

Radiogenomics in Hepatocellular Carcinoma: Machine Learning Based CT Texture Analysis in Predicting β -Catenin Mutation Status

Guo ZiYi

First Affiliated Hospital of JinZhou Medical University, China

Jiang LiPeng

First Affiliated Hospital of JinZhou Medical University, China

Zhu ZhiTu (✉ zhu_zhitu@163.com)

First Affiliated Hospital of JinZhou Medical University, China

Research Article

Keywords: Hepatocellular, HCC, patients, clinical

Posted Date: February 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-203679/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

OBJECTIVE.

The purpose of this study is to evaluate the potential value of CT radiomics in predicting the mutation status of β -catenin in patients with hepatic cell cancer (HCC).

MATERIALS AND METHODS.

In this retrospective study, 43 patients with hepatic cell HCC (18 without β -catenin mutation and 15 with β -catenin mutation) were identified in The Cancer Genome Atlas–hepatic liver Cell Carcinoma database (TCGA-LIHC). To create stable models, the data were augmented to a total of 202 labeled samples (131 without β -catenin mutation and 73 with β -catenin mutation) by obtaining up to five different samples per patient. Extraction of large amounts of image features from portal phase contrast-enhanced CT images had been performed on an open-source software package (Pyradiomics, version 2.1.2.). Reproducibility analysis (intraclass correlation, run ICCs in SPSS 18.0) was performed by two radiologists. Classification problem is about β -catenin gene mutation status. Machine Learning based classifications were performed using the Pycaret (version 2.1.2) software. The main performance metric was the AUC value.

RESULTS.

Of 828 extracted texture features, 759 had excellent reproducibility. Using 10 selected features, the Extra Trees Classifier algorithm correctly classified 93.4% of the HCCs in terms of β -catenin mutation status (AUC value, 0.9741); the CatBoost Classifier algorithm correctly classified 91.9% of the HCCs (AUC value, 0.9692); Gradient Boosting Classifier algorithm correctly classified 91.1% (AUC value, 0.9722). All the three advanced algorithms performed above 90% accuracy.

CONCLUSION.

Machine Learning-based high-dimensional quantitative CT radiomics analysis might be a feasible and potential method for predicting β -catenin mutation status in patients with HCC.

Introduction

Hepatocellular carcinoma (HCC) is a major type of primary malignant hepatic tumor and the third cause of cancer-related mortality worldwide ^[1]. The diagnosis and treatment for HCC had been significantly affected by genomic characteristics of the tumors ^{[2][3]}.

WNT/ β -catenin signaling pathway activation has been known as an important signal for hepatic carcinogenesis ^[4]. Abnormal Wnt– β -catenin signaling due to β -catenin mutation had been found in 30–40% of patients with HCC ^[5]. The β -catenin mutation had been proved to accelerate bile production in a higher grade of differentiation HCCs ^[6]. HCC with β -catenin mutation may be a special subtype that

shows specific pathologic and clinical features, so β -catenin mutation might associate with interesting radiologic feature in clinical settings^[7].

Most of HCCs show the decreased uptake of gadoxetic acid disodium or Gd-EOB-DTPA (Primovist, Schering, Berlin, Germany) comparing with normal liver tissue in the hepatobiliary phase in cancer cells^[8]. A small portion of HCC nodules are reported to uptake more EOB^[9]. Therefore, the imaging diagnosis of HCCs with β -catenin mutation may be important in daily clinical practice.

The term 'radiogenomics' is used herein in the context of the mutual relation or connection between the features of tissue-scale cancer imaging and molecular features of malignancies such as gene expressions.^[10] An emerging related quantitative technique is the computed tomography texture analysis (CTTA), which characterizes the heterogeneity of a lesion inside a specific region of interest (ROI), and hence utilizes pixel attributes and image histograms to obtain quantitative texture parameters.^[11] This technique demonstrated its utility as an imaging biomarker, as a predictor of patient outcomes and overall survival, and as an estimator of therapy response for multiple tumors.^[12] We tried to look for a connection between β -catenin mutation and imaging biomarkers from CTTA-based radiomics.

The existence of a mutual correlation between CTTA-based radiomics and β -catenin mutation has not yet been investigated in the open literature, according to the best that we know. We explore whether such a correlation really exists by investigating the possibility of some relation and connection between CTTA-derived quantitative parameters and mutations of β -catenin in patients affected by hepatocellular carcinoma (HCC).

Results

The diameters of the lesions on axial portal phase contrast enhance images were as follow: (1) HCC with β -catenin mutation: mean \pm standard deviation [SD], 4.95 ± 1.38 cm; median, 4.87cm; interquartile range [IQR], 3.88– 5.40cm; (2) HCC without β -catenin mutation: mean \pm SD, 5.02 ± 1.45 cm; median, 5.07 cm; IQR, 4.06–6.66 cm. The numbers of voxels in the ROIs, each containing a tumor, were as follow: (1) HCC with β -catenin mutation: mean \pm standard deviation [SD], 4574.86 ± 6698.65 ; median, 1875; interquartile range [IQR], 860– 5578; (2) HCC without β -catenin mutation: mean \pm SD, 7093.93 ± 5658.79 ; median, 5984; IQR, 1761.5–10430.5.

