Diagnosing mouse liver fibrosis using machine learning models based on T2-weighted MRI radiomics

Lim, Hyun Ji

Department of Medicine

Graduate School

Yonsei University

Diagnosing mouse liver fibrosis using machine learning models based on T2-weighted MRI radiomics

Advisor Lee, Mi-Jung

A Dissertation Submitted
to the Department of Medicine
and the Committee on Graduate School
of Yonsei University in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Medical Science

Lim, Hyun Ji

June 2025

Diagnosing mouse liver fibrosis using machine learning models based on T2-weighted MRI radiomics

This Certifies that the Dissertation of Lim, Hyun Ji is Approved

Committee Chair	**Park, Mi-Suk**
Committee Member	**Lee, Mi-Jung**
Committee Member	**Je, Bo-Kyung**
Committee Member	**Rhee, Hyungjin**
Committee Member	**Koh, Hong**

Department of Medicine
Graduate School
Yonsei University
June 2025

ACKNOWLEDGEMENTS

I wish to foremost express my heartfelt gratitude to Professor Lee, Mi-Jung, who is my thesis director, for supporting my efforts with total commitment and facilitating every step of the dissertation process. My appreciation for her generous guidance and encouragement during my doctoral training is tremendous. I am also indebted to Professor Park, Mi-Suk, Je, Bo-Kyung, Rhee, Hyungjin, and Koh, Hong for sharing their insight and pertinent advice to assure the superior quality of this paper.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1F1A106277411).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Diagnosing mouse liver fibrosis using machine learning models based on T2-weighted MRI radiomics

We aimed to determine whether machine learning models developed using radiomics features of T2-weighted MRI can evaluate fibrosis in a mouse model with liver injury.

In this study, 6-week-old male C57BL/6 mice were fed a special diet of 3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) to induce cholestatic liver fibrosis. T2-weighted images were acquired using a 9.4 T MRI and images from the mice were further classified into the training set and test set to develop machine learning models. The training set consisted of 39 mice on the DDC diet for one to eleven weeks, with imaging conducted both before and after the feeding period. The test set consisted of 20 mice, with ten on a standard diet (feed without DDC) and ten on a two-week DDC diet. The characteristics of the T2-weighted images were analyzed using radiomics features. After MRI scanning, animals were sacrificed to establish a radiologic-pathologic correlation for liver fibrosis. Liver specimens were stained with Sirius Red, and the collagen proportional area (CPA, %) was measured to quantify liver fibrosis. Fibrosis grades were categorized as none, mild, or moderate based on the CPA. A commercial software package (Syngo.via Frontier, version 1.3.0; Siemens Healthineers, Munich, Germany) was used to develop multiple linear regression (MLR) and random forest (RF) models based on radiomics features. Model performance for evaluating the presence/absence of liver fibrosis was assessed using the area under the receiver operating characteristic curve (AUC). Additionally, correlations between the severity of liver fibrosis, assessed using CPA and fibrosis grade, and radiomics variables were analyzed using the Spearman and Kendall methods.

39 mice (baseline MRI: fibrosis MRI = 22: 28) were included in the training set, and 20 mice (control MRI: fibrosis MRI = 8: 10) were included in the test set. From the T2-weighted images of the training set, 845 radiomics features were extracted and ten of the most relevant features were selected. The AUC for diagnosing liver fibrosis was 0.991 for the MLR model (sensitivity 97.5%, specificity 94.2%), and the average AUC was 0.945 (sensitivity 88.9%, specificity 85.5%) for the

RF model after 10-fold cross-validation. When MLR and RF models were applied to the test set, the AUC of the MLR model was 0.645 (sensitivity 90%, specificity 33.3%), and the AUC of the RF model was 0.817 (sensitivity 90%, specificity 41.7%). In both the training and test sets, the key variables from the MLR and RF models showed correlations with liver CPA and fibrosis grade. In the training set, the most significant ten variables of the RF model were all significantly correlated with fibrosis grade (all, $P \leq 0.001$).

In this study, the random forest model derived from the radiomics features of T2-weighted MRI was able to help diagnose liver fibrosis in a mouse model.

