Comparison of Machine Learning-Based Radiomics Models for Early Recurrence Prediction of Hepatocellular Carcinoma

Panyanat Aonpong[1], Qingqin Chen[2], Yutaro Iwamoto[1], Lanfen Lin[3], Hongjie Hu[2], Qiaowei Zhang[2], and Yen-Wei Chen[1,4,3]
[1] College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan
[2] Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China
[3] College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[4] Zhejiang Lab, Hangzhou, China
Email: {llf, hongjiehu}@zju.edu.cn, chen@is.ritsumei.ac.jp

***Abstract*—BACKGROUND & AIMS: Using a radiomics model, we investigated computed tomography images to make a preoperative prediction of the Early Recurrence (ER) of Hepatocellular Carcinoma (HCC). A radiomics model mainly consists of feature extraction, feature selection, and classification. The conventional method used Least Absolute Shrinkage and Selection Operator (LASSO) regression to select the features and the classification. METHODS: We compared the new combination of feature selection and classification methods for the preoperative ER prediction of HCC. The new combination gave a significantly higher accuracy than the conventional method. Twelve combination models were provided by using different combinations of the feature selection methods and the classification methods. We used three classification methods, which included LASSO, linear Support Vector Machines (SVMs), and decision trees. We compared the performance of each method by using the area under the curve of the receiver operating characteristic to show a more appropriate way to detect the ER of HCC. In addition, we compared the efficiency term of each feature elimination method. These two comparisons can measure the quality of feature selection and the compatibility between both the feature selection and classification phases. RESULTS: Approximately 65 features were selected from 300 features. Our proposed combination showed that the accuracy could be improved by using the SVM classification method for new radiomics models. The accuracy of feature selection using LASSO regression with Support Vector Machine classification can reach 0.8918, whereas this accuracy reaches 0.8779 when support vector machine-recursive feature elimination is used with LASSO classification.**

***Index Terms*—hepatocellular carcinoma, early recurrence, preoperative prediction, liver cancers, Radiomics, feature extraction, feature selection, feature elimination, classification, machine learning**

I. Introduction

Hepatocellular Carcinoma (HCC) is a kind of cancer that occurs in the liver. It is the third leading cause of death from all types of cancers. HCC is commonly found in Asia and Africa, where there is a large outbreak of hepatitis B and C. This viral hepatitis plays a major role in causing chronic liver diseases and eventually leads to liver cell cancer [1]-[5].

In their very early stages, tumors are potentially curable using procedures such as surgical resection, liver transplantation, partial hepatectomy, and radiofrequency ablation. Even in the later stages of HCC, such as the intermediate and advanced stages, long-term survival benefits after resection have been reported. Cancer can recur even after a surgery. This recurrence is a main cause of death for HCC patients. In addition, 60%–80% patients die within five years of recurrence. The period from resection to recurrence is a factor for providing a prognosis and predicting the survival time of patients. HCC that is prognosed at the beginning of a period (less than 1 year) is usually worse than the HCC prognosed later. Even in patients with a small HCC tumor (less than 3cm), Early Recurrence (ER) is not a rare symptom after surgery. Therefore, HCC patients must be well stratified by clinicians. Alternative treatment strategies and preoperative treatments should be considered for patients who are at a high risk of ER. To ensure sustainable surgical management, patients need to be closely monitored and need to be provided post-operative neo-adjuvant treatment. The pathological characteristics associated with ER can be examined by a biopsy; however, a biopsy is not usually recommended because of the risk of cancer cells spreading and the inconsistency between the biopsy findings with the final pathology report [1], [2]. In this paper, we use Computed Tomography (CT)-based imaging for preoperative ER prediction of HCC to better stratify HCC patients with surgical therapies and adjunctive treatments. The use of CT-based imaging helps to obtain a higher degree of accuracy, which also reflects the underlying heterogeneity to be observed.

In [1], the researchers proposed a radiomics method that could help doctors diagnose ER and Non-early Recurrence (NER) of HCC. The radiomics method

Manuscript received July 1, 2019; revised November 7, 2019.

doi: 10.18178/joig.7.4.117-125

focuses on extracting a large number of quantitative properties from medical images by using data classification algorithms [1], [6]. This method can reveal information about the images that are not visible or are difficult to view with the naked eyes. The conventional radiomics methods have high accuracy classification. A radiomics algorithm can be developed in many possible ways.

To employ the radiomics method, we extracted features from the source images. The extracting methods vary depending on the functions. In this study, we used a histogram and the Gray-Level Co-occurrence Matrix (GLCM) as the main techniques. Next, we qualified the quality and importance of each feature and removed the features that were not so important to reduce the redundancy and complexity in the classification phase [1], [6].

In this study, we overcome the challenges of finding the best combination between the various feature elimination methods, including the Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine-Recursive Feature Elimination (SVM-RFE), and the chi-squared test, by using different classification methods, including LASSO, linear Support Vector Machine (SVM), and Decision Trees (DTs). This comparison measures the accuracy term and the efficiency term. We used the Area under the Curve (AUC) to measure the performance in the discrimination of ER and NER groups and to ensure that the accuracy was not obtained randomly.