Dimension Reduction by Reproducibility Analysis and Feature Selection

The study flow-chart is displayed in Fig. 1.

In this study, the target variable is the β -catenin mutation status (yes = 1, no = 0). The data are radiomics parameters, which had been selected with high ICC above 0.9. Total 759 columns and 192 samples (ID)

had been included in the following analysis. 58 samples had been used as test/hold-out set (train/test = 70/30).

Total 18 models had been trained and evaluated using cross validation. In order to compare all models about their performances, all models in the model library had been trained and scored them using stratified cross validation for metric evaluation.

The top 3 models with a score grid that shows average accuracy, AUC, recall, precision, F1, kappa, and MCC with training dataset as shown in Table 2. Of 828 texture features, 759 had excellent reproducibility (intraclass correlation coefficient, ≥ 0.9). Hence, these features were included in the additional dimension reduction steps.

After classifier-specific feature selection algorithms, the number of selected features was reduced to 10. In order to estimate the predict model function on unseen data, 10 sample records (0.05% fraction of total 202 segmentations) had been withheld from the original dataset to be used for predictions.

The plot function takes a trained model object and returns a plot based on the test / hold-out set can be used to analyze the performance across different aspects such as AUC, confusion matrix, decision boundary.

The AUC plot, Precision-Recall Curve and Feature Importance Plot from extra trees classifier, CatBoost classifier, and gradient boosting classifier models had been displayed in Fig. 2 and Fig. 3.

Figure 2 displays receiver operating characteristic (ROC) curves obtained from CTTA of portal phase enhancement CT of HCC using Extra Trees, CatBoost, Gradient Boosting Classifier. For top three models, the top 10 selected features were all wavelet-transformed images parameters. The wavelet derived features had been proved as the most significant features in all the three top models.

The selected features important plots for Extra Trees Classifier, CatBoost Classifier, and Gradient Boosting Classifier models are presented in Fig. 3.

Machine Learning–Based Classifications

In order to predict the test/hold-out set and reviewing the evaluation metrics, the finalized models function fit the model onto the complete dataset including the test/hold-out sample (30% in this study, 58 samples) and unseen samples.

The two good performance models are Extra Trees Classifier and Cat Boost models with accuracy on test data and unseen data as follow: (1) the accuracy of Extra Trees Classifier on test/hold-out set is 0.9138 compared to 0.9342 achieved on the Cross Validation results, accuracy of predicting on unseen data (10 samples) is 0.9; (2) the accuracy of Cat Boost on test/hold-out set is 0.9138 compared to 0.911 achieved on the Cross Validation results, accuracy of predicting on unseen data (10 samples) is 1.0.

Discussion

We devote this work to exploring the effectiveness of the machine learning (ML) methodology of high-dimensional CT radiomics in making a prudent or educated guess of the β -catenin mutation status of HCC patients. Our results indicate that CT radiomics using different ML classifiers (the extra trees classifiers, and the CatBoost classifiers) is potentially useful for predicting HCCs irrespective of whether the β -catenin mutation exists or not.

We recall that radiomics is a medical technique that applies algorithms of data characterization to radiographic medical images for extracting a large number of features^[23]. CT-based radiomics analysis had been used to predict survival of patients with metastatic colorectal cancer^[24]. Radiomics could also be used to predict response of individual HER2-amplified colorectal cancer liver metastases, as well as the biomarkers of molecular subtype prognosis^[25].

As radiogenomics could reveal the relationship between imaging features and genomic features^[26], radiogenomics could be used to bridge imaging and genomics. Our current study may have important practical and clinical implications.

The β -catenin mutation in HCCs may promote immune escape and might affect responsiveness to therapeutic procedures^[27]. The evaluation of genetic mutations of liver cancer could prove impractical if implemented for every patient. Nevertheless, the radiomic features derived from CT texture analysis might provide potential biomarkers for predicting HCCs (whether the β -catenin mutation exists or not) after the validation of such biomarkers in larger datasets. Moreover, we anticipate that new biomarkers and models could be developed through forthcoming research that might involve larger datasets and different feature selection algorithms as well as supporting ML schemes.

In our analysis, the radiomics parameters in each ML-based model were similar. We used cross validation, testing on unseen data methods to optimize the model performance. Total 18 types of classifiers had been selected during the feature selection process. Several experiments with various ML classifiers might be needed to find the best ML scheme, when less data available.

One of the well-known challenges in the field of radiomic is interpretation of the selected features in model development, even if they were validated^[28]. Regarding radiomics of HCCs for identifying β -catenin mutation status^[29], the selected features might represent some kinds of information that are associated with pathological stage, or differentiation grade, which are correlated with β -catenin mutation status^[30].

