HEPATOBILIARY

Application of a machine learning approach to characterization of liver function using ^{99m}Tc-GSA SPECT/CT

Masatoyo Nakajo[1] · Megumi Jinguji[1] · Atsushi Tani[1] · Daisuke Hirahara[2] · Hiroaki Nagano[1] · Koji Takumi[1] · Takashi Yoshiura[1]

Abstract

Purpose To assess the utility of a machine-learning approach for predicting liver function based on technetium-99 m-galactosyl serum albumin (^{99m}Tc-GSA) single photon emission computed tomography (SPECT)/CT.
Methods One hundred twenty-eight patients underwent a ^{99m}Tc-GSA SPECT/CT-based liver function evaluation. All were classified into the low liver-damage or high liver-damage group. Four clinical (age, sex, background liver disease and histological type) and 8 quantitative ^{99m}Tc-GSA SPECT/CT features (receptor index [LHL15], clearance index [HH15], liver-SUVmax, liver-SUVmean, heart-SUVmax, metabolic volume of liver [MVL], total lesion GSA [TL-GSA, liver-SUVmean × MVL] and SUVmax ratio [liver-SUVmax/heart-SUVmax]) were obtained. To predict high liver damage, a machine learning classification with features selection based on Gini impurity and principal component analysis (PCA) were performed using a support vector machine and a random forest (RF) with a five-fold cross-validation scheme. To overcome imbalanced data, stratified sampling was used. The ability to predict high liver damage was evaluated using a receiver operating characteristic (ROC) curve analysis.
Results Four indices (LHL15, HH15, heart SUVmax and SUVmax ratio) yielded high areas under the ROC curves (AUCs) for predicting high liver damage (range: 0.89–0.93). In a machine learning classification, the RF with selected features (heart SUVmax, SUVmax ratio, LHL15, HH15, and background liver disease) and PCA model yielded the best performance for predicting high liver damage (AUC = 0.956, sensitivity = 96.3%, specificity = 90.0%, accuracy = 91.4%).
Conclusion A machine-learning approach based on clinical and quantitative ^{99m}Tc-GSA SPECT/CT parameters might be useful for predicting liver function.

Keywords Asialoglycoprotein receptor · ^{99m}Tc-GSA · SPECT/CT · Machine learning · Support vector machine

Introduction

In a patient with a liver tumor, a pretreatment estimation of the remnant liver function is essential for reducing the morbidity and mortality associated with treatment [1]. The asialoglycoprotein receptor (ASGPR) is expressed on the sinusoidal surfaces of hepatocytes and contributes to the clearance of circulating glycoproteins containing terminal galactose residues [2]. Technetium-99m-galactosyl serum albumin (^{99m}Tc-GSA), an asialoglycoprotein analog, is taken up exclusively by the liver and binds specifically to ASGPR on hepatocytes [3, 4]. Decreased ASGPR expression has been reported in patients with liver damage [5]. Thus, ^{99m}Tc-GSA scintigraphy enables the direct estimation of functioning hepatocytes and is very useful for evaluations of liver function [6].

The receptor index (LHL15) and clearance index (HH15), which are calculated from liver and heart time-activity data, are the most commonly used parameters in liver function assessments [7–9]. However, these parameters are obtained on planar images and do not provide sufficient anatomical information for an accurate liver function assessment because they lack three-dimensional (i.e., depth) information.

✉ Masatoyo Nakajo
toyo.nakajo@dolphin.ocn.ne.jp

1 Department of Radiology, Graduate School of Medical and Dental Sciences, Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima 890-8544, Japan

2 Department of Management Planning Division, Harada Academy, 2-54-4 Higashitaniyama, Kagoshima 890-0113, Japan

Single photon emission computed tomography (SPECT)/ computed tomography (CT) is a relatively new imaging modality that enables the near-simultaneous co-registration of functional and anatomical information acquired during the same imaging session and while the patient remains in the same bed position [10]. The main benefits of SPECT/CT is the improved localization of radioactivity. The standardized uptake value (SUV), a quantitative parameter used commonly during positron emission tomography, has recently been applied to bone SPECT/CT evaluations [11–13], and one report has investigated the SUV during [99mTc]-GSA SPECT/CT evaluations of liver fibrosis [14]. Although reports have described various quantitative liver function analyses based on [99mTc]-GSA scintigraphy, it remains unclear which methods are most useful in this context [7]. Therefore, some kind of new reliable method for evaluating or predicting of the liver function is required.

Machine learning is a field of computer science based on pattern recognition and computational learning that can identify patterns and relationships formed from complex multidimensional databases [15]. It relies on computer algorithms to learn and identify complex interactions among all variables by minimizing the error between predicted and observed outcomes. Compared with conventional statistical methods, which rely on predetermined models, machine learning can deeply detect the interactions among variations and iteratively learn from data to update algorithms [16].