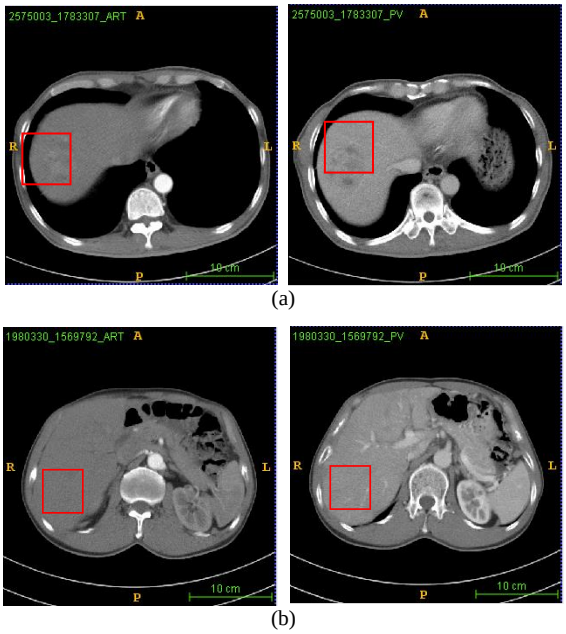
II. Materials

From 2012 to 2016, we collected data from 331 HCC patients at the Sir Run Run Shaw Hospital, Zhejiang University, China. Based on our inclusion criteria, we selected those patients who had (1) a local surgical resection, (2) a negative surgical margin, (3) a follow-up history for at least one year, and (4) CT images before their surgeries. Using these criteria, we finally obtained data on 167 patients. Among these, 65 patients had ER after surgery, and 102 patients were NER cases.

A. Multi-Phase CT Images

Contrast-enhanced CT is the most important imaging modality employed to detect and characterize focal liver lesions. Contrast-enhanced CT scans are divided into four phases before and after the injection of a contrast. A Non-contrast Enhanced (NC) scan is performed before contrast injection. The after-injection phases included the arterial (ART) phase (30–40 s after contrast injection), the PORTAL venous (PV) phase (70-80 s after contrast injection), and the Delay (DL) phase (3-5 min after contrast injection). In our research, we focus on the NC, ART and PV phases, which were scanned by GE LightSpeed VCT and Siemens SOMATON Definition AS. The resolution of the CT images was 512×512 pixels, and the thickness of each slice was 5-7 mm.

The bounding box of the tumor was manually outlined by a radiologist having three years of experience and was checked and revised by a radiologist having six years of experience. The labeling and modification was done using the ITK-SNAP application. After we obtained the region of interest, we performed the registration according to their center points. Raw slices of the CT images in ART (left) and PV (right) phases are shown in Fig. 1. HCC tumors are indicated by red bounding boxes.



(a) (b)

Figure 1. Examples of datasets. For (a) ER and (b) NER. The image on the left is the ART phase, and the image on the right is the PV phase

III. Methods

We used the following methods in this paper.

A. Radiomics Feature Extraction

From the raw dataset, we extracted 300 radiomics features from one CT-image using GLCM and a histogram [1]. The extracted radiomics features consisted of 150 radiomics features from the arterial phase and 150 radiomics features from the PV phase. To focus on the arterial phase, we used the Laplace of Gaussian (LoG) to generate five different images using five different Gaussian filters (σ= 0, 1, 1.5, 2, 2.5).

The LoG is a combination between the Laplacian filter and the Gaussian filter. The Gaussian filter smooths an image to reduce the noise. The Laplacian filter highlights the regions of rapid intensity change, or they work similar to edge detection [6], [7]. For the Gaussian filter, the Gaussian function is calculated as follows:

$$G(x, y; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

where x and y are the coordinates on the x and y axes, respectively, and σ is the standard derivation or the biometrics [6] filter parameter [7]. The Gaussian scale-space representation of the image $f(x, y)$ can be computed as follows:

$$L(x, y; \sigma) = f(x, y) * G(x, y; \sigma) \quad (2)$$

where $L(x, y; \sigma)$ is the Gaussian scale space of the image $f(x, y)$, and $*$ is a convolutional operator [7]. In Eq. (1) and Eq. (2), we apply the Laplacian operator ∇^2 to the Gaussian scale–space representation of the image. We can calculate ∇^2 as follows:

$$\nabla^2 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3}$$

Then, the LoG can be written as follows [7]:

$$\nabla^2 G(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{\pi\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{4}$$

The typical filtered images with the LoG are shown in Fig. 2.

