Article

Chemoembolization for Hepatocellular Carcinoma including Contrast Agent-Enhanced CT: Response Assessment Model on Radiomics and Artificial Intelligence

Sungjin Yoon 1,*, Youngjae Kim 2,†, Juhyun Kim 3, Yunsoo Kim 3, Ohsang Kwon 3, Seungkak Shin 3, Jisoo Jeon 2 and Seungjoon Choi 1,*

1 Department of Radiology, Gachon University College of Medicine, 21 Namdong-daero 774beon-gil, Namdong-gu, Incheon 21565, Republic of Korea
2 Department of Biomedical Engineering, Gachon University College of Medicine, 21 Namdong-daero 774beon-gil, Namdong-gu, Incheon 21565, Republic of Korea
3 Division of Gastroenterology, Department of Internal Medicine, Gachon University College of Medicine, 21 Namdong-daero 774beon-gil, Namdong-gu, Incheon 13120, Republic of Korea
† These authors contributed equally to this work.

Abstract: Purpose: The aim of this study was to assess the efficacy of an artificial intelligence (AI) algorithm that uses radiomics data to assess recurrence and predict survival in hepatocellular carcinoma (HCC) treated with transarterial chemoembolization (TACE). Methods: A total of 57 patients with treatment-naïve HCC or recurrent HCC who were eligible for TACE were prospectively enrolled in this study as test data. A total of 100 patients with treatment-naïve HCC or recurrent HCC who were eligible for TACE were retrospectively acquired for training data. Radiomic features were extracted from contrast-enhanced, liver computed tomography (CT) scans obtained before and after TACE. An AI algorithm was trained using the retrospective data and validated using the prospective test data to assess treatment outcomes. Results: This study evaluated 107 radiomic features and 5 clinical characteristics as potential predictors of progression-free survival and overall survival. The C-index was 0.582 as the graph of the cumulative hazard function, predicted by the variable configuration by using 112 radiomics features. The time-dependent AUROC was 0.6 ± 0.06 (mean ± SD). Among the selected radiomics features and clinical characteristics, baseline_glszm_SizeZoneNonUniformity, baseline_glszm_ZoneVariance and tumor size had excellent performance as predictors of HCC response to TACE with AUROC of 0.853, 0.814 and 0.827, respectively. Conclusions: A radiomics-based AI model is capable of evaluating treatment outcomes for HCC treated with TACE.

Keywords: hepatocellular carcinoma; chemoembolization; radiomics; artificial intelligence; contrast-enhanced CT; treatment response; overall survival

1. Introduction

Hepatocellular carcinoma (HCC) is the fifth most common malignant tumor and is responsible for a half million deaths annually worldwide [1]. Curative treatment options such as surgical resection, local ablation and liver transplantation have been the most effective treatments for HCC. However, less than 20% of HCC patients are treated surgically, mainly because of associated cirrhosis and advanced stage of cancer at diagnosis [2,3]. Based on the survival benefits with increasing evidence [4,5], transarterial chemoembolization (TACE) is the widely-used, standard treatment modality for intermediate-stage HCC.

Multi-detector computed tomography (CT) is the most widely used imaging tool for assessing therapeutic response after TACE [6]. However, previous studies have reported
that the detection of viable nodules, using lipiodol, is usually ambiguous since hepatic lesions are frequently hyper attenuated in arterial phase images, such as arterio-portal shunts [7,8]. Manual assessment, which is observer-dependent, is based on the tumor diameter measured in only one axial plane rather than its full three-dimensional nature [9]. Certainly, these methods are not always applicable considering the various conditions that HCC tumors can present with after TACE treatment, including those with irregular shapes, ill-defined margins, or heterogeneous necrosis [10]. Previous studies have shown that overall response rates range from 15% to 85% after TACE and cumulative local tumor progression rates at 1 and 5 years are 33% and 73%, respectively [11,12]. Therefore, it is important to assess tumor response to TACE treatment, which can help guide subsequent treatment strategies.

Recent emerging technologies in quantitative computational image analysis offer promising opportunities. Bioinformatic analysis transforms images into mineable data, enabling the characterization of lesions beyond visual recognition. Radiomics, a high-dimensional quantitative analysis approach, can compute a set of features that uniquely characterize a tumor [13,14]. Radiomics has been successfully applied to predict outcomes through tumor imaging for various cancers [15,16]. However, using radiomics to predict survival in HCC has not been explored.