A few limitations of this study should be addressed. First, as a retrospective study design, this study provided some inferior level of evidence. Second, ML-based classifiers might risk overfitting induced by the small and imbalanced patient population. We strived to reduce this expected overfitting problem through the application of data augmentation techniques to increase the number of the labeled samples, a truly fruitful method for overcoming overfitting in ML-based classification. Third, although 3D

segmentation could represent radiomics information more effectively, we just used the largest 2D slice and its adjacent consecutive upper and lower slices for CT radiomics, because most former clinical research on HCCs had been based on a single segmentation or segmentations of a few slices. Fourth, we derived the imaging data from TCGA-LIHC on The Cancer Imaging Archive (TCIA) website, which includes patients from different centers and sources using different image acquisition protocols, just the same as in standard clinical practice. To minimize various kinds of variabilities, all image samples underwent normalization and pixel rescaling procedures as shown in the methods section. The current technique had been proved to reduce both variabilities and bias. Fifth, we used the same dataset for training, validation and testing, an action that could certainly be viewed as a bias, and hence we implemented a 10-fold cross-validation procedure to minimize such a potential bias. It is obvious that independent external datasets are needed to validate the performance of the classifiers in any further exploration. Sixth, we included the portal phase images only in the analysis since they are widely available. Further research is warranted for the unenhanced CT or arterial phase-enhanced CT. Seventh, we evaluated only the β -catenin mutation status because the corresponding patient group possessed sufficient imaging data that satisfied our criteria and attained clinical usefulness with an effective prognostic value in this study. Ninth and finally, all radiogenomic studies suffer from the same common problem, namely the possibility of some discrepancy between the data present on imaging studies and the small sample used for genomic analysis^[31].

In conclusion, CT radiomics based on machine learning is shown to be a feasible and potentially successful method for predicting β -catenin mutation status in HCC patients. Due to the advantage of routine acquisition of enhanced CT images, we prudently propose that this radiogenomics approach could be used as a future clinical decision support tool in larger and prospective trials.

Materials And Methods Ethics

All materials and methods were performed in accordance with relevant guidelines and regulations. All experimental protocols were approved by the institutional review board at the JinZhou Medical University (Authorization Number: JMU20210217). All patients in the study were deidentified. The data were publicly available for scientific purposes.

Data Source

All genomic and clinical data in this study had been obtained from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) database^[13]. Available pretreatment imaging studies were downed in DICOM format from Cancer Imaging Archive website^[14]. TCGA- LIHC database included 429 patients with HCC (total number of β -catenin mutation 103, frequency of mutation about 26.8%). However, the imaging data of only 97 patients were available for use.

Inclusion Criteria

To create a uniform imaging protocol, the study included only those patients with available preoperative portal phase contrast-enhanced CT (CECT) images that were acquired with maximum tube voltage (140 kV), slice thickness (5 mm or less), no slice overlap. Because the arterial phase and delay phase was not available in the patients' imaging data, we only include portal phase images.

Exclusion Criteria

Patients were excluded from study due to poor-quality CECT images (significant image noise, significant artifacts, or other quality issues), multiple tumors (as a result of information in the database creating uncertainty about which tumor had the β -catenin mutation). Fewer than five patients' imaging studies would be excluded due to minimize the heterogeneity of the imaging protocol.

Patient Population

Total 43 patients with total 43 HCCs tumors had met the eligibility criteria of the study. All the information of demographics and clinical characteristics of the patients had been displayed in Table 1. The list of included patients, in addition to their corresponding patient codes both in the TCGA-LIHC database and on The Cancer Imaging Archive website, are presented in Appendix.

Data Augmentation

Data augmentation has been proved and is considered a powerful method for avoiding overfitting when there is a small amount of data. It has been successfully applied in many different machine learning-based classification tasks ^[15]. Given the small number of patients in our study might lead to potential overfitting, we naturally augmented the labeled data in our study by obtaining samplings from different levels of the tumors (as shown in Fig. 1).

HCCs were sampled by 3–5 different and consecutive slices around the largest diameter center slice, unless the last slice of the tumor was affected by partial volume ^[15]. The augmentation resulted in 202 labeled segmentation data (131 without the β -catenin mutation and 71 with the β -catenin mutation) from 43 HCCs (28 without the β -catenin mutation and 15 with the β -catenin mutation). We considered using actual data derived from the multiple segmentations or samplings, rather than artificial or synthetic data.

Reference Standard

The reference standard for classification was the presence or absence of a β -catenin mutation as reported in the TCGA-LIHC database ^{[14][16]}. Of the 43 patients included in the present study, 15 had a β -catenin mutation. No β -catenin mutation was identified in 28 patients.