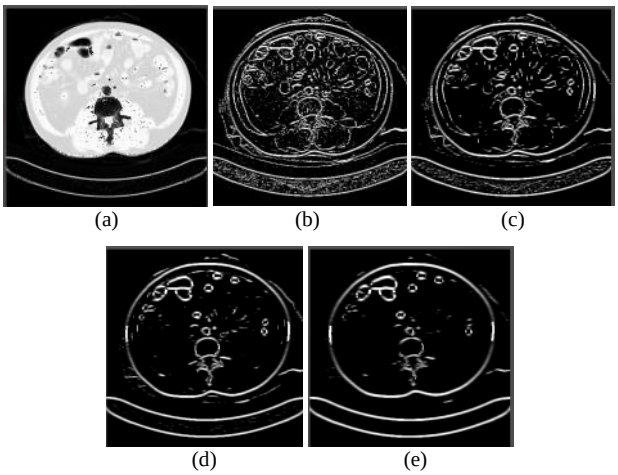


Figure 2. Laplace of Gaussian in different filters. (a) σ=0, (b) σ=1, (c) σ=1.5, (d) σ=2, (e) σ=2.5

We extracted a four-way GLCM [8] from each LoG filter; the four-way GLCM involves angles at 0 °, 45 °, 90 °, and 135 °. Then, we calculated the following features for every GLCM result:

- Contrast
- Entropy
- Correlation
- Homogeneity
- Energy

We used the original histogram for feature extraction and determined the following:

- Percentile S.D. (10, 25, 50 top percentiles)
- Percentile mean (10, 25, 50 top percentiles)
- Mean
- Standard Derivation (S.D.)
- Kurtosis
- Skewness

Finally, we obtained 150 radiomics features in the ART phase. Next, we performed the same procedure with the PV phase to obtain another 150 radiomics features. Finally, from one raw medical image, we obtained 300 radiomics features.

B. Radiomics Feature Selection

In this sub-section, we describe the three feature selection methods used in this study to eliminate unimportant radiomics features from the data pool.

After extracting all 300 features, we eliminated the features that did not affect the classification phase; this helped to reduce the data complexity and redundancy. We used the following methods for this elimination process:

1) Least absolute shrinkage and selection operator

LASSO is a method of linear regression in which data values are shrunk toward a central point; it is used in the fields of statistics and machine learning [6]. LASSO minimizes the slope of the linear function with the lowest residual calculated from the inspected data. The LASSO regression model works based on the following Lagrangian form:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{N} \left(\left(\sum_{i=1}^{N} y_i - \sum_{j=1}^{M} \beta_j X_{ij} \right) \right)^2 - \lambda \sum_{j=1}^{M} |\beta_j| \right\} \tag{5}$$

here, N is the number of samples, and M is the number of features. X_{ij} is the j-th radiomics feature at the i-th sample; y_i is the outcome (with the class label 1 for ER and 0 for NER) of the i-th sample; β_j is the slope of the graph or the coefficient at the j-th feature, and λ is the effectiveness required to balance the first term and the second term [6]. The second term is a sparse term. To solve Eq. (1), we first determine the best λ by using the AUC of the Receiver Operating Characteristic (ROC). Then, we determine the optimum value of the coefficient (β_j) by minimizing Eq. (5). Because of the sparse second term, only the important features that strongly correlate with the class label (y) have non-zero coefficients; other features having zero coefficients are considered to be unimportant features (i.e., they do not have any relationship with the class label (y)). In this paper, the features with the non-zero coefficients (β) are selected as important features for the prediction (i.e., calculating the score or class label) of y. This will be further described in this section under Classification Methods (see C.1).

2) Support vector machine-recursive feature elimination

SVM-RFE is the next generation SVM, which we explained in the Classification Methods section of [9]-[12]. SVM-RFE eliminates insignificant features based on the SVM classification and its accuracy. SVM-RFE can only function with linear SVMs because of its simplicity. SVM-RFE has the following work sequence [13]-[15].

- Estimate the weight ($\mathbf{w}$) of the linear SVM using the loss function (Eq. (6)). Our dataset had high complexity; therefore, we needed to use a soft-margin equation. To do this, we minimized as follows:

$$\min_{\mathbf{w},\mathbf{b}} \left\{ \left[\frac{1}{N} \sum_{i=1}^{N} \left(\max\ 0, 1 - \left(y_i \sum_{j=1}^{M} w_j \cdot x_{ij} - b \right) \right) \right] + \lambda \sum_{j=1}^{M} {w_j}^2 \right\} \tag{6}$$

where N is the number of samples; M is the number of features; w is the slope of the linear graph or the weight; x_{ij} is the j-th radiomics feature at the i-th sample; λ is the trade-off

between the margin size, and b is the bias or the y-axis intersection [9]-[12]

- Calculate the ranking criteria based on the SVM weight as follows:

$$C_j = w_j^2 \tag{7}$$

where C_j represents the criteria based on the SVM weight of the j-th feature [13].

- Eliminate the lowest criteria rank feature (i.e., the lowest C).
- Repeat the steps 1–3 until the required number of features are obtained.

3) Chi-Squared test

The chi-squared test is used to determine the dependency between two variables [15]-[17]. In our case, we used the chi-squared test to check the dependency between the same features from every dataset and the expected ground truth. The chi-squared test can measure the dependency index of the j-th feature in the form of $\chi(j)^2$. The features have high complexity; therefore, we need to normalize all the data from each feature to give it the same range value by using the following equation:

$$O'_{ij} = \frac{O_{ij} - min(\mathbf{O}_j)}{max(\mathbf{O}_j) - min(\mathbf{O}_j)} \tag{8}$$

here, O_{ij} and O'_{ij} denote the j-th radiomics feature observed at the i-th data before and after normalization, and $\mathbf{O}j$ denotes all the data in the j-th radiomics feature. We can write the chi-squared test as follows:

$$\chi(j)^2 = \sum_{i=1}^{N} \left(\frac{O'_{ij} - E_i}{E_i}\right)^2 \tag{9}$$

where E_i denotes the expected value or the target class [16], [17], that is, the expected ground truth of the i-th data. Using the definition of the chi-squared test, we can easily apply the chi-squared technique in the radiomics feature selection by calculating the chi-squared statistics using Eq. (9) for the radiomics features and the target classes ER and NER. Then, we can observe the existence of a relationship between the variables and the target. If the target class is related to the radiomics feature, it will show a low χ^2 value. From this property, we can rank the relationship between the radiomics feature and the target class. Then, we discard a number of high-ranking radiomics features, that is, the features have a low relationship with the target class.