Recently, some studies have proposed classification methods based on a machine learning approach into the research of nuclear medicine [17–19]. However, to our knowledge, no study has previously investigated the efficacy of a quantitative [99mTc]-GSA SPECT/CT analysis using a machine learning approach for assessing the liver function. Therefore, this study aimed to assess the utility of a machine learning approach based on [99mTc]-GSA SPECT/CT for predicting the liver function.

Materials and methods

Patients

Our institutional review board approved this retrospective study and waived the requirement for written informed consent. From October 2016 to December 2019, 166 consecutive patients with liver tumors underwent pretreatment [99mTc]-GSA SPECT/CT to evaluate the liver function. The following inclusion criteria were applied: (1) measurement of three blood biochemical indices (albumin, total bilirubin and prothrombin time activity) within 1 week before or after [99mTc]-GSA SPECT/CT and (2) performance of an indocyanine green clearance test (ICGR15: ICG retention rate at 15 minutes) within 1 week before or after [99mTc]-GSA SPECT/CT.

Imaging protocols

Thoraco-abdominal planar and SPECT/CT images were acquired using a dual-head gamma camera with low-energy, high-resolution collimators and a multidetector (16-row) spiral CT scanner (Siemens Intevo SPECT/CT system; Siemens Medical Solutions USA, Inc., Hoffmann Estates, IL, USA). Immediately after the intravenous injection of 185 MBq of [99mTc]-GSA (Nihon Medi-Physics Co., Ltd. Tokyo, Japan), an anterior planar image was dynamically acquired in 30-second frames for 30 minutes. Data were digitally collected using a 128×128 matrix on a dedicated computer system (E-soft, Siemens Healthcare GmbH, Erlangen, Germany). SPECT (45 steps at 15 s/step, 360°, 128×128 matrix) was initiated 30 minutes after [99mTc]-GSA injection. Subsequently, CT images were acquired using a tube voltage of 130 kV and a dose-modulation algorithm with a quality reference mAs setting of 15 (CAREDose 4D; Siemens Medical Solutions USA, Inc.). SPECT data were reconstructed using Siemens xSPECT Quant with an Ordered Subset Conjugate-Gradient Minimizer (OSCGM) algorithm and the following settings: one subset, 60 iterations and a 10-mm Gaussian filter.

Image analysis

The planar, SPECT and SPECT/CT images were displayed on a workstation (Syngo.via; Siemens Healthcare GmbH, Erlangen, Germany). The anterior planar images were analyzed quantitatively by an experienced nuclear medicine radiologist who was aware of the study purpose but was blinded to the patients' clinical information. The liver and heart time activity curves were generated using regions of interests (ROIs) that covered the whole liver and heart, respectively. The LHL15 and HH15 were determined according to a previously reported method as follows: LHL15 = liver ROI count at 15 min/(liver ROI count + heart ROI count at 15 min); HH15 = heart ROI count at 15 min/ heart ROI count at 3 minutes [8, 20].

The following quantitative SPECT image analyses were performed by another nuclear medicine radiologist. First, a volume of interest (VOI) that contained the whole liver was placed manually and used to determine the liver SUVmax. Next, a threshold of 40% SUVmax was set to automatically delineate the VOI that met or exceeded this 40% threshold, and this VOI was used to calculate the liver SUVmean, metabolic liver volume (MLV) and total lesion [99mTc]-GSA uptake (TL GSA). The TL GSA was calculated as the liver SUVmean multiplied by the MLV. The heart SUVmax was also recorded, using the CT images as a reference. The software (Syngo.via; Siemens Healthcare GmbH, Erlangen, Germany) automatically calculated the liver SUVmax, liver SUVmean, heart SUVmax, MLV and TL GSA. The SUVmax ratio (liver SUVmax/heart SUVmax) was calculated manually.

Assessment of liver function

This study used the liver function criteria corresponding to liver damage types A, B and C, as proposed by the Liver Cancer Study Group of Japan (Table 1) [21]. The patients were subsequently classified into low liver damage (grade A) and high liver damage (grade B and C) groups.

Machine-learning approach

Four clinical (age, sex, background liver disease and tumor histological type) and 8 quantitative ^{99m}Tc-GSA (LHL15, HH15, liver SUVmax, liver SUVmean, MLV, TL GSA, heart SUVmax and SUVmax ratio) features were used to predict high liver damage employing machine-leaning approaches.

The machine learning classifiers in this study included a support vector machine (SVM) and random forest (RF). The SVM used the L2 penalty, while the RF used 100 estimators and a select entropy criterion. The ROC analysis was performed to compare the predictive performances of the models and the areas under the ROC curves (AUCs) were calculated.

To overcome imbalanced data, stratified sampling was used [22, 23]. To minimize the negative influence of overfitting, 5-fold cross-validation and feature selection were performed in this study [24, 25]. The 5-fold cross-validation randomly split the dataset into 5 subsets. For each repeated time, four subsets were used as the training group and the remaining subset was used as the testing data. This procedure was repeated 5 times, and each subset should be used exactly once as the testing group.