Key words : liver fibrosis, magnetic resonance imaging, radiomics, mice

1. Introduction

The importance of non-invasive imaging for evaluating liver fibrosis in patients with chronic liver disease has been increasingly recognized, particularly in pediatric patients.[1,2] In children, radiation-free methods for diagnosing liver fibrosis include ultrasound, biopsy, and MRI. While measuring elasticity through ultrasound is well-established, its results are subject to variability depending on the ultrasound machine used and the skills of its operator. Thus, biopsy remains the gold standard for diagnosing liver fibrosis, despite its limitations, including sampling error, invasiveness, and the need for anesthesia in pediatric patients.[3] MRI, specifically MR elastography (MRE), has shown high accuracy for predicting fibrosis and can be performed in pediatric patients.[2,4-7] However, MRE requires specialized equipment that may not be available in every hospital. Also, while possible, performing MRE in infants and young children is challenging because mechanical drivers are not fit to their smaller bodies, making effective vibration transmission difficult. This challenge is further compounded by the rapid breathing of young children compared to adults.[4]

Biliary atresia is a rare but critical cholestatic liver disease where early diagnosis and treatment are essential. In patients suffering from this condition, liver cholestasis can begin prenatally or perinatally, and progressive liver fibrosis is sometimes observed at birth. Making an accurate and timely diagnosis is critically important for these younger patients despite its difficulty since early diagnosis of liver fibrosis can aid treatment planning and improve prognosis. If liver fibrosis can be diagnosed and quantified with conventional T2-weighted imaging—through a basic anatomical sequence of MRI that is routinely performed to assess pediatric liver disease—it would significantly benefit patient care.

Radiomics analyzes relationships between pixels in medical images such as CT, MRI, and PET to extract features ranging from first-order characteristics like entropy and skewness to higher-order features like texture analysis.[8-10] It quantifies complex image information, reflecting the heterogeneity of pathological tissues, and applies it to diagnosis, prognosis prediction, and treatment planning. Radiomics encompasses morphological, histogram, texture, and higher-order features, leveraging data mining and artificial intelligence techniques like machine learning and deep learning to optimize clinical decision-support systems.[11] By detecting subtle differences in shape, size, and

texture, it enhances the ability to classify diseases, determine disease presence and severity, and even predict critical clinical outcomes.[12] By combining radiomics with machine learning, medical imaging is advancing towards more precise and data-driven diagnostics and prognostics.[13-16] Technological developments so far suggest that radiomics using conventional MRI sequence images can be a potential alternative to conventional methods of evaluating liver fibrosis in children. However, there are limited studies on using T2-weighted MRI radiomics to predict liver pathology in children.[13,14,17-19]

We aimed to determine whether machine learning models derived using radiomics features of T2-weighted MRI can evaluate liver fibrosis in mice with cholestatic liver injury. If hepatic fibrosis can be evaluated using T2-weighted images in mice, these findings might be a springboard for the development of further diagnostic approaches using T2-weighted MRI for liver fibrosis in pediatric patients, as well as patients in hospitals without access to MRE equipment.

2. Materials and methods

2.1. Animal model

All animal research was conducted in accordance with the Laboratory Animals Welfare Act, the Guide for the Care and Use of Laboratory Animals, and the Guidelines and Policies for Rodent Experimentation provided by the Institutional Animal Care and Use Committee of the Yonsei University Health System (IACUC, 2022-0131). Radiologists and veterinarians monitored the condition of the studied mice.

The mouse model of this study consisted of mice that were orally administered a special rodent diet including 0.1% 3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) (DooYeol Biotech, Seocho-gu, Seoul, South Korea). The DDC model is a well-known in vivo model for cholestatic liver fibrosis caused by intraductal porphyrin plugs and obstruction of the intrahepatic bile ducts.[20] The mice were allowed free access to food, and no other feed was provided apart from the special feed containing DDC. In the training set, 39 male C57BL/6 mice, aged 6 weeks, were included. The mice were housed three per cage and underwent a one-week acclimatization period before receiving the special feed or undergoing MRI imaging. Mice underwent baseline imaging

before starting the DDC diet. Following the baseline scans, they were fed the DDC diet for durations ranging from 1 to 11 weeks. The mice were either sacrificed after the initial MRI imaging, or if their conditions were stable, they continued to be fed the DDC diet followed by additional MRI imaging. In the test set, 20 male C57BL/6 mice, aged 6 weeks, were included. The mice were housed five per cage and underwent a one-week acclimatization period before receiving the special diet or undergoing MRI imaging. In the test set, ten mice in the control group underwent MRI after two weeks on a normal diet, while ten mice in the fibrosis group underwent MRI after two weeks on the DDC diet (Figure 1).