Segmentation

The tumor segmentations were manually performed using 3D Slicer software (version 4.8.1) ^[17]. Up to five segmentations were obtained for each lesion with about 2 mm of margin shrinkage from the lesion contour. The initial segmentation was done on the axial image slice representing the largest cross-

sectional area of the tumor. The additional segmentations were then performed on the adjacent consecutive upper and lower slices. Shrinkage was performed using the margin shrinkage function of the software that creates the procedure equally in every direction.

Image Processing

A DICOM image format was used in each step of the analysis. Before texture feature extraction, all images were normalized, rescaled, and discretized^[18]. To minimize inter scanner effects, all datasets were normalized by centering the pixel image intensity values at the mean value with SD. We set the scaling factor to 1. Pixel spaces in all image slices were resampled and rescaled to the resolution of $1 \times 1 \text{ mm}^2$ in order to be considered into many texture features necessitate the same spatial resolution and require the pixel size to be comparable. Cubic B-spline interpolation was used for rescaling. The gray-level discretization was done in the matrix representation of the gray levels in the segmentation with a bin width value of 0.01.

Feature Extraction

Texture features were extracted using an open source software package for the extraction of radiomic data from medical images (Pyradiomics, version 2.2.0)^[18]. The features were extracted from the original, filtered, and wavelet-transformed images^[19]. The Laplacian of Gaussian filter was used for image filtration, with values of 1 mm, 3 mm, and 5 mm denoting fine, medium, and coarse patterns, respectively.

The extracted texture features included first-order features, the gray-level dependence matrix, gray-level co-occurrence matrix, gray-level run-length matrix, gray-level size zone matrix, neighboring gray-tone difference matrix, and wavelet-based texture features. The total number of features extracted per lesion was 828. Detailed descriptions and mathematic formulas for these features have been described elsewhere^[18].

Feature Reduction by Reproducibility Analysis

To assess the reproducibility of the texture features^[20], two radiologists independently segmented 43 randomly selected tumors. Both radiologists were blinded to the β -catenin mutation status. Intra-class correlation coefficient values^[21] were calculated for each texture feature with the use of statistical software (run ICCs in SPSS 18.0). Only the features with an intra-class correlation coefficient of 0.9 or greater, which indicated excellent reproducibility, were included in additional dimension reduction steps.

Feature Reduction and Prediction by Machine Learning–Based Classifications

Machine Learning based classifications were performed using the Pycaret software, version 2.1.2^[22]. Total 18 models of classifier had been made with the 10-fold cross-validations. The performance of classifiers had been mainly evaluated and compared basing on AUC value. The values of accuracy,

sensitivity, specificity, precision, the F-measure, and the Matthews correlation coefficient (MCC) had all been calculated.

The classification module in Pycaret is a supervised machine learning module, which is used for classifying the elements into a binary group based on various techniques and algorithms. In the current study of classification problem include β -catenin gene mutation detection found (positive vs. negative).

In order to demonstrate the prediction function on unseen data, sample of 10 records (5% total samples) has firstly been with-hold from the original dataset to be used for predictions. The second step is to creates the transformation pipeline to prepare the data for modeling and deployment. The target column indicated status of β -catenin gene mutation. The third step is comparing all models to evaluate performance. The output shows average Accuracy, AUC, Recall, Precision, F1, Kappa, and MCC crossing the 10-folds along with training times. Total over 15 models using cross validation had been trained. Accuracy (highest to lowest) is used to model selection. AUC, feature importance plot and confusion matrix had been used to analyze the performance across different aspects based on the test / hold-out set. The last step is finalizing the model and predicting on unseen data.

References

1. Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology*. 2019;156:477-491.e1. doi: 10.1053/j.gastro.2018.08.065.
2. Ahn JC, Teng PC, Chen PJ, etal. Detection of Circulating Tumor Cells and Their Implications as a Biomarker for Diagnosis, Prognostication, and Therapeutic Monitoring in Hepatocellular Carcinoma. *Hepatology*. 2020 Feb 4. doi: 10.1002/hep.31165.
3. Pinyol R, Montal R, Bassaganyas L, etal. Molecular predictors of prevention of recurrence in HCC with sorafenib as adjuvant treatment and prognostic factors in the phase 3 STORM trial. *Gut*. 2019;68:1065-1075. doi: 10.1136/gutjnl-2018-316408.
4. Vilchez V, Turcios L, Marti F, etal. Targeting Wnt/beta-catenin pathway in hepatocellular carcinoma treatment. *World J Gastroenterol*. 2016;22:823-32. doi: 10.3748/wjg.v22.i2.823.
5. Perugorria MJ, Olaizola P, Labiano I, etal. Wnt-beta-catenin signalling in liver development, health and disease. *Nat Rev Gastroenterol Hepatol*. 2019;16:121-136. doi: 10.1038/s41575-018-0075-9.
6. Charawi S, Just PA, Savall M, Abitbol S, etal. LKB1 signaling is activated in CTNNB1-mutated HCC and positively regulates beta-catenin-dependent CTNNB1-mutated HCC. *J Pathol*. 2019;247:435-443. doi: 10.1002/path.5202.
7. Auer TA, Fehrenbach U, Grieser C, etal. Hepatocellular adenomas: is there additional value in using Gd-EOB-enhanced MRI for subtype differentiation? *Eur Radiol*. 2020;30:3497-3506. doi: 10.1007/s00330-020-06726-8.
8. Kitao A, Matsui O, Yoneda N, etal. Hepatocellular Carcinoma with beta-Catenin Mutation: Imaging and Pathologic Characteristics. *Radiology*. 2015;275:708-17. doi: 10.1148/radiol.14141315.