C. Classification Methods

After feature selection, the selected features are used for machine learning–based classification or prediction. We used three methods as classifiers or predictors.

1) Least absolute shrinkage and selection operator

In this paper, we used LASSO regression for identifying the importance of radiomics features, and we used it as one of the predictions (or classification) methods by using the following equation:

$$y_{score} = \boldsymbol{X} \cdot \boldsymbol{\beta} \tag{10}$$

here, y_{score} represents a radiomics score (i.e., an estimation of y for a new test sample) based on linear regression with the selected features ($\mathbf{X}$) and their coefficients ($\boldsymbol{\beta}$).

LASSO can give us the radiomics score of all the data. We could apply this property with the threshold to allow it to make a binary classification using the following equation:

$$y_{test} = \begin{cases} 1, if\ y_{score} > 0.5 \\ 0, if\ y_{score} \leq 0.5 \end{cases} \tag{11}$$

here, y_{test} is the result of the test sample, and y_{score} is obtained score.

2) Support Vector Machine (SVM)

The SVM model is a kind of supervised machine learning. It is one of the most useful classification techniques [9]-[11]. The SVM model is fast, accurate, and flexible for simple studies. The SVM model represents the widest area possible on the graph for separating groups of classes from one another [9]. In this paper, we use linear SVM because it is simple and has high accuracy; in particular, it can be used in the SVM-RFE method.

The SVM model consists of three main lines in a two-dimension graph, including the optimal hyperplane, which is the middle line that separates the groups of features from one another. In the linear SVM, the hyperplane is represented as follows:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{12}$$

where $\boldsymbol{w}$ is the weight vector; $\boldsymbol{x}$ is the input vector; and b is the bias or the y-axis intersection [9]-[12].

The two lines of the support vector touch one or two data points. These two lines need to maximize the length of last line (i.e., the margin), which indicates the distance between the two support vector lines. In the linear SVM, the two support vector lines are linear lines in the form of the following linear equations:

$$\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + b = +1 \tag{13}$$

$$\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + b = -1 \tag{14}$$

Therefore, when d_i is the margin of separation or the expected class from the prediction [9]-[12], we can state the following:

$$\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + b \geq 0 \text{ for } d_i = +1 \tag{15}$$

$$\mathbf{w}^{\mathbf{T}} \cdot \mathbf{x} + b < 0 \text{ for } d_i = -1 \tag{16}$$

In our case, $+1$ and -1 refer to the ER and NER groups of the HCC patients.

3) Decision Tree [18]

DT is a hierarchical data structure. It looks like an upside-down tree that has root nodes at the top and leaves at the bottom. The DT contains many nodes. Each node is responsible for testing. The branches represent the possible choices, and the leaves at the bottom of the tree represent the results of the prediction. DT is one of the simplest algorithms used for classification, and it

does not need a high-performance computer. To train a DT, the DT node is split using an algorithm. In this study, we used one of the most popular node-splitting algorithms named ID3 algorithm [17]-[19].

The ID3 algorithm [19] needs information gain calculation to make the decision to split a node. This study used binary classification; therefore, we can focus on p (positive) and n (negative) classes at the inspected node. We can calculate the entropy value as follows:

$$I(positive, negative) = -\frac{N_{positive}}{N_{positive}+N_{negative}}\log_2\frac{N_{positive}}{N_{positive}+N_{negative}} - \frac{N_{negative}}{N_{positive}+N_{negative}}\log_2\frac{N_{negative}}{N_{positive}+N_{negative}} \quad (17)$$

here, $I(positive, negative)$ denotes the entropy between the positive and negative class; $N_{positive}$ denotes the number of positive data; and $N_{negative}$ denotes the number of negative data at the inspected node [19]. The information gain value allows us to calculate the average entropy and to measure the randomness as follows:

$$E(A_s) = \sum_{k=1}^{G_s}\frac{N_k}{N} I(N_{positive_k}, N_{negative_k}) \quad (18)$$

where G_s is the group of data we want to separate using the A_s attribute. N_{p_k} and N_{n_k} denote the number of positive and negative elements, respectively, in the group k [18]. Now, we can calculate the information gain from the average entropy as follows:

$$gain(A_s) = I(positive, negative) - E(A_s) \quad (19)$$

We can first separate the attribute into different groups by selecting the attribute that has the highest information gain value. Then, we change the inspected node to the child node and follow this method again until no more attribute remains to make a decision [17], [18].

