, the RF and LR models achieved the same overall accuracy of 78% (sensitivity 62.5%, specificity 82.1%, PPV 50.0%, NPV 88.5% for both). All results were obtained through LOOCV (Figure 2).

Discussion

The main result of this study is the successful application of AI and ML techniques to predict TACE treatment outcomes pre-procedurally using clinical and imaging features. The essence of ML is training a computer to recognize patterns in data to predict a specific outcome. For this study, the input data consisted of clinical and imaging information and the predicted outcome was response to TACE treatment. ML maps the input features to the outcome under a specific model. LR models linearly weigh each feature to predict the outcome, while RF models average a set of decision trees to account for nonlinear relationships between the features.

LR and RF models were trained using baseline clinical and imaging data, treatment characteristics (how the patient was treated), and treatment outcomes derived from a retrospective cohort of patients with HCC treated with TACE. To apply the model to a new HCC patient being considered for TACE, this same baseline clinical and imaging information would be fed to the model along with the potential treatment being considered. The model would then predict the treatment outcome. Classifications of patients as treatment responders or non-responders can be used as an indication to choose or not choose a potential treatment in clinical practice (Figure 3).

The best overall accuracy was achieved with the two-feature model trained using one image-derived feature (pre-TACE tumor signal intensity > 27.0) and one clinical feature (presence of cirrhosis). One hypothesis for the synergistic role these features play in determining TACE outcomes relates perfusion to cirrhosis. The qEASL criterion measures the change in enhancing tumor volume. A tumor with higher mean signal intensity before TACE will likely undergo a greater change in enhancement following TACE. Hepatic arterial flow in cirrhotic patients is known to undergo compensation when portal venous flow decreases (14). The increased tumor intensity in the presence of cirrhosis may be due to hepatic arterial compensation. Embolization of compensated arteries would therefore theoretically result in a higher net decrease in tumor blood flow compared to embolization of non-compensated arteries. While the two-feature model outperformed the five-feature model, it is not possible to conclude that the dropped features were insignificant to TACE outcomes; merely that their contribution provided less insight into the study-specific dataset. There may exist synergy between these features when applied to a larger dataset.

Several ML models solve classification problems in a supervised learning context, including RF, LR, support vector machines, and Bayesian networks (15–17). Generally, the process of training the learning model is more important than the type of ML model selected (18). For this study, a LR model was selected because it is simple to implement. However, LR assumes a linear relationship between variable and outcome (17). This is often not the case,

especially when integrating diverse variables from clinical and imaging sources. Relative to LR, RF models are better able to handle nonlinear relationships between variables and outcomes (17). Therefore, a RF model was also selected to be trained as a second ML model.

The ideal ML model includes the most relevant features to the outcome that are also generalizable beyond the training set. However, it is not possible to identify these features with certainty. Certain features may appear important, but only because of the small dataset size. Given that the dataset was only 36 patients, inclusion of 6 binary features would allow 64 possible feature vectors, substantially increasing the likelihood of overfitting. For this reason, the number of features included in the model was reduced from 25 to 5. The variance threshold was applied first, and the p-value threshold was set to arrive at exactly five features.

The primary limitation of this study is the very small patient cohort with which the models were trained. In order to increase the accuracy of treatment response prediction, the models should be trained on a larger patient cohort curated to represent both typical and atypical HCC tumors. It was also necessary to combine multiple treatment modalities in the training set (e.g. conventional TACE and TACE with drug-eluting embolic agents) to increase the size of the training set. Future studies are warranted that include patients from multiple centers to evaluate treatment modality as a feature. Emerging modalities could also be included as input to the model. The strong predictive value of cirrhosis and tumor enhancement alone demonstrate the importance of selecting the correct features to include in a ML model. However, performing an exhaustive search of features to maximize performance on one dataset runs the danger of overfitting. The two-feature model worked best for this dataset, but may not generalize beyond this patient cohort. An additional limitation of this study is the imbalance in the training set; the number of treatment non-responders outweighed the number of treatment responders. The trends in figure 2 are reflective of this imbalance; as overall accuracy grows with each feature added, the accuracy of classifying treatment non-responders increases while treatment responder classification decreases. Additional training of these models with a patient cohort with a greater frequency of treatment responders would most effectively mitigate this bias. Another limitation of this study is a smaller than average tumor size. Tumor size greater than 3 cm was included as a feature, however it was not stratified beyond this binary cutoff. A direction for future work would be to include additional size-cutoffs as features to see if their contributions provide additional insight into TACE outcomes.