9. Bise S, Frulio N, Hocquelet A, et al. New MRI features improve subtype classification of hepatocellular adenoma. *Eur Radiol.* 2019;29:2436-2447. doi: 10.1007/s00330-018-5784-5.
10. Renzulli M, Biselli M, Brocchi S, et al. New hallmark of hepatocellular carcinoma, early hepatocellular carcinoma and high-grade dysplastic nodules on Gd-EOB-DTPA MRI in patients with cirrhosis: a new diagnostic algorithm. *Gut.* 2018;67:1674-1682. doi: 10.1136/gutjnl-2017-315384.
11. Zhou M, Leung A, Echegaray S, et al. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. *Radiology.* 2018;286:307-315. doi: 10.1148/radiol.2017161845.
12. Brenet Defour L, Mulé S, Tenenhaus A, Piardi T, et al. Hepatocellular carcinoma: CT texture analysis as a predictor of survival after surgical resection. *Eur Radiol.* 2019 ;29:1231-1239. doi: 10.1007/s00330-018-5679-5.
13. <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/studied-cancers/liver>
14. <https://wiki.cancerimagingarchive.net/display/Public/TCGA-LIHC>
15. Kocak B, Durmaz ES, Ates E, et al. Radiogenomics in Clear Cell Renal Cell Carcinoma: Machine Learning-Based High-Dimensional Quantitative CT Texture Analysis in Predicting PBRM1 Mutation Status. *AJR Am J Roentgenol.* 2019;212:W55-W63. doi: 10.2214/AJR.18.20443.
16. Luke JJ, Bao R, Sweis RF, et al. WNT/beta-catenin Pathway Activation Correlates with Immune Exclusion across Human Cancers. *Clin Cancer Res.* 2019;15;25:3074-3083. doi: 10.1158/1078-0432.CCR-18-1942.
17. <https://www.slicer.org/>
18. Kulkarni A, Carrion-Martinez I, Dhindsa K, et al. Pancreas adenocarcinoma CT texture analysis: comparison of 3D and 2D tumor segmentation techniques. *Abdom Radiol (NY).* 2020 Sep 16. doi: 10.1007/s00261-020-02759-1.
19. van Griethuysen, J. J. M., Fedorov, A., Parmar, C., et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 2017; 77, e104–e107. doi: 10.1158/0008-5472.CAN-17-0339.
20. Fornaçon-Wood I, Mistry H, Ackermann CJ, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol.* 2020 ;30:6241-6250. doi: 10.1007/s00330-020-06957-9.
21. Warrens MJ. Transforming intraclass correlation coefficients with the Spearman-Brown formula. *J Clin Epidemiol.* 2017 ;85:14-16. doi: 10.1016/j.jclinepi.2017.03.005.
22. <https://pycaret.org/>
23. Gillies RJ, Kinahan PE, Hricak H. "Radiomics: Images Are More than Pictures, They Are Data". *Radiology.* 278 (2): 563–77. doi:10.1148/radiol.2015151169.
24. Mühlberg A, Holch JW, Heinemann V, et al. The relevance of CT-based geometric and radiomics analysis of whole liver tumor burden to predict survival of patients with metastatic colorectal cancer.

- Eur Radiol. 2021 ;31:834-846. doi: 10.1007/s00330-020-07192-y.
25. Giannini V, et al. Radiomics predicts response of individual HER2-amplified colorectal cancer liver metastases in patients treated with HER2-targeted therapy. *Int J Cancer*. 2020; 147:3215-3223. doi: 10.1002/ijc.33271.
 26. Hong EK, Choi SH, Shin DJ, et al. Radiogenomics correlation between MR imaging features and major genetic profiles in glioblastoma. *Eur Radiol*. 2018;28:4350-4361. doi: 10.1007/s00330-018-5400-8.
 27. Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, et al. beta-Catenin Activation Promotes Immune Escape and Resistance to Anti-PD-1 Therapy in Hepatocellular Carcinoma. *Cancer Discov*. 2019;9:1124-1141. doi: 10.1158/2159-8290.CD-19-0074.
 28. Bakr S, Echegaray S, Shah R, et al. Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study. *J Med Imaging (Bellingham)*. 2017;4:041303. doi: 10.1117/1.JMI.4.4.041303.
 29. Nault JC, Martin Y, Caruso S, et al. Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology*. 2020;71:164-182. doi: 10.1002/hep.30811.
 30. Kim E, Lisby A, Ma C, et al. Promotion of growth factor signaling as a critical function of beta-catenin during HCC progression. *Nat Commun*. 2019;23;10:1909. doi: 10.1038/s41467-019-09780-z.
 31. Arefan D, Chai R, Sun M, et al. Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features. *Med Phys*. 2020;47:6334-6342. doi: 10.1002/mp.14538.