D. Prediction Models

In this section, we explain the models used in the experiment. We used LASSO regression, SVM-RFE and CHI-2 to measure the importance of each feature. LASSO regression will automatically select and assign the weight for each feature but SVM-RFE and CHI-2 will not do this because almost all the weights from SVM-RFE and CHI-2 are non-zero. We need to assign the number of features that we want to use for each method. To solve this problem, we employed the accuracy of the classification method used each time a feature was eliminated. We collected all the accuracy results and used the best result as the most suitable number of features.

Then, we used LASSO, SVM, and DT to classify the weighed radiomics features for each feature selection method. The algorithm is shown in the block diagram in Fig. 3.

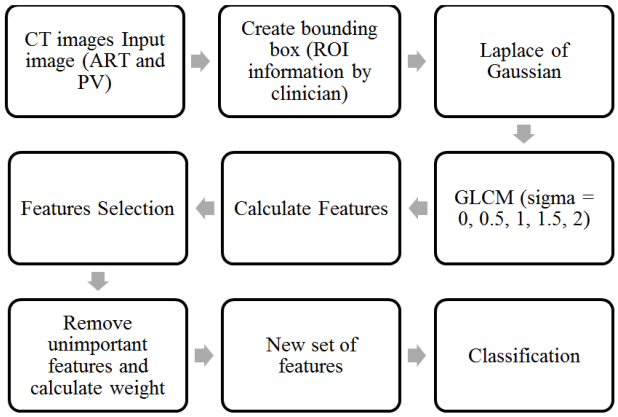


Figure 3. Block diagram of proposed algorithm

In each radiomics classification model, we made many combination models between the feature selection methods and the classification methods. All the combined models are shown in Table I. They are divided into four groups: Group A (without feature selection), Group B (with LASSO feature selection), Group C (with SVM-RFE feature selection); and Group D (with the chi-squared test feature selection).

TABLE I. ALL MODELS COMBINED BETWEEN FEATURE SELECTION METHODS AND CLASSIFICATION METHODS IN THE EXPERIMENT

Model	Feature Selection methods			Classification methods		
	LASSO	SVM-RFE	CHI-2	LASSO	SVM	DT
A1				•		
A2					•	
A3						•
B1	•			•		
B2	•				•	
B3	•					•
C1		•		•		
C2		•			•	
C3		•				•
D1			•	•		
D2			•		•	
D3			•			•

E. Evaluation

In this paper, we use AUC of ROC to evaluate the accuracy of each method.

The ROC curve is a graph that shows the diagnostic capabilities of the binary classifier system. It shows the relationship between the sensitivity (true positive rate) and 1-specificity (1-false positive rate) [6]. When the classification method works very well, the ROC graph will shift to the top, that is, the mean AUC of the ROC will increase. Therefore, we can conclude that the method having more AUC should have more accurate classification performance [20].

IV. Results and Discussion

To evaluate the selection quality of the appropriate features, we can estimate the accuracy of the algorithms from the analysis used in the classifications. In this article, we use LASSO regression, linear SVM, and DT because they have high performance with the radiomics method; they do not take much time to process and have high precision. Furthermore, because they are simple, it is easy to modify the parameters.

A. Feature Selection Results

The chi-squared test by itself could not determine the quality of feature elimination; therefore, we used the linear SVM to determine the quality of the iterations. We used the SVM to test the quality of the remaining features in the training set each time the feature was cut out using K-fold validation with the data in the training set. Then, we collected all the accuracy values to check and select the number of features that provided suitable accuracy values. In this experiment, we normalized all the accuracy values as integers in the range from 0 to 100; for example, if we had an accuracy factor of 0.8283 from the K-fold validation, it was normalized to 82. Then, we chose the highest accuracy, which had the smallest number of features. We found that each method had a different ability to cut the features. Cutting efficiency had been used to indicate the cutting ability of each method, which calculated by the equation (20).

$$C.E. = \frac{N_{features_{all}} - N_{features_{remain}}}{N_{features_{all}}} \times 100 \qquad (20)$$

where $C.E.$ is cutting efficiency, $N_{features_{all}}$ is the number of all features, in this study $N_{features_{all}} = 300$ and $N_{features_{remain}}$ is the number of remaining features from feature selection methods that we observing. The cutting efficiency of each feature selection method is shown in Fig. 4.

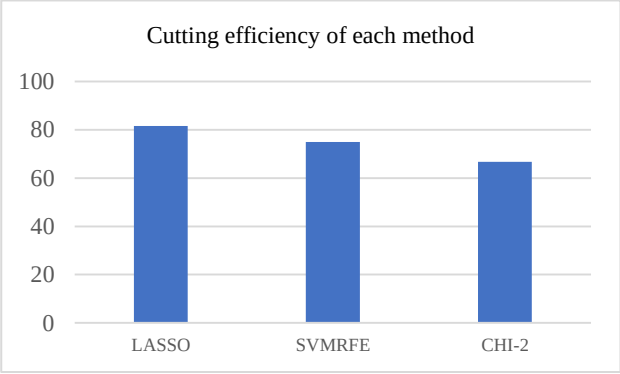


Figure 4. Cutting efficiency comparison between all the feature selection methods

From the experiment, we found that LASSO had the best cutting efficiency (average 81.56%), followed by SVM-RFE (average 75% for the best determining results); the worst was Chi-2 (average 66.67% for the best determining results). Some of the highest probability features that all three methods had are summarized in Table II.