Gini impurity criterion was applied to determine the relative importance of individual features. The Gini impurity is the probability of an unseen case being incorrectly classified for a given decision or rule. Features with high Gini impurity (or low Gini importance) split the data into impure categories, while features that decrease Gini impurity are able to partition the data into purer classes with larger members. Thus, features with large decreases in Gini rank higher in importance for the model. The importance of a feature is defined as the decrease in the Gini impurity [26–28].

Not only feature selection method but also principal component analysis (PCA) method were conducted to compare the results of the classification based on the presence or absence of feature selection or dimensionality reduction, and the following 4 different models were obtained for the machine learning approaches; First, neither feature selection method nor PCA were performed, Second, feature selection was not performed, but PCA was conducted, Third, feature selection method using 5 best-ranked features was only performed, but PCA was not performed, Forth, both feature selection method using 5 best-ranked features and PCA were performed.

These machine learning approaches were performed using Orange version 3.24.1 (Bioinformatics Lab at University of Ljubljana, Slovenia), an open-source software package for data mining and visualization [29].

Statistical analysis

The Mann–Whitney U-test or chi-squared test was used to assess the difference between two quantitative variables or compare categorical data, as appropriate. A ROC analysis was performed to examine the diagnostic performance of each parameter for the prediction of high liver damage. The DeLong method was used to analyze the statistical significance of differences between AUCs [30].

Data are presented as medians and interquartile ranges (IQRs). All p values are two-sided, and a p value < 0.05 is considered to indicate a statistically significant difference. The statistical analysis was performed using MedCalc Statistical Software (MedCalc Software, Mariakerke, Belgium).

Results

Patient characteristics

Of the 166 initially identified patients, 38 were excluded from the analysis because they had not undergone an ICG clearance test. Finally, 128 patients (97 males, 31 females; mean [± SD] age, 71 ± 8 years; range, 39–90 years) were eligible for the analyses.

Seventy-six patients underwent liver resection, and the liver tumors received the following pathological diagnoses: hepatocellular carcinoma (HCC), 66 patients; cholangiocellular carcinoma, four patients and others, six patients. The remaining 52 patients underwent transcatheter arterial chemoembolization (TACE) for HCC.

Table 1 Liver damage criteria presented by the Liver Cancer Study Group of Japan

Variables	Liver damage		
	A	B	C
Ascites	None	Controllable	Uncontrollable
Total bilirubin (mg/dl)	< 2.0	2.0–3.0	> 3.0
Albumin (g/l)	> 3.5	3.0–3.5	< 3.0
ICGR15 (%)	< 15	15–40	> 40
Prothrombin activity (%)	> 80	50–80	< 50

A liver damage type of A, B, or C is determined according to the fulfillment of more than two categories. A case that fulfills more than two categories corresponding to two liver damage types should be classified as the higher liver damage type. For example, if a case fulfills three categories associated with liver damage A and two categories associated with liver damage B, the case is considered the latter type

ICGR15 indocyanine green retention rate at 15 min

Table 2 Clinical characteristics of patients in the low and high liver damage groups

Characteristic	Low liver damage	High liver damage	Total	p value
Number	100	28	128	
Age range (median)	39–90 (73)	54–82 (69)		0.054
Sex				0.62
Male	77	20	97	
Female	23	8	31	
Background liver disease				0.005
Hepatitis B	27	0	27	
Hepatitis C	19	5	24	
Non B, Non C	54	23	77	
Histology				0.88
HCC	92	26	118	
Others	8	2	10	

HCC hepatocellular carcinoma

Table 3 Comparison of HH15, LHL15 and SUV-related parameters between patients with low and high liver damage

	Low liver damage (n = 100)			High liver damage (n = 28)			p value
	Median	IQR	Range	Median	IQR	Range	
HH15[a]	0.59	0.56–0.63	0.46–0.75	0.70	0.66–0.75	0.63–0.85	< 0.001
LHL15[b]	0.92	0.89–0.93	0.79–0.97	0.85	0.81–0.87	0.72–0.91	< 0.001
Liver SUVmax	57.3	51.6–64.7	29.1–105.5	53.7	44.0–59.1	22.4–84.5	0.016
Liver SUVmean	37.3	33.9–42.1	19.9–57.7	33.2	28.1–38.7	15.4–44.2	0.005
MLV (cm³)	897.7	782.5–1006.3	547.9–1429.3	881.6	648.6–1008.5	524.6–1808.7	0.31
TL GSA	34176.3	28406.7–41204.3	182923.4–400096.7	27871.0	21509.8–34703.3	16035.2–46612.1	0.001
Heart SUVmax	5.4	4.3–6.7	2.5–11.0	8.1	7.4–8.7	5.8–14.7	< 0.001
SUVmax ratio	10.9	8.8–13.3	5.4–32.5	6.6	5.3–7.5	2.6–9.7	< 0.001

SUV standardized uptake value, *SUVmax* maximum SUV, *MLV* metabolic liver volume, *TL GSA* total lesion [99mTc]GSA uptake, *IQR* interquartile range

[a]The clearance index, HH15, was calculated as the heart ROI count at 15 min/heart ROI count at 3 min

[b]The receptor index, LHL15, was calculated as the liver region of interest (ROI) count at 15 min/(liver ROI count + heart ROI count at 15 min)

The extent of liver damage in each patient was determined during the pre-treatment period. One hundred patients presented with low liver damage, while 28 presented with high liver damage (grade B, $n = 26$ and grade C, $n = 2$). Patients with hepatitis B and hepatitis C were significantly less likely to be positive for high liver damage, compared to those with hepatitis nonB/nonC disease. No significant differences in age, sex and liver tumor histology were observed between patients with low and high liver damage. The patients' characteristics are summarized in Table 2.