Tables

Table 1
Demographic characteristics of the sample population (total 43 Patients).

Characteristics	Value
Mean age (year)	59.74
Sex	
female	14
male	29
Disease stage	
Stage I-II	29
Stage II-IV	14
Histopathologic nuclear grade	
grades 1–2 (low)	27
grades 3–4 (high)	16
Beta-Catenin mutation	
Absent	28
present	15
Vascular tumor invasion	
Present	10
Absent	33

Table 2

Classification results for hepatocellular carcinoma (HCC) with β -catenin mutation and without β -catenin mutation using texture features and Extra Trees Classifier, catboost and Gradient Boost analysis.

Model	Accuracy	AUC	Recall	Precision	F1 score	Kappa	MCC
Extra Trees Classifier	0.9341	0.9739	0.88	0.935	0.9022	0.8531	0.8586
CatBoost Classifier	0.9187	0.9692	0.84	0.935	0.8744	0.8166	0.8285
Gradient Boosting Classifier	0.911	0.9722	0.9	0.8798	0.8843	0.8124	0.819
Note: TP = True Positive; TN = True Negative; FP = False Positives; FN = False Negatives							

Accuracy = $\frac{TP + TN}{TP + FP + FN + TN}$;

AUC = area under the curve;

Recall = $TP / (TP + FN)$;

Precision = $TP / (TP + FP)$;

Specificity = $TN / (TN + FP)$;