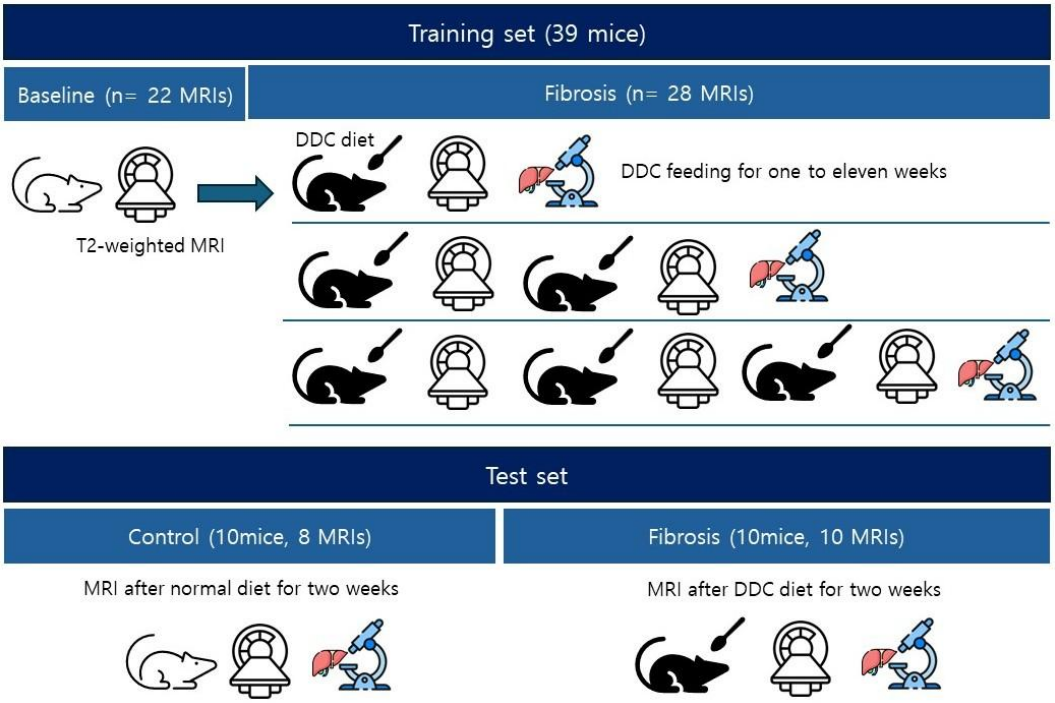


Figure 1. Mouse model in the training and test sets. DDC: 3,5-diethoxycarbonyl-1,4-dihydrocollidine

2.2. MRI scanning

All MRI experiments were performed with a 9.4 T Bruker BioSpec scanner (Ettlingen, Germany). Fasting was not conducted before MRI imaging. Mice were anesthetized using 1–2%

isoflurane and their heart rates were monitored during the MRI scan. Their body temperatures were maintained by circulating warm water. The MRI scans were conducted contingent upon the physical condition of the mice, with scans excluded if respiratory rates were unstable (less than 30 breaths per minute) following anesthesia. T2-weighted images were acquired using a 72 mm trans-receiver coil. Imaging parameters were as follows: matrix = 192×192, field of view = 35×35 mm, 30 slices, slice thickness 0.5 mm, effective echo time = 22 ms, repetition time = 2000 ms, and number of average = 2.

2.3. Fibrosis grading

Liver specimens were collected only if the mice could be dissected immediately after death to minimize tissue damage. Mice that died before or during MRI imaging were excluded. In the training set, liver tissues were obtained from 18 mice. In the test set, liver specimens were obtained from 18 mice.

The liver specimens were fixed in a formaldehyde solution for over 24 hours, embedded in paraffin blocks, and stained with Sirius Red. Using image analysis software (Image J, version 1.52a; National Institutes of Health), areas stained with Sirus Red were quantitatively assessed.[21,22] The collagen proportional area (CPA) was defined as the percentage of the area stained for Sirus Red. Images at 20 × magnification were converted to red-green-blue (RGB) images using the same operator-dependent threshold to detect areas of stained collagen. The researcher determined the threshold at which fibrosis appeared as red on the converted image. Additionally, areas stained red but not representing fibrosis, such as bile stains, were manually removed. Tissue within 5 mm of the capsule and large blood vessels were avoided when drawing region of interest. The severity of liver fibrosis was expressed with the CPA and classified into three grades (no fibrosis in cases with CPA < 4.8%; mild fibrosis with CPA of 4.8–10.3%; and moderate fibrosis with CPA > 10.3%).[23]

2.4. Segmentation

Axial liver MRI images without motion artifacts were included in the study in both the training and test sets. The liver was manually segmented on T2-weighted MR images by a radiologist (with 7 years of experience) using a commercial software package (Syngo.via Frontier, version 1.3.0; Siemens Healthineers, Munich, Germany). The regions of interest (ROIs)

were drawn three times: first at the right liver, second at the left liver, and third at the whole liver, excluding the hepatic hilar vascular structures. The ROIs were selected from the largest available cross-section within the entire volume of axial images (Figure 2).