Comparison of HH15, LHL15 and SUV-related parameters between low and high liver damage (Table 3)

As shown in Table 3, the patients with high liver damage had a significantly lower LHL15, liver SUVmax, liver SUVmean, TL GSA or SUVmax ratio and had a significantly higher HH15 or heart SUVmax, compared to those with low liver damage ($p < 0.05$ for all). No significant difference in the MLV was observed between the high and low liver damage groups ($p = 0.31$).

Diagnostic performances of each index for predicting high liver damage (Table 4)

As shown in Table 4, each index yielded the following AUC for the ability to predict high liver damage: HH15, 0.930 ($p < 0.001$); LHL15, 0.909 ($p < 0.001$); liver SUVmax, 0.650 ($p = 0.012$); liver SUVmean, 0.673 ($p = 0.004$); TL GSA, 0.713 ($p = 0.001$); heart SUVmax, 0.894 ($p < 0.001$) and SUVmax ratio, 0.917 ($p < 0.001$). The parameters yielded sensitivity rates ranging from 50.0% (liver SUVmean) to 100% (HH15), specificity rates ranging from 42.0% (liver

Table 4 The abilities of HH15, LHL15 and SUV-related parameters to predict high liver damage

Index	Cut off value	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p value
HH15[a]	> 0.62	100 (28/28) 87.7–100[c]	75.0 (75/100) 65.3–83.1[c]	80.4 (103/128) 72.5–86.9[c]	0.930 0.871–0.967[c]	< 0.001
LHL15[b]	< 0.88	85.7 (24/28) 67.3–96.0[c]	86.0 (86/100) 77.6–92.1[c]	85.9 (110/128) 78.7–91.4[c]	0.909 0.846–0.953[c]	< 0.001
Liver SUVmax	< 60.2	85.7 (24/28) 67.3–96.0[c]	42.0 (42/100) 32.3–52.3[c]	51.6 (66/128) 42.6–60.5[c]	0.650 0.561–0.732[c]	0.012
Liver SUVmean	< 32.9	50.0 (14/28) 30.6–69.4[c]	80.0 (80/100) 70.8–87.3[c]	73.4 (94/128) 64.9–80.9[c]	0.673 0.584–0.753[c]	0.004
MLV	< 652.4	28.6 (8/28) 13.2–48.7[c]	92.0 (92/100) 84.8–96.5[c]	78.1 (100/128) 70.0–84.9[c]	0.563 0.472–0.650[c]	0.36
TL GSA	< 29388.9	64.3 (18/28) 44.1–81.4[c]	72.0 (72/100) 62.1–80.5[c]	70.3 (90/128) 61.6–78.1[c]	0.713 0.627–0.790[c]	0.001
Heart SUVmax	> 7.4	82.1 (23/28) 63.1–93.9[c]	89.0 (89/100) 81.2–94.4[c]	87.5 (112/128) 80.5–92.7[c]	0.894 0.828–0.942[c]	< 0.001
SUVmax ratio	< 7.75	82.1 (23/28) 63.1–93.9[c]	87.0 (87/100) 78.8–92.9[c]	85.9 (110/128) 78.7–91.4[c]	0.917 0.855–0.959[c]	< 0.001

SUV standardized uptake value, *SUVmax* maximum SUV, *MLV* metabolic liver volume, *TL GSA* total lesion [99mTc]GSA uptake, *AUC* area under the ROC curve

[a]The clearance index, HH15, was calculated as the heart ROI count at 15 min/heart ROI count at 3 min

[b]The receptor index, LHL15, was calculated as the liver region of interest (ROI) count at 15 min/(liver ROI count + heart ROI count at 15 min)

[c]95% confidence interval

SUVmax) to 89% (heart SUVmax) and accuracy rates of 51.6% (SUVmax) to 87.5% (heart SUVmax) for predicting high liver damage. The representative 99mTc-GSA SPECT/CT images of low and high liver damages were shown in Figs. 1 and 2, respectively.

The AUCs of the HH15, LHL15, heart SUVmax and SUVmax ratio were significantly higher than those of the liver SUVmax, liver SUVmean and TL GSA ($p < 0.05$ for all). No significant differences in the AUCs was observed between other pairs of 7 parameters ($p \geq 0.05$ each).