F1 score = $2 * Precision * Recall / (Precision + Recall)$

$$MCC = \left(\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \right)$$

Figures

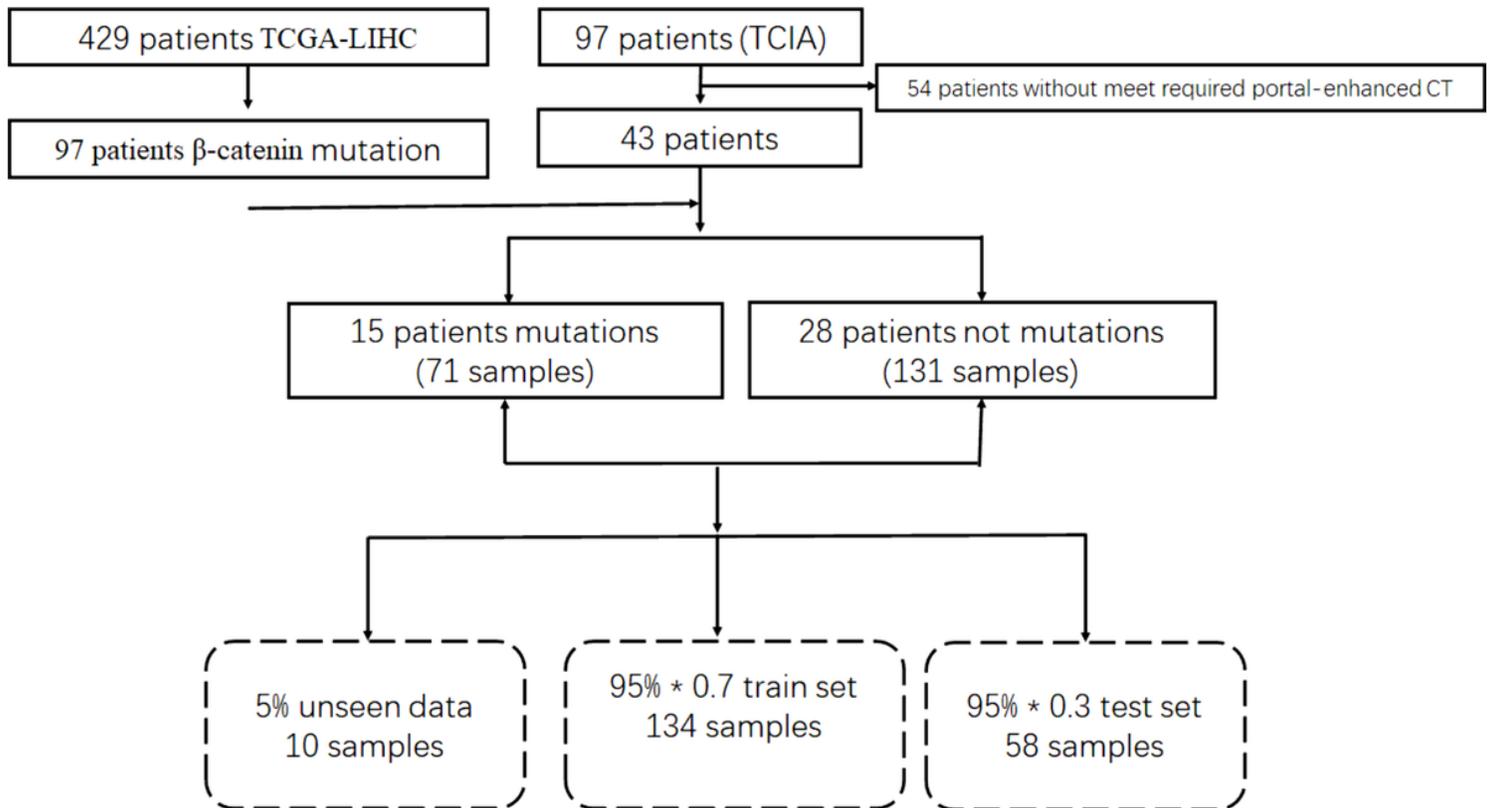


Figure 1

flowchart of the study.

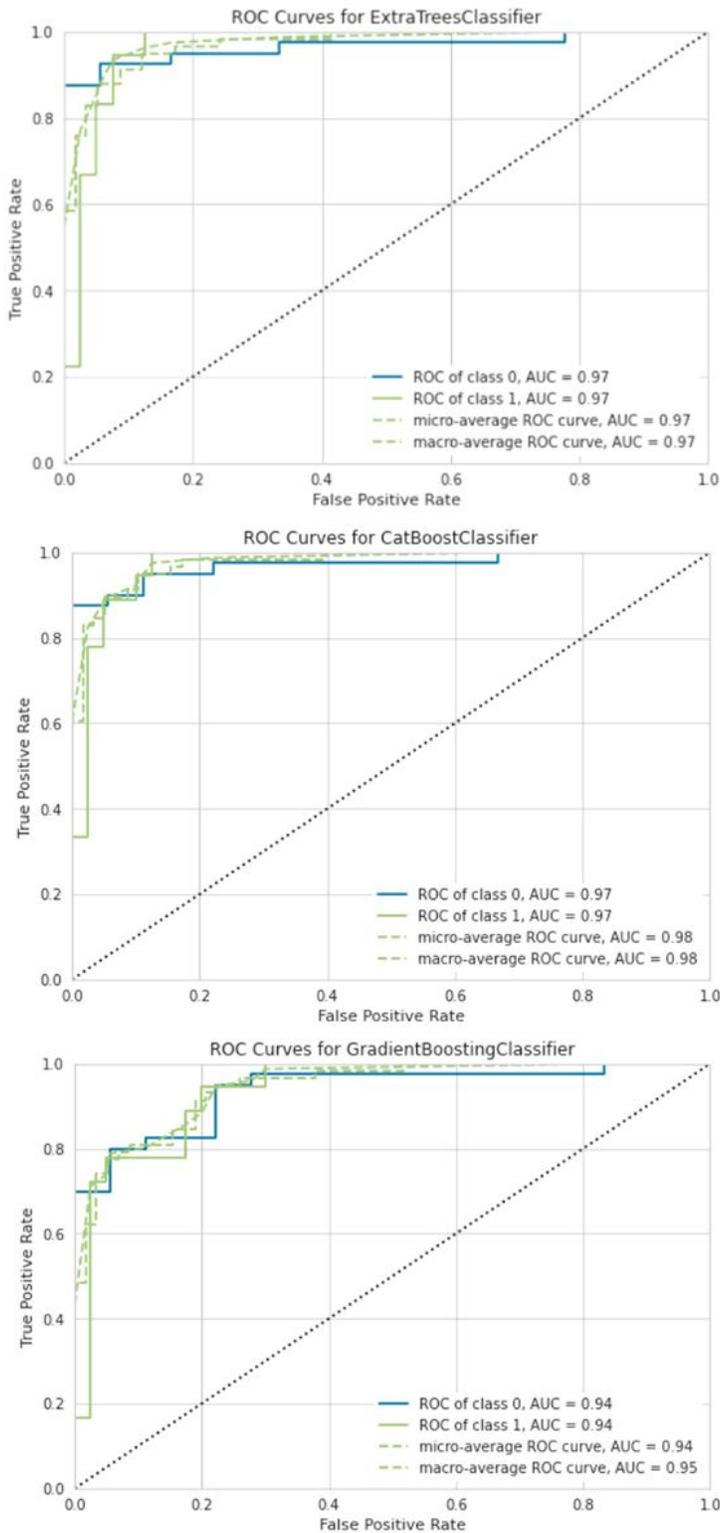


Figure 2

ROC curves obtained from CTTA of HCC portal phase enhancement CT imaging using Extra Trees, Cat Boost, Gradient Boost Classifier. (a) Classification between β -catenin mutation and without β -catenin mutation using Extra Trees Classifier. The areas under the curve (AUC) are 0.97 and 0.97, respectively; (b) Classification between β -catenin mutation and without β -catenin mutation using Cat Boost Classifier. The areas under the curve (AUC) are 0.97 and 0.97, respectively; (c) Classification between β -catenin mutation

and without β -catenin mutation using Gradient Boosting Classifier. The AUCs are 0.94 and 0.94, respectively.