Recently, deep convolutional neural networks have been gaining recognition in imaging research [17]. A neural network is a modality used for artificial intelligence (AI), and convolutional layers are an effective tool for imaging pattern recognition. Previous studies demonstrated that deep learning with convolutional neural networks has achieved good performance in imaging pattern recognition [18,19]. Therefore, further studies are needed to support the robustness of radiomics approaches for predicting recurrence in HCC patients after HCC treatment.

The aim of this study was to evaluate an AI-based radiomics model applying CT studies after TACE that could predict HCC recurrence and be a prognostic biomarker for the survival of HCC patients.

2. Materials and Methods
2.1. Patients

This study was approved by the institutional review board of the Gil Medical Center (GAIIRB2019-038) and was performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. Between May 2019 and August 2021, we enrolled 57 patients with treatment-naïve intrahepatic HCC, or recurrent HCC (no marginal recurrence), and were eligible for TACE as their first-line therapy. Exclusion criteria were as follows; (1) Tumor thrombus in main portal vein; (2) Infiltrative HCC; (3) Extra-hepatic tumor spread.

2.2. Methods Including Dose of Contrast Agent and CT Protocol

All patients received baseline CT scans within one month prior to TACE and follow-up CT scans 4-12 weeks after TACE. The CT liver protocol was used for all patients by including non-enhanced, arterial, portal venous, and wash-out phases. After enhancement in the descending aorta had reached 100 Hounsfield units, images of the arterial and portal venous phases were obtained with 18 and 50 s delays, respectively. The equilibrium phase was obtained with a fixed delay of 180 s after initiating the contrast injection. Contrast material, containing iobitridol (Xenetics; Guerbet, Aulnay sous Bois, France), was injected at 1 mL/kg of body weight (to a maximum of 150 mL) via 18-gage, peripheral venous access (generally an antecubital vein), at a flow rate of 4 mL/s with a power injector. No side effects from the CT contrast media were reported by any patients.

2.3. TACE
TACE treatment was administered with a mixture of 5 mL iodized oil contrast medium, lipiodol (Guerbet, Aulnay sous Bois, France), and 30–50 mg of adriamycin. We selectively embolized the feeding artery using absorbable gelatin sponge particles until flow stasis. TACE was performed repeatedly in an “on-demand” fashion during follow-up examinations 4–12 weeks after the initial TACE treatment.

2.4. Assessment of Actual Treatment Responses
The modified RECIST (mRECIST) guideline was used for response evaluations [20]. According to mRECIST, the longest diameter of the viable portion of the tumor with viability is defined as tumor tissue with enhancement in the arterial phase of contrast-enhanced CT. Complete response (CR) is defined as the disappearance of any intratumoral arterial enhancement in all target lesions; partial response (PR) is considered when the sum of the longest viable tumor diameters of target lesions decrease at least 30% from their baseline sum; progressive disease (PD) occurred when the sum of the longest viable tumor diameters increase by at least 20% to the baseline sum of diameters of target lesions since the beginning of treatment; and stable disease (SD) is any case that shows neither a sufficient decrease or increase in the viable tumor diameter sum to qualify for PR or PD, respectively.

2.5. Radiomics and Artificial Intelligence
Radiomics approaches include texture analysis as a subset and compute hundreds of features to describe tumor characteristics. The tumor phenotype was quantified and images were converted into mineable data with high-dimensional features. We also extracted image features related to tumor intensity (histogram), shape, texture and wavelet (high or low-frequency feature).

To extract image features from CT data, a binary mask is created by specifying the lesion area of the image as an ROI. Using the pyradomics library, we input the original image data and the generated binary mask to extract radiomics features for the lesion area of the original image. Radiomics feature extraction is the process of quantifying the imaging features of a specified region in image data through a mask into quantitative variables such as shape features, first-order features, and second-order features. Shape features are features related to the size of the ROI specified by the mask, which are variables that quantify the size, shape, surface, and orientation of the image. The extracted shape features include the number of pixels contained within the ROI (volume), the number of surface pixels to measure the surface area (surface area), how close to spherical the ROI is (sphericity), and how elongated the shape of the ROI is (elongation). First-order features are features related to the distribution of pixel values within the ROI and are computed using a histogram of the frequencies of pixel values. These features quantify information such as the brightness, contrast, and sharpness of the image. Second-order features are features related to the spatial correlation of pixel values in the ROI of a medical image, and are calculated by constructing a matrix that represents the spatial relationship of each pair of pixels in the region of interest and performing a matrix product with the image data. It is expressed as Gray-Level Dependence Matrix (GLDM), Gray-Level Run Length Matrix (GLRLM). The features computed in this way can quantify and express information such as texture or pattern in the image.