Prediction of high liver damage using machine learning (Table 5)

On the model without feature selection, the 12 features including 4 clinical and 8 quantitative 99mTc-GSA features were used as the input data. The AUC of each classifier for predicting high liver damage in the absence of a PCA was 0.904 for SVM ($p < 0.001$) and 0.939 for RF ($p < 0.001$). No significant differences in the AUCs were observed between SVM and RF ($p \geq 0.05$ each). The sensitivity, specificity and accuracy were 92.9%, 73.0% and 77.3% for SVM, and 92.9%, 85.0% and 86.7% for RF, respectively. On the other hand, the AUC of each classifier for predicting high liver damage in the presence of a PCA was 0.839 for SVM ($p < 0.001$) and 0.896 for RF ($p < 0.001$). No significant differences in the AUCs were observed between SVM and RF ($p \geq 0.05$ each). The sensitivity, specificity and accuracy

were 78.6%, 84.0% and 82.8% for SVM, and 89.3%, 78.0% and 80.4% for RF, respectively.

On the model with feature selection, the following 5 best features which were ranked using the decrease in Gini impurity (Supplemental Table 1) were used as the input data; Heart SUVmax, SUVmax ratio, LHL15, HH15, and background liver disease. The AUC of each classifier for predicting high liver damage in the absence of a PCA was 0.942 for SVM ($p < 0.001$) and 0.933 for RF ($p < 0.001$). No significant differences in the AUCs were observed between SVM and RF ($p \geq 0.05$ each). The sensitivity, specificity and accuracy were 89.3%, 89.0% and 89.1% for SVM, and 92.9%, 85.0% and 86.7% for RF, respectively. On the other hand, the AUC of each classifier for predicting high liver damage in the presence of a PCA was 0.945 for SVM ($p < 0.001$) and 0.956 for RF ($p < 0.001$). No significant differences in the AUCs were observed between SVM and RF ($p \geq 0.05$ each). The sensitivity, specificity and accuracy were 85.7%, 91.0% and 89.8% for SVM, and 96.3%, 90.0% and 91.4% for RF, respectively.

On both models with or without feature selections, we did not observe significant differences in the AUCs between each classifier and four of the quantitative 99mTc-GSA parameters (HH15, LHL15, heart SUVmax and SUVmax ratio; $p \geq 0.05$). No significant differences were also observed either absence or presence of the PCA. However, the AUC of each classifier was significantly higher than that of the remaining three parameters (liver SUVmax, liver SUVmean and TL GSA; $p < 0.05$) on both model with or

Fig. 1 A 74-year-old male patient with hepatocellular carcinoma (HCC) and low liver damage. An early-phase enhanced computed tomography (CT) image (**a**) shows enhancement of the HCC. A ^{99m}Tc-GSA planar image (**b**) obtained 15 minutes after the injection of ^{99m}Tc-GSA shows hepatic uptake but no retention of ^{99m}Tc-GSA in the cardiac blood pool. A ^{99m}Tc-GSA SPECT/CT image (**c**) reveals homogeneous ^{99m}Tc-GSA uptake in the liver. The receptor index (LHL15), clearance index (HH15), liver maximum standardized uptake value (SUVmax), liver SUVmean, metabolic liver volume (MLV), total lesion ^{99m}Tc-GSA uptake (TL GSA), heart SUVmax and SUVmax ratio were 0.54, 0.95, 77.1, 46.76, 799.1 cm^3, 37364.5, 4.3 and 17.76, respectively. All parameters failed to meet the threshold for high liver damage, and the patient underwent liver resection to treat HCC
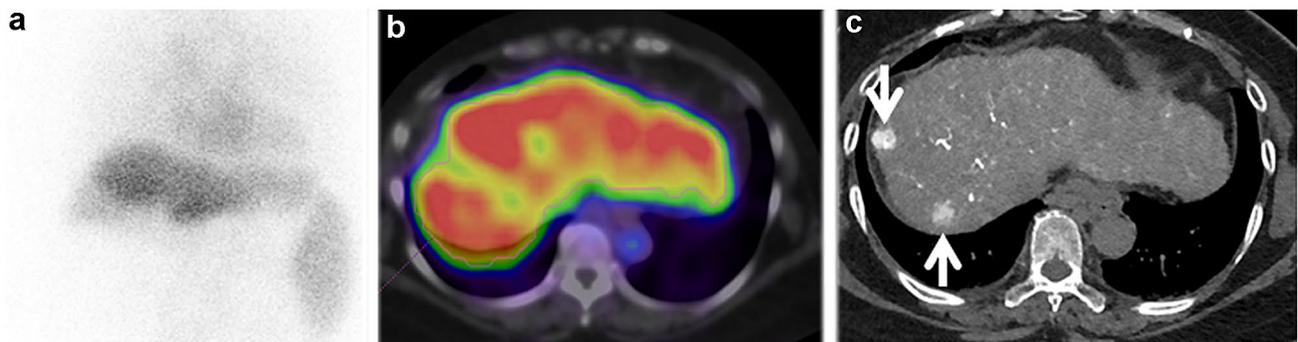
Fig. 2 A 58-year-old female patient with hepatocellular carcinoma (HCC) and high liver damage (Damage type C). A ^{99m}Tc-GSA planar image (**a**) obtained 15 minutes after the injection of ^{99m}Tc-GSA depicts the retention of ^{99m}Tc-GSA in the cardiac blood pool with hepatic and splenic uptake. The ^{99m}Tc-GSA single photon emission computed tomography (SPECT)/computed tomography (CT) image (**b**) shows heterogeneous ^{99m}Tc-GSA uptake in the liver. The receptor index (LHL15), clearance index (HH15), liver maximum standardized uptake value (SUVmax), liver SUVmean, metabolic liver volume (MLV), total lesion ^{99m}Tc-GSA uptake (TL GSA), heart SUVmax and SUVmax ratio were 0.72, 0.85, 55.2, 35.0, 567.9 cm^3, 19888.7, 14.05 and 3.93, respectively. All these parameters except for liver SUVmean met the respective thresholds for high liver damage. CT hepatic arteriography (**c**) shows the enhanced tumors, and the patient underwent transcatheter arterial chemoembolization (TACE) to treat HCC