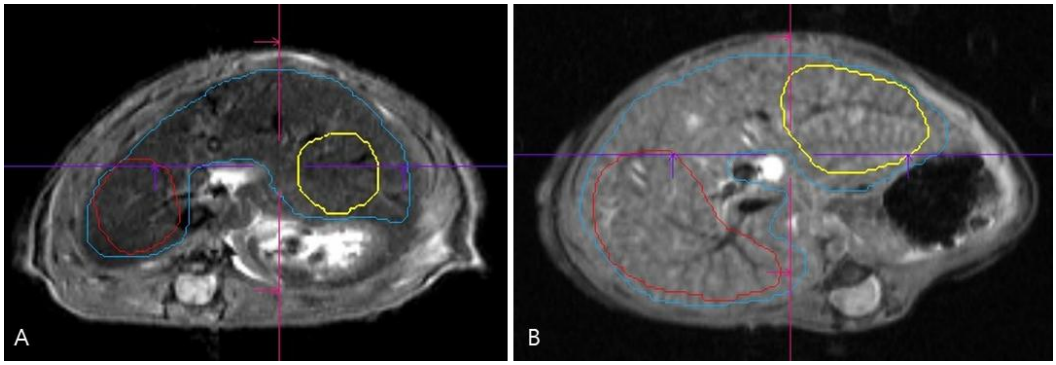


Figure 2. Regions of interest drawn at the right liver, left liver, and whole liver on T2-weighted axial images taken (A) at baseline and (B) after six weeks of the DDC diet. Pathology on the same day of imaging (B) confirmed moderate fibrosis. DDC: 3,5-diethoxycarbonyl-1,4-dihydrocollidine

2.5. Normalization

All MR images were resampled with a spatial resolution of 1×1×1 mm^3 using the linear interpolation method run by the software package for normalization before radiomics analysis.

Equation for Normalized Intensity: $f(x)= \frac{s(x-\mu x)}{\sigma x}$

In the above equation, f(x) is the normalized intensity, x is the original intensity, μx is the mean, and σx is the standard deviation of the image signal intensity of each drawn ROI.

2.6. Extraction, selection, and modeling of radiomics features

Radiomics models were derived using T2-weighted images from mice (Figure 3). Radiomics features were analyzed in a binary manner based on the presence or absence of liver fibrosis.

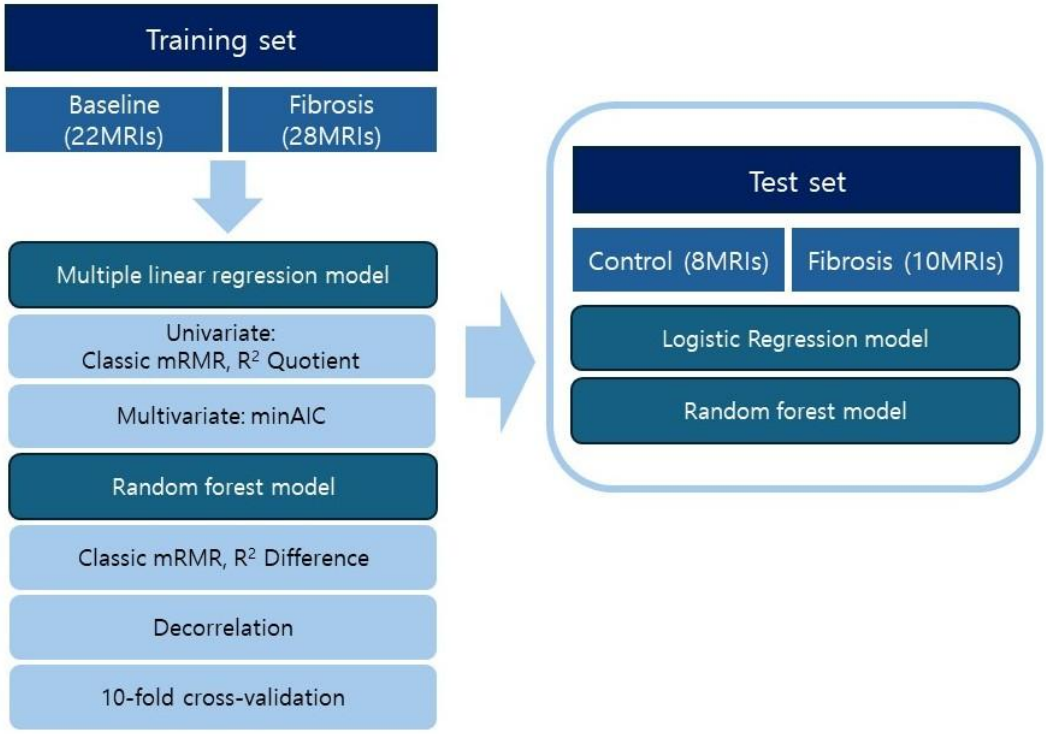


Figure 3. Development and training flow charts for the radiomics models.

mRMR: minimum Redundancy Maximum Relevance, minAIC: minimum Akaike Information Criterion.