The extracted radiomics variables consist of 107 features, including shape features (14 features), first-order features (18 features), and second-order features including GLCM and GLDM (75 features). The authors added 5 clinical variables to this, for a total of 112 features used in the study.

We conducted the training phase using retrospective CT data (100 patients, from January 2015 to February 2016) and performed the prediction phase using prospective data (57 patients) (Table 1). Radiomic features were extracted from CT images of the arterial
The location of the tumor was marked by the radiologist, and the tumor was segmented by the medical imaging engineer. Finally, the segmentation result was confirmed by the radiologist.

**Table 1. Baseline Patient Characteristics of Prediction and Training Set.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prediction Set (n = 49)</th>
<th>Training set (n = 98)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (range)</td>
<td>65.9 (10.4)</td>
<td>63.5 (9.8)</td>
</tr>
<tr>
<td>Sex, male/female</td>
<td>36/13</td>
<td>77/21</td>
</tr>
<tr>
<td>Etiology of liver cirrhosis, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td>30 (61)</td>
<td>63 (63)</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td>7 (14)</td>
<td>12 (12)</td>
</tr>
<tr>
<td>Alcoholic</td>
<td>9 (19)</td>
<td>16 (16)</td>
</tr>
<tr>
<td>Hepatitis B virus + Alcoholic</td>
<td>2 (4)</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Hepatitis C virus + Alcoholic</td>
<td>1 (2)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>HCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>46 (94)</td>
<td>81 (82)</td>
</tr>
<tr>
<td>2</td>
<td>3 (6)</td>
<td>17 (18)</td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 cm</td>
<td>17 (34)</td>
<td>34 (35)</td>
</tr>
<tr>
<td>2 ≤ x &lt; 5 cm</td>
<td>27 (55)</td>
<td>46 (47)</td>
</tr>
<tr>
<td>≥5 cm</td>
<td>5 (11)</td>
<td>18 (18)</td>
</tr>
<tr>
<td>Child–Pugh score, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>41 (84)</td>
<td>84 (86)</td>
</tr>
<tr>
<td>B</td>
<td>8 (16)</td>
<td>14 (14)</td>
</tr>
<tr>
<td>Survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progression-free survival (months)</td>
<td>16.2 (12.2)</td>
<td>16.1 (17.9)</td>
</tr>
<tr>
<td>Overall survival (months)</td>
<td>31.4 (13.3)</td>
<td>50.7 (26.7)</td>
</tr>
</tbody>
</table>

We used radiomics to extract variables from the baseline and follow-up CT data (107 variables each). The variables extracted from the baseline and follow-up data were combined with clinical variables to create one dataset (112 variables in total, 98 cases). The model was trained using the scikit-survival framework, which was analyzed using the Random Forest Survival method. A total of 98 cases was used as the training set, and a prediction group of 48 cases was used as the test set. We used TensorFlow 2.6 framework.

2.6. Statistical Analysis

We used a Concordance Index (C-index) score to measure the performance of the model. The C-index is a performance metric used in survival analyses, which measures the degree of agreement between the relative order of events predicted by a model and the actual order in which they occur. The higher the C-index score, the better the model’s predictions [21,22]. We also measured the time-dependence area under the receiver operating characteristics (AUROC) by comparing the prediction of survival over time with the actual event occurrence. Analyses were performed by an independent investigator using scikit-survival 0.22.1. The AUROC and optimal thresholds were obtained by the multipleROC package in R. All reported p values are two-sided and considered statistically significant at <0.05. Statistical analyses were performed using R software/environment (R version 2.9.1).
3. Results

3.1. Patient Demographics

Patient demographic information is presented in Table 1. Of the 57 patients in the prediction group, two were lost to follow-up and one patient expired before undergoing TACE. Five additional patients were excluded from the analysis because tumor segmentation was not possible. As a result, 49 patients were included in the prediction phase of the analysis. Of the 100 patients in the training group, two were excluded from the analysis because tumor segmentation was not possible. As a result, 98 patients were included in the training phase of the analysis (Figure 1).