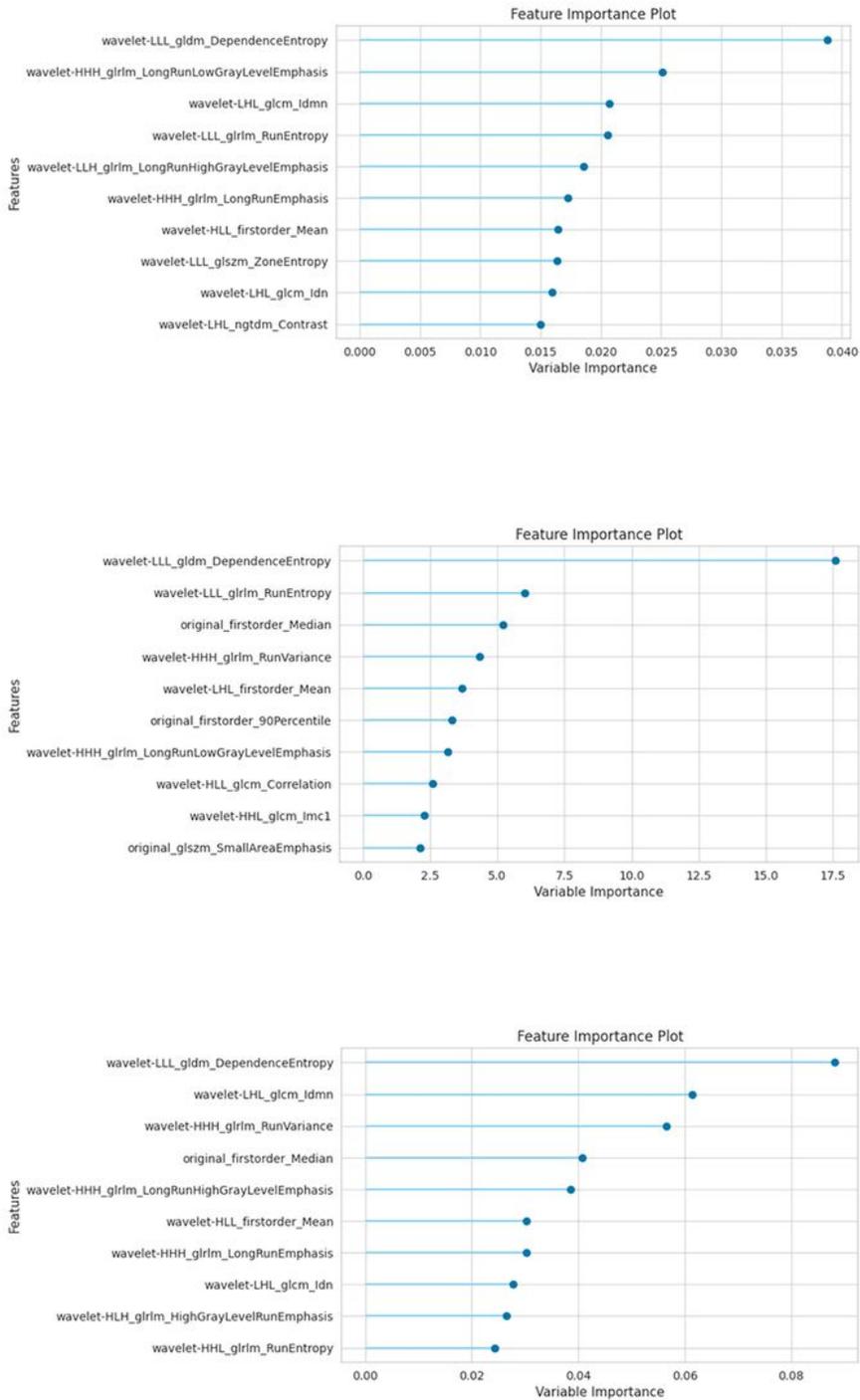


Figure 3

The selected features important plots for Extra Trees Classifier, Cat Boost Classifier, and Gradient Boosting Classifier models. (a) listing of 10 features for the classification between β -catenin mutation and without β -catenin mutation using Extra Trees Classifier; (b) listing of 10 features for the classification

between β -catenin mutation and without β -catenin mutation using Cat Boost Classifier. (c) listing of 10 features for the classification between β -catenin mutation and without β -catenin mutation using Gradient Boosting Classifier.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformationfile20210216.docx](#)