TABLE II. Example of the Highest Probability Features That All Three Method Had Usually Called

Coef#	Phase	LoG filter	GLCM Degree	Computation
8	ART	0	135	Correlation
33	ART	0	1	Energy
52	ART	1.5	135	Entropy
53	ART	1.5	0	Energy
60	ART	1.5	135	Homogeneity
61	ART	2	0	Contrast
79	ART	2	90	Homogeneity
82	ART	2.5	45	Contrast
99	ART	2.5	90	Homogeneity
101	ART	-	0	mean
122	ART	-	1.5	std
139	ART	2	-	std25
140	ART	2	-	std50
141	ART	2.5	-	mean
151	PV	0	0	Contrast
193	PV	1.5	90	Contrast
195	PV	1.5	0	Correlation
201	PV	1.5	90	Entropy
204	PV	1.5	45	Energy
210	PV	1.5	135	Homogeneity
212	PV	2	45	Contrast
224	PV	2	45	Energy
249	PV	2.5	90	Homogeneity
271	PV	1.5	-	mean
272	PV	1.5	-	std
289	PV	2	-	std25

B. Prediction Results

We compare the accuracy of the analysis before and after selecting the radiomics features. In addition, we also compared the different methods used to show the differences in the properties of the radiomics methods. Every value was measured in the AUC of the ROC to avoid correction from randomness. The results of the five-fold validation that was averaged 10 times are shown as Fig. 5.

From Fig. 5, the model names on the *x*-axis refer to the feature selection methods and the classification methods (see Table I). The *y*-axis shows the measured precision in the AUC of the ROC form. Fig. 5 shows that the SVM classification method works very well with the feature selection using LASSO regression on the B2 model. This model can reach an accuracy of 0.6704. In the C1 model, the LASSO regression classification method with feature selection using SVM-RFE can reach an accuracy of 0.6623. This value is higher than the conventional method that uses LASSO for both feature selection and classification, which can reach a maximum accuracy of 0.6085.

From Fig. 5, it is clear that radiomics feature extraction using LASSO can significantly boost the performance of SVM. Using LASSO regression, $\boldsymbol{\beta}$ was calculated, and it was assigned to the weight of each radiomics feature. The $\boldsymbol{\beta}$ values varied depending on the relationship between the radiomics feature and the expected target class. After using SVM for classification, we were able to easily analyze these expanded features.

In the process of feature elimination, SVM-RFE will eliminate the radiomics features and simultaneously assign the weight for each feature using the criteria (C_i). When using LASSO regression for classification, LASSO uses Eq. (5) to calculate the radiomics score of the inspected sample. The multiplier ($\boldsymbol{\beta}$) which was trained before behaves like other screening layers on its own. The features that have low relationship indexes will be rejected.

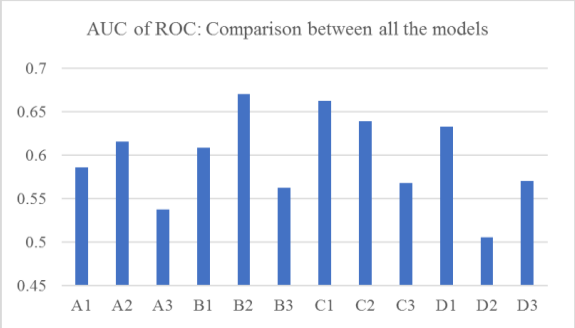
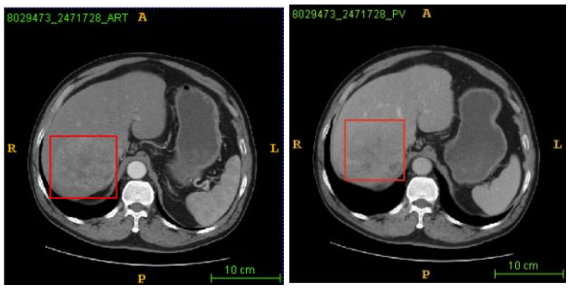


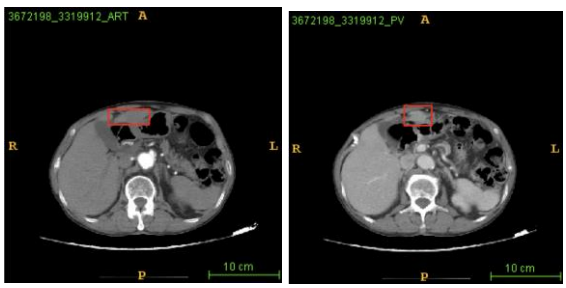
Figure 5. Comparison of AUC between all the models

In our experimental results, we found that some CT-images were correctly predicted by all 12 radiomics models. On the other hand, some CT-images were mis-predicted by more than 8 models. Typical images for both cases are shown in Fig. 6-Fig. 9.