without feature selections, and these significances were also observed either absence or presence of the PCA. The RF with 5 selected features and PCA model obtained the best performance for predicting high liver damage (AUC = 0.956, sensitivity = 96.3%, specificity = 90.0%, accuracy = 91.4%).

Discussion

Although the LHL15 and HH15 are the most common parameters used in liver function assessments, these indices are calculated from planar images and do not provide sufficient anatomical information to enable an accurate assessment [7]. SPECT/CT generates image voxels in units of radioactivity per volume (i.e., kBq/ml) based on the results of an algorithm of the CT-based attenuation correction and scatter correction [31]. This technique differs from previous nuclear imaging techniques such as planar scintigraphy, SPECT or non-quantitative SPECT/CT, which generate imaging units in counts per second. Quantitative SPECT/CT allows the normalization of lesion radioactivity with respect to the injected radioactivity and thus yields quantitative parameter values such as the SUV [32, 33]. In this context, we evaluated the utility of an SUV-related ^{99m}Tc-GSA SPECT/CT analysis of liver function. To our knowledge, only one other report has discussed the use of the SUV obtained during ^{99m}Tc-GSA SPECT/CT to evaluate liver

Table 5 Classification performance for predicting the high liver damage

Index	Without PCA					With PCA				
Model without feature selection	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p value	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p value
SVM	92.9 (26/28) 76.5–99.1[a]	73.0 (73/100) 63.2–81.4[a]	77.3 (99/128) 69.1–84.3[a]	0.904 0.839–0.949[a]	< 0.001	78.6 (22/28) 59.0–91.7[a]	84.0 (84/100) 75.3–90.6[a]	82.8 (106/128) 75.1–88.9[a]	0.839 0.764–0.898[a]	< 0.001
RF	92.9 (26/28) 76.5–99.1[a]	85.0 (85/100) 76.5–91.4[a]	86.7 (111/128) 79.6–92.1[a]	0.939 0.882–0.973[a]	< 0.001	89.3 (25/28) 71.8–97.7[a]	78.0 (78/100) 68.6–85.7[a]	80.4 (103/128) 72.5–86.9[a]	0.896 0.830–0.943[a]	< 0.001
Model with feature selection	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p value	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p value
SVM	89.3 (25/28) 71.8–97.7[a]	89.0 (89/100) 81.2–94.4[a]	89.1 (114/128) 82.3–93.9[a]	0.942 0.886–0.976[a]	< 0.001	85.7 (24/28) 67.3–96.0[a]	91.0 (91/100) 83.6–95.8[a]	89.8 (115/128) 83.3–94.5[a]	0.945 0.890–0.977[a]	< 0.001
RF	92.9 (26/28) 76.5–99.1[a]	85.0 (85/100) 76.5–91.4[a]	86.7 (111/128) 79.6–92.1[a]	0.933 0.875–0.970[a]	< 0.001	96.3 (27/28) 81.7–99.9[a]	90.0 (90/100) 82.4–95.1[a]	91.4 (117/128) 85.1–95.6[a]	0.956 0.905–0.984[a]	< 0.001

SVM support vector machine, *RF* random forest, *PCA* principal component analysis, *AUC* area under the ROC curve

[a]95% confidence interval

fibrosis [14], and the authors reported that the SUVmean was a highly accurate predictor of severe liver fibrosis.

In our study, patients with high liver damage had a significantly lower LHL15, liver SUVmax, liver SUVmean, TL GSA and SUVmax ratio and a significantly higher HH15 and heart SUVmax than patients with low liver damage. Three indices (HH15, LHL15 and SUVmax ratio) yielded a high AUC (>0.90), while the heart SUVmax most accurately differentiated between low and high liver damage. These results suggest that in addition to the previously reported quantitative parameters HH15 and LHL15, SUV-related parameters derived via ^{99m}Tc-GSA SPECT/CT can be used to evaluate liver function. One advantage of the SUVmax over previously reported quantitative analyses is the ease of calculation, as this value can be determined by simply placing a VOI within the liver or heart without introducing inter- or intra-observer variability.