Given the lack of evidence-driven guidelines supporting the use of one intra-arterial therapy over others, ML based models predicting treatment response could assist physicians making decisions about loco-regional treatment selection in patients with liver cancer. Treatment-selection would require additional studies that include multiple treatment modalities as input features with the goal of maximizing the responder class probability. The widespread adoption of the electronic health records has aggregated many types of disparate medical information and vastly improved their accessibility. The field of interventional oncology should exploit the volume and diversity of data from clinical records to design predictive clinical tools. Enabled by machine learning methods, reliable insight into treatment

outcomes will empower physicians to make more informed recommendations for the management of hepatic malignancies.

References

1. Benson AB 3rd, Abrams TA, Ben-Josef E, Bloomston PM, Botha JF, Clary BM, et al. NCCN clinical practice guidelines in oncology: hepatobiliary cancers. *J Natl Compr Canc Netw*. 2009; 7(4):350–91. [PubMed: 19406039]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016; 66(1):7–30. [PubMed: 26742998]
3. Gillmore R, Stuart S, Kirkwood A, Hameeduddin A, Woodward N, Burroughs AK, et al. EASL and mRECIST responses are independent prognostic factors for survival in hepatocellular cancer patients treated with transarterial embolization. *J Hepatol*. 2011; 55(6):1309–16. [PubMed: 21703196]
4. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45(2):228–47. [PubMed: 19097774]
5. Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. *J Hepatol*. 2001; 35(3):421–30. [PubMed: 11592607]
6. European Association For The Study Of The L, European Organisation For R, Treatment Of C. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2012; 56(4):908–43. [PubMed: 22424438]
7. Tacher V, Lin M, Duran R, Yarmohammadi H, Lee H, Chapiro J, et al. Comparison of Existing Response Criteria in Patients with Hepatocellular Carcinoma Treated with Transarterial Chemoembolization Using a 3D Quantitative Approach. *Radiology*. 2016; 278(1):275–84. [PubMed: 26131913]
8. Bishop, CM. Pattern recognition and machine learning. New York: Springer; 2006. p. xxp. 738
9. Alpaydin, E. Introduction to machine learning. 3. Cambridge, Massachusetts: The MIT Press; 2014. p. xxiip. 613
10. Sahu S, Scherthaner R, Ardon R, Chapiro J, Zhao Y, Sohn JH, et al. Imaging Biomarkers of Tumor Response in Neuroendocrine Liver Metastases Treated with Transarterial Chemoembolization: Can Enhancing Tumor Burden of the Whole Liver Help Predict Patient Survival? *Radiology*. 2017; 283(3):883–94. [PubMed: 27831830]
11. Duran R, Chapiro J, Frangakis C, Lin M, Schlachter TR, Scherthaner RE, et al. Uveal Melanoma Metastatic to the Liver: The Role of Quantitative Volumetric Contrast-Enhanced MR Imaging in the Assessment of Early Tumor Response after Transarterial Chemoembolization. *Transl Oncol*. 2014; 7(4):447–55. [PubMed: 24953419]
12. Chockalingam A, Duran R, Sohn JH, Scherthaner R, Chapiro J, Lee H, et al. Radiologic-pathologic analysis of quantitative 3D tumour enhancement on contrast-enhanced MR imaging: a study of ROI placement. *Eur Radiol*. 2016; 26(1):103–13. [PubMed: 25994198]
13. Lin M, Pellerin O, Bhagat N, Rao PP, Loffroy R, Ardon R, et al. Quantitative and volumetric European Association for the Study of the Liver and Response Evaluation Criteria in Solid Tumors measurements: feasibility of a semiautomated software method to assess tumor response after transcatheter arterial chemoembolization. *J Vasc Interv Radiol*. 2012; 23(12):1629–37. [PubMed: 23177109]
14. Iranpour P, Lall C, Houshyar R, Helmy M, Yang A, Choi JI, et al. Altered Doppler flow patterns in cirrhosis patients: an overview. *Ultrasonography*. 2016; 35(1):3–12. [PubMed: 26169079]
15. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One*. 2017; 12(5):e0177726. [PubMed: 28542342]
16. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc Inform Res*. 2010; 16(4):253–9. [PubMed: 21818444]

17. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal.* 2012; 16(5):933–51. [PubMed: 22465077]
18. Hastie, T., Tibshirani, R., Friedman, JH. *The elements of statistical learning : data mining, inference, and prediction.* 2. New York, NY: Springer; 2009. p. xxiip. 745

Appendix

Thresholds for clinical variables (e.g. serum albumin >3.5 g/dL) were selected based on Barcelona Clinic Liver Cancer cutoffs of clinical parameters associated with overall survival. Other imaging variables were placed into binary categories (e.g. pre-TACE number of tumors >2 or <2) based on thresholds identified by bucketing data into two histogram bins and identifying the cutoff value between bins.

For computational efficiency, the following pre-TACE imaging traits were transformed into binary features: enhancing tumor volume, liver volume, mean liver signal intensity, tumor signal intensity, number of tumors, standard deviation of liver signal intensity, standard deviation of tumor signal intensity, tumor volume, and tumor diameter. After bucketing each aforementioned variable into 3 bins, values in the uppermost bin above a defined threshold were assigned a value of one, otherwise zero.

The logistic regression and random forest classifiers were trained using standard formulations with specific values provided in Table 3 of the supplemental materials.

Figure 1. Image Processing Workflow

Workflow describing MR image pre-processing and feature selection steps. After segmentation of tumor boundaries, qEASL enhancing tumor volumes were obtained before and after treatment and used to determine patient classification as treatment response or non-response. The qEASL treatment outcomes were combined with clinical and imaging features passing selection criteria to train logistic regression and random forest models to predict treatment outcomes. The accuracy of the models compared to qEASL assessment of treatment response was evaluated using leave-one-out cross-validation.

Abbreviation: quantitative European Association for the Study of the Liver, qEASL; trans-arterial chemoembolization, TACE; magnetic resonance, MR

Figure 2. Performance of Machine Learning Algorithms

Logistic regression (A and B) and random forest (C and D) classifier accuracies as features are added. Features are added according the following order: 1: ethiodized oil, 2: Sorafenib, 3: Cirrhosis, 4: Pre-TACE tumor signal intensity > 27.0, 5: Number of tumors > 2).

Abbreviation: trans-arterial chemoembolization, TACE

Figure 3. Clinical Application of Machine Learning Models for Response Prediction

Figure illustrating the clinical application of machine learning models trained to predict response to loco-regional treatment and thereby inform treatment selection. Learning model classification of patient treatment response using pre-treatment data in combination with a planned treatment protocol can be used as an indication to treat or not to treat with TACE. Abbreviation: trans-arterial chemoembolization, TACE