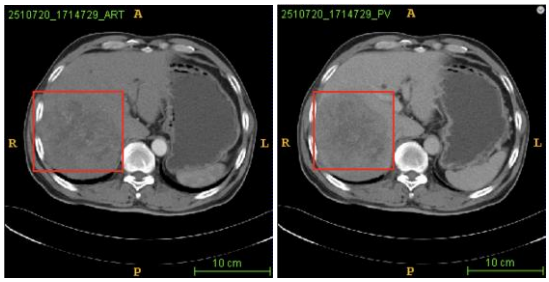
(a) (b)

Figure 6. An example of NER, which was mis-predicted as ER by more than eight radiomics medels. (a) ART (b) PV



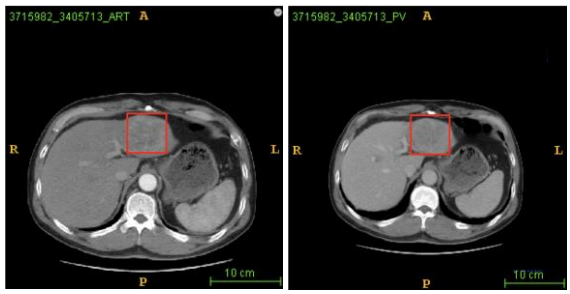
(a) (b)

Figure 7. An example of ER, which was mis-predicted as NER by more than eight radiomics models. (a) ART (b) PV



(a) (b)

Figure 8. An example of ER, which was correctly predicted by all 12 radiomics models. (a) ART (b) PV



(a) (b)

Figure 9. An example of NER, which was correctly predicted by all 12 radiomics models. (a) ART (b) PV

V. Conclusion

Our experiments prove that using LASSO regression for radiomics feature selection together with SVM for classification provides significantly higher ER for HCC classification as compared with conventional radiomics methods that use only LASSO regression for both phases. Combining both LASSO and SVM reduces each other's weaknesses making it possible to extract the highest efficiency for each algorithm. This leads us to the next generation of ER for the HCC detection method using the new combination. Furthermore, this highest combined method can also reduce the number of radiomics features by up to 81%.

The methods used in this experiment are simple and easy to study. Therefore, there is a possibility of developing a new radiomics method by using the structure to increase the work efficiency and accuracy and to reduce the number of features.

Acknowledgment

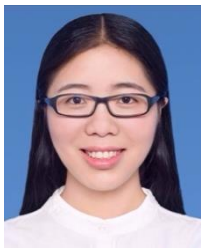
This work was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 18H03267 and No.17H00754, in part by Zhejiang Lab Program under the Grant No.2018DG0ZX01, and in part by the Key Science and Technology Innovation Support Program of Hangzhou under the Grant No.20172011A038.

References

[1] Y. Zhou, *et al.*, "CT-based radiomics signature: A potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma," *Abdom Radiol*, February 8, 2017.

[2] A. W. H. Chan, *et al.*, "Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection," *Journal of Hepatology*, August 27, 2018.
[3] H. Tsukama, H. Tanaka, W. Ajiki, and A. Oshima, "Liver cancer and it's prevention," *Asian Pasific Journal of Cancer Prevention*, vol. 6, no. 3, pp. 244-250, March 2005.
[4] T. T. Cheung, *et al*., "The Asia pacific consensus statement on laparoscopic liver resection for hepatocellular carcinoma: A report from the 7th Asia-Pacific primary liver cancer expert meeting held in Hong Kong," *Liver Cancer*, pp. 28-39, December 9, 2017.
[5] R. J. Gillies, P. E. Kinahan, Hricak, and Hedvig, "Radiomics: Images are more than pictures, they are data," *Radiological Society of North America*, pp. 563-577, November 18, 2015.
[6] V. Kumar, *et al*., "Radiomics: The process and the challenges," *European Journal of Cancer*, pp. 441-446, August 13, 2012.
[7] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," *IEEE Transaction on Cybernatics*, vol. 43, no. 6, pp. 1719-1733, December 2013.
[8] S. D. M. Raja and A. Shanmugam, "ANN and SVM based war scene classification using invariant moments and GLCM features: A comparative study," *IJMLC*, vol. 2, no. 6, December 2012.
[9] T. Evgeniou and P. Massimiliano, "Workshop on support vector machines: Theory and applications," *Machine Learning and Its Applications: Advanced Lectures*, pp. 249-257, January 2001.
[10] C. Cortes, Vapnik, and N. Vladimir, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, September 1995.
[11] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," *Journal of Theoretical and Applied Information Technology*, 2005.
[12] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," *Department of Computer Science and Information Engineering, National Taiwan University*, 2003.
[13] M. L. Huang, H. Hung, W. M. Lee, R. K. Li, and B. R. Jiang, "SVM-RFE based feature selection and taguchi parameters optimization for multiclass SVM classifier," *Scientific World Journal*, September 10, 2014.
[14] K. B. Duan, J. C. Rajapakse, H. Wang, and F. Azuaje, "Multiple SVM-RFE for gene selection in cancer classification with expression data," *IEEE Transactions on NanoBioscience*, vol. 4, no. 3, pp. 228-234, September 29, 2005.
[15] J. Wang, X. Li, H. Fan, and F. Li, "Classification of lip color based on multiple SVM-RFE," in *Proc. IEEE International Conference on Bioinformatics and Biomedicine Workshops*, USA, November 2011.
[16] H. O. Lancaster, *The Chi-squared Distribution*, Wiley, 1969.
[17] M. L. McHugh, "The Chi-square test of independence," *Biochem Med (Zagreb)*, pp. 143-149, 2013.
[18] P. Aonpong, T. Kasetkasem, P. Rakwatin, I. Kumazawa, and T. Chanwimaluang, "The Random forest approach for land cover mapping," in *Proc. IC-ICTES*, May 12, 2016.
[19] J. R. Quinlan, "Induction of decistion trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letter*, pp. 861-874, December 19, 2005.