Some recent studies have proposed classification methods based on machine learning approaches such as SVM and RF [17–19]. SVM is commonly considered superior for classification, as it provides optimal results based on existing information [34]. RF is a very effective model-free approach in terms of feature screening and classification [26]. Sakai et al. examined the usefulness of a machine learning approach by using the radiomics features of digital breast tomosynthesis to automatically classify benign and malignant breast lesions. In that study, the adoption of SVM as a classifier yielded the highest classification performance (AUC = 0.798) [35]. Ahn et al. used a machine learning approach to examine the prognostic value of FDG-PET-based radiomics in patients with non-small cell lung cancer; in that study, the RF model best predicted disease recurrence (AUC = 0.956) [18]. Rajula et al. [36] reviewed the difference between machine learning and conventional statistical method in healthcare, and reported that one main difference between machine learning and conventional statistical method lies in their purposes, and machine learning is focused on making predictions as accurate as possible, while conventional statistical models are aimed at inferring relationships between variables. They also reported that the benefits of machine learning comprise flexibility and scalability compared with conventional statistical approaches, which make it deployable for several tasks, such as diagnosis and classification. To our knowledge, however, no study has previously investigated the efficacy of a quantitative analysis of ^{99m}Tc-GSA SPECT/CT data via a machine learning approach for the assessment of liver function.

In our study, the adoption of RF with 5 selected features and PCA model yielded the best liver damage classification performance, although no significant differences were observed between the AUCs of each classifier and four quantitative ^{99m}Tc-GSA parameters (HH15, LHL15, heart SUVmax and SUVmax ratio; $p \geq 0.05$ for all). Even though these four quantitative ^{99m}Tc-GSA parameters showed high diagnostic performance for discriminating between low and high liver damage, it might be difficult to correctly classify low and high liver damage by using only one each parameter. Machine learning algorithms have the ability to analyze various data types (for instance, imaging data and laboratory findings) and integrate them into predictions for diagnosis [37], and our findings indicate that the machine learning approach with combining information on clinical and quantitative ^{99m}Tc-GSA features has the potential to correctly predict the liver function.

This study had some limitations that warrant discussion. First, this was a retrospective study with a relatively small sample, and case-selection bias was unavoidable. Our patient population included only a few patients with high liver damage because those with severe dysfunction do not usually receive ^{99m}Tc-GSA SPECT/CT evaluations, when therapeutic surgeries or liver tumor interventions are considered. Stratified sampling was used to overcome these imbalanced data, and cross-validation, feature selection and PCA were performed to minimize the negative influence of overfitting in this study. Although internal validation showed high diagnostic performance in RF with selected features and PCA model, the lack of external validation limits the generalizability of our result. Thus, a training/test scheme, which would require a large number of samples, would be preferable for the validation of classifiers, and a prospective study of a much larger population with external validation test is needed to validate and confirm our findings. Second, we did not examine the relationship between the post-therapeutic liver function and pretreatment ^{99m}Tc-GSA SPECT/CT parameters due to differences in the treatment procedures. A follow-up study is needed to assess the predictive value of the machine learning approach based on ^{99m}Tc-GSA SPECT/CT with respect to clinical outcomes.

In conclusion, a machine learning approach based on clinical and quantitative ^{99m}Tc-GSA SPECT/CT parameters might be useful for predicting liver function.

Compliance with ethical standards

References

1. Hoekstra LT, de Graaf W, Nibourg GA, et al (2013) Physiological and biochemical basis of clinical liver function tests: a review. Ann Surg 257:27-36

2. Ashwell G, Morell AG (1974) The role of surface carbohydrates in the hepatic recognition and transport of circulating glycoproteins. Adv Enzymol Relat Areas Mol Biol 41: 99-128

3. Sobue G, Kosaka A (1980) Asialoglycoproteinemia in a case of primary hepatic cancer. Hepatogastroenterology 27:200-203.

4. Marshall JS, Green AM, Pensky J, Williams S, Zinn A, Carlson DM (1974) Measurement of circulating desialylated glycoproteins and correlation with hepatocellular damage. J Clin Invest 54:555-562.

5. Sawamura T, Nakada H, Hazama H, Shiozaki Y, Sameshima Y, Tashiro Y (1984) Hyperasialoglycoproteinemia in patients with chronic liver diseases and/or liver cell carcinoma. Asialoglycoprotein receptor in cirrhosis and liver cell carcinoma. Gastroenterology 87:1217-1221

6. Okabayashi T, Shima Y, Morita S, et al (2017) Liver function assessment using technetium 99m-Galactosyl single-photon emission computed tomography/CT fusion imaging: a prospective trial. J Am Coll Surg 225:789-797

7. Yoshida M, Beppu T, Shiraishi S, et al (2015) (99m)Tc-GSA SPECT/CT fused images for assessment of hepatic function and hepatectomy planning. Ann Transl Med 3:17