The mean age of the prediction and training groups were 65.9 ± 10.4 and 63.5 ± 9.8 (mean ± standard deviation [SD]) years, respectively. Etiology of liver cirrhosis in the prediction and training groups included hepatitis B virus (61% and 63%), hepatitis C virus (14% and 12%), and alcoholic (19% and 16%). Some viral carriers in the prediction and training groups were also chronic alcoholics; hepatitis B virus (4% and 6%); and hepatitis C virus (2% and 3%), respectively. In the prediction and training groups, Child–Pugh Class A was 84% and 86% and Child–Pugh Class B was 16% and 14%, respectively. The number of HCCs in the prediction and validation group included single (94% and 82%) and two (6% and 18%) HCCs, respectively. The HCC tumor sizes in the prediction and validation groups included <2 cm (34% and 35%), 2 ≤ x < 5 cm (55% and 47%) and ≥5 cm (11% and 18%), respectively. Progression-free survival (PFS) of the prediction and validation groups were 16.2 and 16.1 months and overall survival (OS) was 31.4 and 50.7 months, respectively (Table 1).

3.2. Survival Models

We compared 107 radiomics features between the baseline and follow-up data and analyzed 5 clinical characteristics (gender, age, tumor number, tumor size, and Child–Pugh score) as potential predictors of PFS and OS by the scikit-survival framework model, using the Random Forest Survival method. We calculated the C-index score to measure the performance of the model. The C-index was 0.582 (p = 0.020) as the graph of the cumulative hazard function, predicted by the variable configuration by using 112 features. (Figure 2). The time-dependent AUROC was 0.6 ± 0.06 (mean ± SD, p = 0.014) when comparing the prediction of survival over time with the actual event occurrence (Figure 3). Feature selection was performed based on the C-score of the model, and finally, four variables were extracted. The selected four variables were follow-up_glszm_LargeAreaHighGrayLevelEmphasis (weight 0.088), baseline_glszm_SizeZoneNonUniformity (weight 0.007), baseline_glszm_ZoneVariance (weight 0.006), baseline_firstorder_Kurtosis (weight 0.005) (Figure 4). We further determined whether the selected factors were good predictors of tumor response in HCC treated with TACE. According to mRECIST criteria, 33 patients (68%), achieved a CR and 15 (32%) had a PR, SD and PD. We then evaluated whether radiomics features and clinical factors perform optimally in the diagnosis of viable tumors. The AUROC of Baseline_glszm_SizeZoneNonUniformity, baseline_glszm_ZoneVariance and tumor size was 0.853 (p < 0.001), 0.814 (p < 0.001) and 0.827
(p < 0.001), respectively, and showed excellent performance as predictors of HCC response treated with TACE, while follow-up_glszm_LargeAreaHighGrayLevelEmphasis and kurtosis performed relatively well (AUROC, 0.812 (p < 0.001) and 0.681 (p < 0.001), respectively) (Figure 5).

**Figure 2.** Graph of the cumulative hazard function predicted by the variable configuration for each test case. The colored lines are the 107 radiomics features and 5 clinical characteristics analyzed as potential predictors of overall survival, represented by the scikit-survival framework model. C-index scores were calculated to measure the performance of the model.
Figure 3. Graph of the time-dependent AUROC predicted by the variable configuration for each test case.

<table>
<thead>
<tr>
<th>feature</th>
<th>weight</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>FU_glszm_LargeAreaHighGrayLevelEmphasis</td>
<td>0.088724</td>
<td>0.021511</td>
</tr>
<tr>
<td>base_glszm_SizeZoneNonUniformity</td>
<td>0.007235</td>
<td>0.023347</td>
</tr>
<tr>
<td>base_glszm_ZoneVariance</td>
<td>0.006428</td>
<td>0.029546</td>
</tr>
<tr>
<td>base_firstorder_Kurtosis</td>
<td>0.005649</td>
<td>0.018157</td>
</tr>
</tbody>
</table>

Figure 4. Result of feature selection.
Figure 5. Summary of the diagnostic performance of the radiomics features and clinical factors for the prognosis of HCC response to TACE treatment.