Panyanat Aonpong was born in Suphanburi Province, Thailand, in 1993. He received a Bachelor of Engineering degree, Computer Engineering, from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2015. Master of Engineering degree, Information Communication Technology for Embedded Systems from Kasetsart University, Bangkok, Thailand, in 2017. Now, he is studying at the doctoral degree in Department of Information Science and Engineering, Ritsumeikan University, Shiga, Japan, since 2018. He is interested in computer image processing. He has experience working in satellite imagery when he was working in Geoinformatics Information and Space Technology Development Agency (GISTDA), the Thai government organization which works directly to the satellites and images.

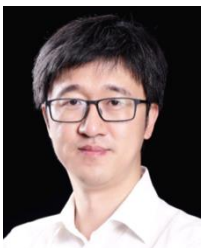


Qingqing Chen was born in Wenzhou, China in 1993. She received her bachelor's degree of medicine from Wenzhou medical university in 2016. And now she is studying at Zhejiang University School of medicine, China, for her doctoral degree. Her research mainly focuses on the artificial intelligence on liver lesion imaging, such as classification, li-rads category, prognosis and so on. She participated and presented at radiology society of North America (RSNA 2017) at Chicago and received a short-term exchange to Department of Information Science and Engineering, Ritsumeikan university for three months.

Yutaro Iwamoto received the B.E. and M.E., and D.E. degree from Ritsumeikan University, Kusatsu, Japan in 2011 and 2013, and 2017, respectively.
He is currently an Assistant Professor at Ritsumeikan University, Kusatsu, Japan. His current research interests include image restoration, segmentation, classification of medical imaging and deep learning

Lan-fen Lin was born in Pingyang, China in 1969. She received her B.S. and Ph.D. degrees in Aircraft Manufacture Engineering from Northwestern Polytechnical University in 1990, and 1995 respectively. She was post doctor in the college of Computer Science and Technology of ZJU form Jan. 1996 to Dec. 1997. Now she is a full professor at the college of Computer Science and Technology, Zhejiang University, China, vice director of Artificial Intelligence Institute in ZJU., and member of Advisory Expert Group for manufacturing industry informatization in Zhejiang Province.
Her research interests include Big data Analysis, Data Mining, Knowledge Management, etc. She is currently leading a research group to exploring AI technologies for medical imaging. She is the leader of more than 40 projects, include the subproject of Major State Basic Research Program (973) of china, and the National High Technology Research and Development Program of China. Lan-fen Lin has published more than 140 research papers in the journals or international conference proceedings.

Hongjie Hu, Ph.D., a chief physician and doctoral supervisor in Department of Radiology, Sir Run Run Shaw hospital, Zhejiang University school of medicine since 2015. He has engaged in medical imaging for over 30 years. He achieved his master's degree of medine (1992) and doctor's degree of medicine (1998) in Zhejiang Medical University. He went to some place for further education at Loma Linda University medical center in 2000, Myo clinic in 2008 and Cleveland University clinic in USA. He has rich clinical experience in cardiothoracic diagnosis and interventional radiotherapy.



Qiaowei Zhang was born in China in 1973. He received his Ph.D. and MD. degree in 2005 from Zhejiang university, Zhejiang, China. He was trained as resident and fellow doctor in Sir Run Run Shaw hospital, Hangzhou, China. He received short term fellowship/research training in Loma Linda University medical center, USA, and Grosshadern hospital, LMU, Germany. He is attending doctor of Sir Run Run Shaw hospital school of medicine, Zhejiang University, assistant director of radiology department.

Yen-Wei Chen was born in Hangzhou, China in 1962. He received his B.E. degree in 1985 from Kobe Univ., Kobe, Japan. He received his M.E. degree in 1987, and his D.E. degree in 1990, both from Osaka Univ., Osaka, Japan.

He was a research fellow at the Institute of Laser Technology, Osaka, from 1991 to 1994. From Oct. 1994 to Mar. 2004, he was an associate Professor and a professor in the department of Electrical and Electronic Engineering, Univ. of the Ryukyus, Okinawa, Japan. He is currently a professor at the college of Information Science and Engineering, Ritsumeikan University, Japan. He is also a visiting professor at the college of Computer Science and Technology, Zhejiang University, and Zhejiang Lab, Hangzhou, China. His research interests include medical image analysis, pattern recognition.

Yen-Wei Chen has published more than 300 research papers. He has received many distinguished awards. He is the Principal Investigator of several projects in bio-medical engineering and image analysis, funded by Japanese Government.