8. Sasaki N, Shiomi S, Iwata Y, et al (1999) Clinical usefulness of scintigraphy with 99mTc-galactosyl-human serum albumin for prognosis of cirrhosis of the liver. J Nucl Med 40:1652-1656

9. Shiomi S, Sasaki N, Tamori A, et al (1999) Use of scintigraphy with 99mtechnetium galactosyl human serum albumin for staging of primary biliary cirrhosis and assessment of prognosis. J Gastroenterol Hepatol 14:566-571

10. Buck AK, Nekolla S, Ziegler S, et al (2008) SPECT/CT. J Nucl Med 49:1305-1319.

11. Kaneta T, Ogawa M, Daisaki H, Nawata S, Yoshida K, Inoue T (2016) SUV measurement of normal vertebrae using SPECT/CT with Tc-99 m methylene diphosphonate. Am J Nucl Med Mol Imaging 6:262-268

12. Suh MS, Lee WW, Kim YK, Yun PY, Kim SE (2016) Maximum standardized uptake value of (99 m)Tc hydroxymethylene diphosphonate SPECT/CT for the evaluation of temporomandibular joint disorder. Radiology 280:890-896

13. Beck M, Sanders JC, Ritt P, Reinfelder J, Kuwert T (2016) Longitudinal analysis of bone metabolism using SPECT/CT and (99 m) Tc-diphosphono-propanedicarboxylic acid: comparison of visual and quantitative analysis. EJNMMI Res 6:60

14. Tokorodani R, Sumiyoshi T, Okabayashi T, et al (2019) Liver fibrosis assessment using 99 mTc-GSA SPECT/CT fusion imaging. Jpn J Radiol 37:315-320

15. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine learning for medical imaging. Radiographics 37:505–515

16. Waljee AK, Higgins PD (2010) Machine learning in medicine: a primer for physicians. Am J Gastroenterol 105:1224–1246

17. Gao X, Chu C, Li Y, et al (2015) The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. Eur J Radiol 84:312-317

18. Ypsilantis PP, Siddique M, Sohn HM, et al (2015) Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. PLoS One 10:e0137036

19. Ahn HK, Lee H, Kim SG, Hyun SH (2019) Pre-treatment ^{18}F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. Clin Radiol 74:467-473

20. Nakajima K, Kinuya K, Mizutani Y, et al (1999) Simple scintigraphic parameters with Tc-99m galactosyl human serum albumin for clinical staging of chronic hepatocellular dysfunction. Ann Nucl Med 13:5-11

21. Ikai I, Arii S, Kojiro M, et al (2004) Reevaluation of prognostic factors for survival after liver resection in patients with hepatocellular carcinoma in a Japanese nationwide survey. Cancer 101:796-802

22. Sian H, Purnami SW (2015) Combine sampling support vector machine for imbalanced data classification. Procedia Computer Science 72:59-66

23. Xie Y, Jiang B, Gong E, et al (2019) Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. AJR Am J Roentgenol 212:44-51

24. Jung Y (2018) Multiple predicting K-fold cross-validation for model selection. J Nonparametr Stat 30:197-215

25. Cook JA, Ranstam J (2016) Overfitting. Br J Surg 103:1814

26. Breiman L (2001) Random forests. Mach Learn 45:5-32

27. Rahman R, Kodesh A, Levine SZ, Sandin S, Reichenberg A, Schlessinger A (2020). Identification of newborns at risk for autism using electronic medical records and machine learning. Eur Psychiatry 63:e22

28. Hotta M, Minamimoto R, Miwa K (2019) 11C-methionine-PET for differentiating recurrent brain tumor from radiation necrosis: radiomics approach with random forest classifier. Sci Rep 9:15666.

29. Demsar J, Curk T, Erjavec A, et al (2013) Orange: data mining toolbox in Python. J Machine Learn Res 14:2349-2353

30. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837-845

31. Kitajima K, Futani H, Fujiwara M, et al (2018) Usefulness of quantitative bone single photon emission computed tomography/computed tomography for evaluating response to neoadjuvant chemotherapy in a patient with periosteal osteosarcoma. Cureus 10:e3655

32. Zeintl J, Vija AH, Yahil A, Hornegger J, Kuwert T (2010) Quantitative accuracy of clinical 99 mTc SPECT/CT using ordered-subset expectation maximization with 3-dimensional resolution recovery, attenuation, and scatter correction. J Nucl Med 51:921-928

33. Bailey DL, Willowson KP (2013) An evidence-based review of quantitative SPECT imaging and potential clinical applications. J Nucl Med 54:83-89

34. Asl BM, Setarehdan SK, Mohebbi M (2008) Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artif Intell Med 44:51-64

35. Sakai A, Onishi Y, Matsui M, et al (2020) A method for the automated classification of benign and malignant masses on digital breast tomosynthesis images using machine learning and radiomic features. Radiol Phys Technol 13:27-36

36. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V (2020) Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas) 56:455

37. Ngiam, KY Khor, IW (2019) Big data and machine learning algorithms for health-care delivery. Lancet Oncol 20: e262–e273