4. Discussion

This study evaluated the utility of radiomics derived from contrast-enhanced CT scans and AI models in predicting the response and prognosis to TACE treatment in patients with HCC. We conducted a comparative analysis between tumor responses to TACE predicted by the radiomics model and actual responses measured using mRECIST guidelines. Additionally, comparisons of PFS and OS were made between the mRECIST guidelines and radiomics model prediction. Variables were extracted using radiomics from both baseline and follow-up contrast-enhanced CT scans, and ultimately, four variables were selected through feature selection calibration. Ultimately, the C-index of the radiomics prediction model for survival analysis yielded a value of 0.582.

TACE is the established standard treatment for intermediate-stage HCC, and CT serves as the primary imaging modality for evaluating treatment response. Lipiodol is commonly utilized in TACE, exhibiting high attenuation on CT scans. The region exhibiting high attenuation due to lipiodol uptake demonstrates a robust correlation with tumor necrosis [23]. However, the high attenuation displayed by lipiodol on CT scans creates challenges in evaluating small enhancing viable lesions, as it gives rise to beam-hardening artifacts in the surrounding area [24]. The sensitivity of CT in detecting residual or recurrent disease after TACE with lipiodol is less than 50% when MR is considered the reference standard [25,26]. These findings indicate constraints in the assessment of TACE treatment response on CT scans through conventional visual assessment.

Radiomics is a technique employed for the quantitative characterization of medical images, involving high-dimensional, quantitative data [27]. In contrast to conventional methods that treat medical images as visual data for manual inspection, radiomics presents a novel approach to extracting information embedded within the medical images [28], making it possible to identify high-dimensional variables beyond semantic features obtained through conventional visual assessments.

There have been several radiomics studies aiming to identify variables associated with the response to TACE in patients with HCC. Several studies have shown that a radiomics model based on MRI effectively predicts the response to TACE [29–32]. Bernatz et al. found that a model, incorporating radiomics of post-TACE CT and clinical scores, effectively predicted the prognosis after TACE [33]. This model includes a single radiomic feature, Large Dependence High Gray Level Emphasis, and achieved a C-index of 0.67 for overall survival. This is in line with the results of our study that Large Dependence High Gray Level Emphasis in follow-up CT scans was the most significant feature associated with overall survival. This feature assesses the joint distribution of large dependence in-
volving higher gray-level values [34]. Furthermore, it was one of the key features for distinguishing between histologic grades 1 and 3 in an MRI radiomics study of HCC [35]. In our study, the model was created incorporating a total of four features from baseline and post-TACE CT and demonstrated a C-index score of 0.582 and time-dependence AUROC of 0.6 for overall survival. Although the C-index and AUROC were not excellent, the radiomic features extracted in this process predicted treatment response better than manual analyses.

Our study has several limitations. First, the study was limited to a relatively small number of samples, which may cause instability in the feature values. The sample size, especially the number of patients in the prospective test data (57 cases), may limit the generalizability of the study results. Future studies with larger samples are needed to fully validate the findings. Second, the quality and consistency of radiomic features extracted from various CT scans can be influenced by various factors, including imaging protocols, equipment variations, and image artifacts. Third, for the training phase, the authors used retrospective data. Retrospective data collection can introduce bias and confounding factors that can affect the accuracy and reliability of AI algorithms.

Radiomics features can provide valuable quantitative information, but their clinical interpretation and integration into existing prognostic models and treatment decision-making processes can be challenging. The use of AI algorithms in medical decision-making raises ethical and regulatory considerations regarding patient privacy, consent, and potential biases. Addressing these limitations through rigorous research design, data validation, and collaboration between clinicians, radiologists, and data scientists can improve the reliability and clinical usefulness of AI algorithms in predicting treatment outcomes in HCC patients treated with TACE.

In conclusion, a radiomics-based AI model would be beneficial for evaluating treatment response and predicting overall survival through CT in patients with HCC treated by TACE.

**Author Contributions:** Conceptualization, S.C.; methodology, S.C.; software, Y.K. and J.J.; validation, S.Y. and S.C.; writing—original draft preparation, S.Y. and S.C.; writing—review and editing, all authors; funding acquisition, S.C.; supervision, S.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by Guerbet, but the authors had complete control of the data and information submitted for publication at all times.

**Institutional Review Board Statement:** This study was approved by the institutional review board of the Gil Medical Center (GAIRB2019038) and was performed in accordance with the Declaration of Helsinki.

**Informed Consent Statement:** All participants provided written informed consent.

**Data Availability Statement:** Dataset available on request from the authors. The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.