……

A software package (syngo.via Frontier, version 1.3.0; Siemens Healthineers) was used for the radiomics analysis, which was developed based on the PyRadiomics library, version 3.0.1 (https://github.com/Radiomics/pyradiomics) and scikit-learn machine learning library (https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html). A total of 845 radiomics features—including 101 original and 744 wavelet-filtered— were extracted from the training set. A cluster map was generated through agglomerative hierarchical clustering to display associations between the found clusters of mice and radiomics features.

The multiple linear regression (MLR) model was built by selecting the optimal subset using the minimum Akaike Information Criterion (minAIC) method. The random forest (RF) model was built utilizing decision-making trees with Gini Inpurity. Both the MLR model and the RF

model underwent feature selection before training, and to minimize feature redundancy, decorrelation was performed using classic minimum redundancy maximum relevance (mRMR) and R^2 difference calculations (Figure 3). The 10-fold cross-validation was used to evaluate non-linear relationships and avoid overfitting. The average area under the receiver operating characteristic curves (AUCs) of the MLR model and RF model were provided for the training set. Both the trained MLR and RF models were tested in the test set.

2.7. Statistical analysis

Statistical analyses were performed using SPSS software (version 25.0; IBM Corp., Armonk, NY, USA) and R software (R Foundation for Statistical Computing). A commercial software package (Syngo.via Frontier, version 1.3.0; Siemens Healthineers, Munich, Germany) was used to analyze univariate and multivariate analyses and to build MLR and RF models. The diagnostic performance of the MLR and RF models for the presence of liver fibrosis was assessed using the AUC, sensitivity, and specificity. The varying degrees of liver fibrosis were represented using the CPA or the three-stage fibrosis grading system, and the correlations between these representations and the significant radiomics features of the MLR and RF models were analyzed using Spearman analysis and Kendall analysis. Statistical analyses accounted for repeated measurements within individual subjects as three ROIs were measured from each mouse. The intraclass correlation coefficient (ICC) was calculated to assess the reproducibility of significant radiomics feature values derived from the MLR and RF models across the three ROIs. P values less than 0.05 were considered statistically significant.

3. Results

3.1. Mice MRI acquisition

In the training set, 39 mice were initially enrolled in the experiment. Baseline imaging was successfully performed for 27 mice, while 12 did not undergo imaging due to unstable heart rates. Of the 27 baseline MRIs, 5 were excluded from analysis due to significant motion artifacts, resulting in 22 baseline MRIs being included in the final analysis. MRIs acquired after the DDC

diet were classified as the fibrosis group, yielding 30 MRIs. However, two of these were excluded due to severe motion artifacts, leaving 28 fibrosis MRIs for analysis. Among them, seven mice in good physical condition underwent MRI scans twice after beginning the DDC diet.

In the test set, ten mice in the control group underwent MRI after two weeks on a normal diet, producing ten control MRIs. Two MRIs were excluded due to motion artifacts. Another ten mice in the fibrosis group underwent MRI after two weeks of the DDC diet, yielding ten fibrosis MRIs.

3.2. Mice hepatic fibrosis results

Pathology specimens were obtained from 18 mice (CPA median 8.0%, ranging from 5.5 to 20.8%), including 11 mice in the mild fibrosis group (CPA median 7.2%, ranging from 5.5 to 8.6%) and 7 mice in the moderate fibrosis group (CPA median 13.3%, ranging from 11.1 to 20.8%) (Figure 4). Liver specimens were not collected from the control group. When analyzing pathologic correlations, the control group was classified as the no fibrosis group.