Table 1

Training Cohort: Patient, Disease, and Treatment Characteristics

Parameter	N (%)	Parameter	N (%)
Number of Patients	36	Sorafenib	8 (22.2)
Treatment Modality		Transplant recipient	5 (13.9)
Conventional TACE	30 (83)	Distant Metastasis	2 (5.6)
DEE-TACE	6 (17)	Lymph Node Metastasis	3 (8.3)
Gender		Hepatitis B	3 (8.3)
Male	31 (86.1)	Hepatitis C	28 (77.8)
Female	5 (13.9)	Cirrhosis	5 (13.9)
Race		Ascites	5 (13.9)
White	23 (63.9)	Encephalopathy	2 (5.6)
African American	8 (22.2)	BCLC stage	
Asian/Pacific Islander	2 (5.6)	0	1 (2.8)
Other	3 (8.3)	A	12 (33.3)
ECOG Performance Status		B	10 (27.8)
0	24 (66.7)	C	13 (36.1)
1	10 (27.8)		
2	2 (5.5)	Parameter	Mean (SD)
Child Pugh Class		Albumin	3.98 (0.5)
A	31 (86.1)	Bilirubin	0.9 (0.5)
B	4 (11.1)	INR	1.1 (0.1)
C	1 (2.8)	Tumor Size (cm)	5.2 (3.0)
		Number of Tumors	2.6 (1.1)
		Age	62.0 (8.9)

Clinical, disease, and treatment characteristics of the patient cohort used to train the machine learning model. Liver transplantation occurred strictly after TACE in all patients who were transplanted.

Abbreviations: trans-arterial chemoembolization, TACE; drug-eluting embolic agents, DEE; standard deviation, SD; Barcelona Clinic Liver Cancer, BCLC

Table 2

Feature Selection Criteria.

Binary feature	Class variance > 20%	Univariate p-value
Clinical / Laboratory		
Albumin > 3.5 g/dL	Yes	0.68
Alcoholic liver disease	Yes	0.96
Ascites present		0.31
Bilirubin > 1.5 mg/dL		0.83
Encephalopathy present		0.47
Hepatitis B		0.59
Hepatitis C	Yes	0.58
Transplant recipient		0.97
Distant (extra-hepatic) metastases present		0.30
Lymph node metastases present		0.59
<i>Cirrhosis present</i>	Yes	0.34
Treatment		
<i>Treated with Lipiodol</i>	Yes	0.51
<i>Sorafenib treatment</i>	Yes	0.51
Demographic		
Male		0.28
White ethnicity	Yes	0.74
Imaging		
Portal vein invasion		0.47
Pre-TACE enhancing tumor volume > 598 cm ³		0.38
Pre-TACE liver volume > 1,990 cm ³	Yes	0.57
Pre-TACE mean liver signal intensity > 15.4		0.97
<i>Pre-TACE tumor signal intensity > 27.0</i>	Yes	0.19
<i>Pre-TACE number of tumors > 2</i>	Yes	0.51
Pre-TACE standard deviation of liver signal intensity > 9.4	Yes	0.83
Pre-TACE standard deviation of tumor signal intensity > 10.6		0.90
Pre-TACE tumor volume > 1,070 cm ³		0.38
Tumor diameter > 3 cm	Yes	0.70

Total features considered and final features selected to train machine learning model after applying selection criteria: a) > 20% variance and b) Univariate *p*-value < 0.55 when associated with treatment response. Five features (italicized) satisfied both criteria.

Table 3

Model Parameters.

Model	Parameter	Value
Logistic regression	Norm	L2
	Formulation	Primal
	Inverse of regularization	10 ⁻¹⁵
	Bias constant added	Yes
	Bias constant scaling	1
	Solver	liblinear
	Tolerance	10 ⁻⁴
Random forest	Tree count	100
	Quality measure	Gini impurity
	Samples per split	2
	Min samples per leaf	1
	Min impurity split	10 ⁻⁷
Parameters in Common	NF	5
	N	36 (28 R-, 8 R+)
	R- probability threshold	0.8
	Validation	LOOCV

Parameters used to train machine learning models. The threshold for non-responder classification was selected as 0.8; the probability of the patient being a non-responder must be determined to be greater than 0.8 by the model, otherwise a responder label is applied. Model implementations were enabled by the freely available scikit-learn Python library.

Abbreviations: number of features, NF; number of patients, N; treatment non-responder, R-; treatment responder, R+; leave-one-out cross-validation, LOOCV