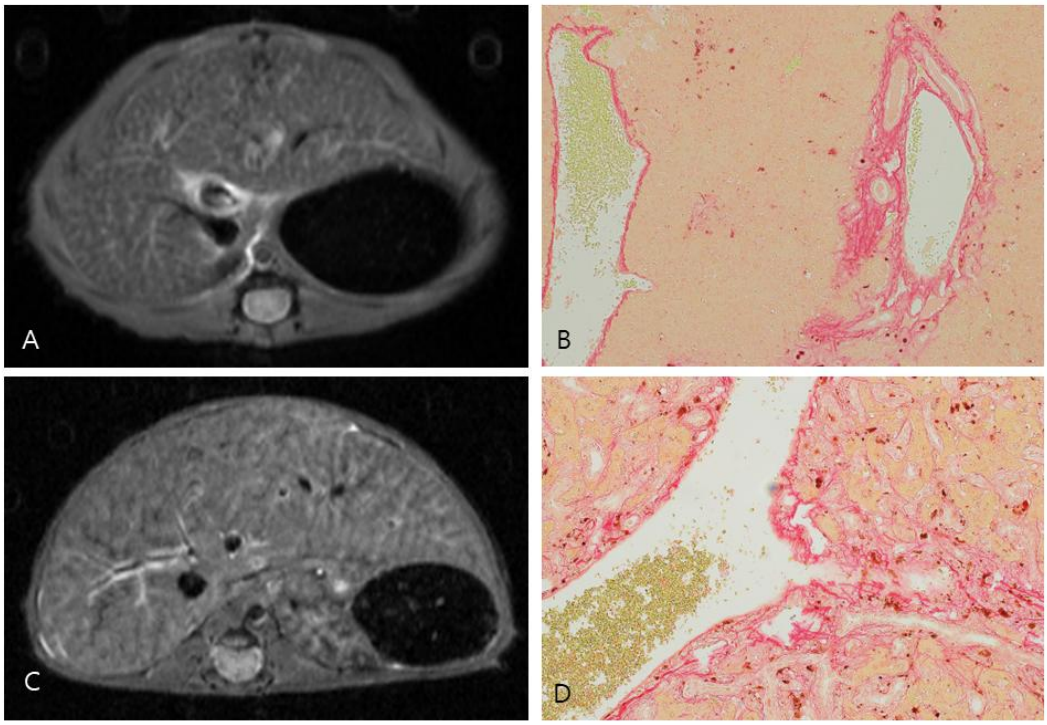


Figure 4. Axial T2-weighted images from (A) a mouse after three weeks of the DDC diet and (C) a mouse after nine weeks of the DDC diet. Liver specimens ([B] for mouse [A] image and [D] for mouse [C] image) were obtained on the same day of MRI and stained with Sirius Red. The CPA was 7.2% in (B), which graded the severity of liver fibrosis as mild and 17.5% in (D), which graded it as moderate. DDC: 3,5-diethoxycarbonyl-1,4-dihydrocollidine, CPA: collagen proportional area

In the test set, pathology specimens were obtained from all 18 mice, including 8 in the control group (CPA median 3.0%, ranging from 0.3% to 4.3%), which were all classified as the no fibrosis group and ten in the fibrosis group (CPA median 9.0%, ranging from 6.2% to 14.6%). In the fibrosis group, eight were in the mild fibrosis group (CAP median 8.2%, ranging from 6.2% to 10.3%), and two were in the moderate fibrosis group (CPA median 12.9%, ranging from 11.2% to 14.6%).

3.3. Development and testing of the radiomics model for the presence of hepatic fibrosis

A total of 845 radiomics features, including 101 original and 744 wavelet-filtered, were extracted from T2-weighted images in the training set (Figure 5).

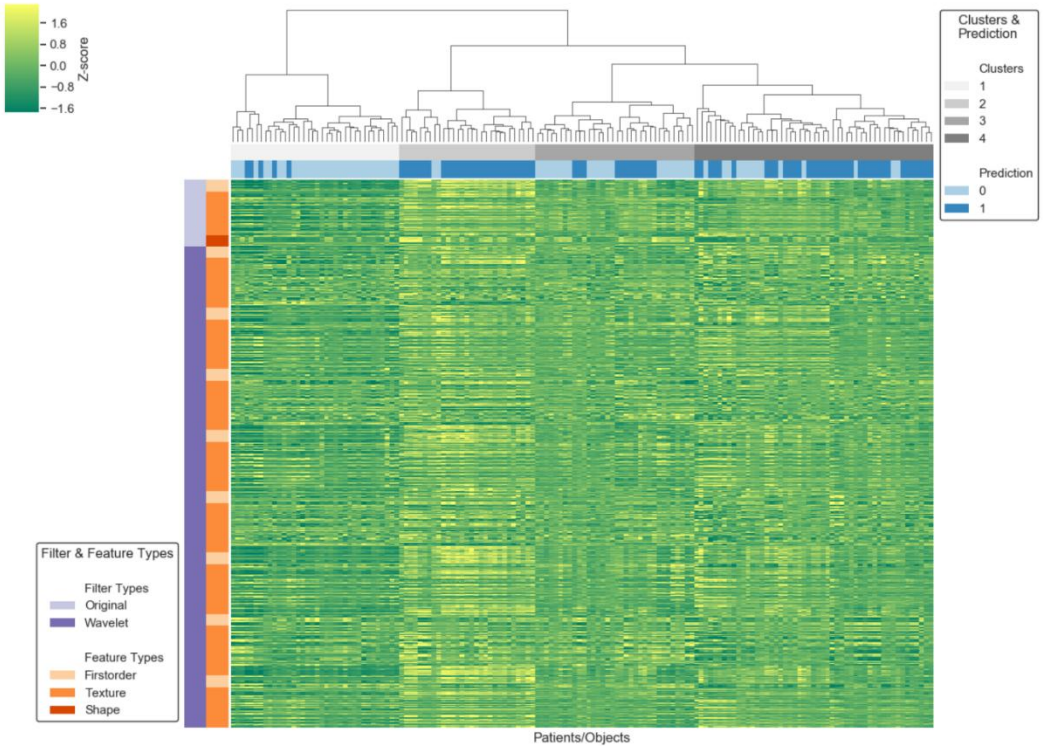


Figure 5. Cluster map from agglomerative hierarchical clustering to visualize the associations between the found clusters of subjects and radiomics features regarding liver fibrosis.

The top 10 most relevant features were selected, including wavelet- HLH_ GLRLM_ShortRunEmphasis, wavelet- LLL_firstorder_Minimum, LLH_firstorder_Entropy, wavelet-HLH_ GLSZM_HighGrayLevelZoneEmphasis, wavelet-HLL_firstorder_Mean, wavelet-LLL_ GLCM_Idn, wavelet-LHL_glcm_MCC, wavelet-HHH_glszm_SmallAreaHighGrayLevelEmphasis, wavelet-HLH_glcm_ClusterProminence, and wavelet-LHH_glszm_GrayLevelNonUniformityNormalized. The heatmap of these radiomics features showed differences between the baseline and fibrosis groups (Figure 6).

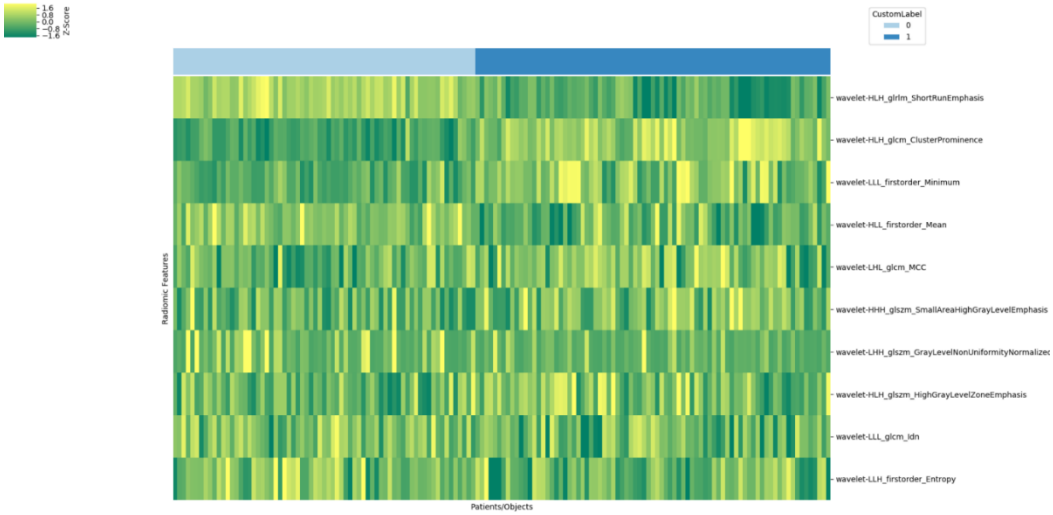


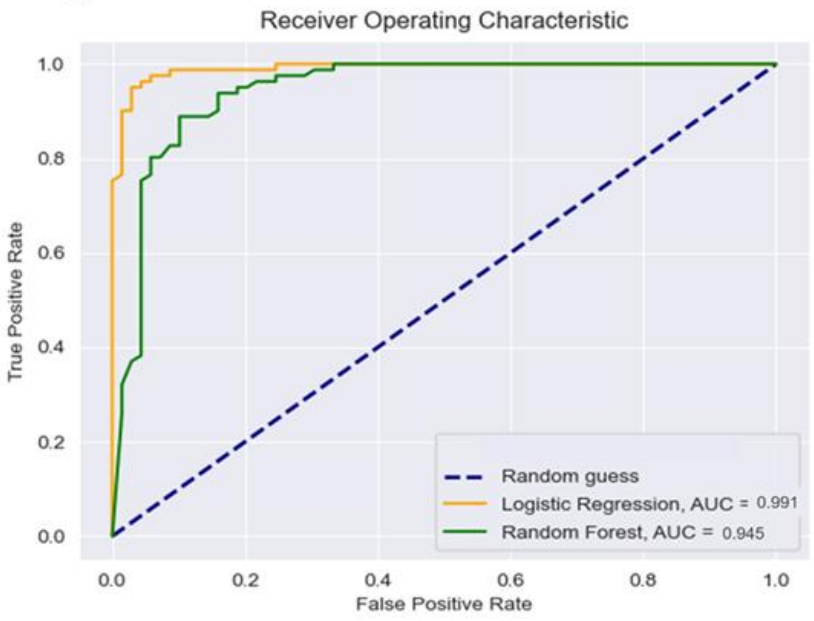
Figure 6. Heatmap of the ten most important radiomics features from T2-Weighted MRI for evaluating the presence of liver fibrosis.

The best subset among these radiomics features was identified using the minAIC method to build the MLR model. The MLR model includes wavelet-HLH_glrlm_ShortRunEmphasis (odds ratios [OR] 749.074, 95% confidence interval [CI] 32.857- 17077.623, P < 0.001), wavelet-LLL_firstorder_Minimum (OR 27.12, 95% CI 3.094-237.737, P < 0.001), wavelet-LLH_firstorder_Entropy (OR 2.561, 95% CI 1.077-6.092, P = 0.033), wavelet-HLH_glszm_HighGrayLevelZoneEmphasis (OR 1.974, 95% CI 0.703-5.539, P = 0.197), wavelet-HLL_firstorder_Mean (OR 3.632, 95% CI 0.981-13.443, P = 0.053), and wavelet-LLL_glcm_Idn (OR 1.533, 95% CI 0.552-4.261, P = 0.412). The MLR model showed excellent diagnostic performance for differentiating the presence and absence of liver fibrosis (AUC 0.991, sensitivity 97.5%, specificity 94.2%). Machine learning using classic mRMR with R^2 difference, followed by decorrelation to reduce feature redundancy, identified the eight most important random forest features for distinguishing the presence and absence of fibrosis from all the extracted radiomics features. These eight features were wavelet-HLH_glrlm_ShortRunEmphasis, wavelet-HLH_glcm_ClusterTendency, wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis, wavelet-HLH_glcm_ClusterProminence, wavelet-LLL_firstorder_Minimum, wavelet-HHH_glcm_DifferenceEntropy, wavelet-LLL_glcm_Correlation, and wavelet-

HLL_firstorder_Mean. The average AUC of the RF model showed excellent diagnostic performance for differentiating the presence and absence of liver fibrosis (AUC 0.945, sensitivity 88.9%, specificity 85.5%)

When the MLR model and RF model were applied to the test set, the MLR model showed poor performance (AUC 0.645, sensitivity 90%, specificity 33.3%), but the RF model showed good performance (AUC 0.817, sensitivity 90%, specificity 41.7%) (Figure 7 and Figure 8).

A



B

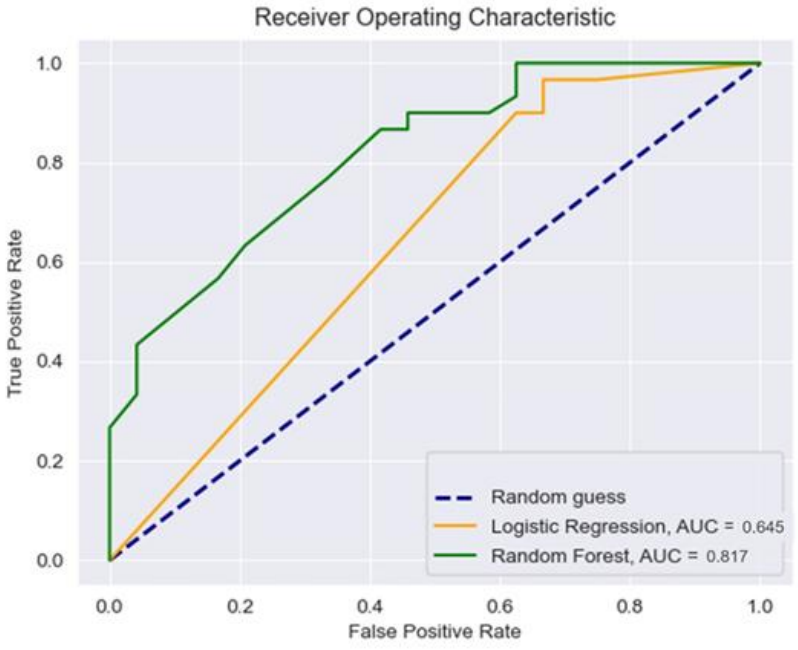


Figure 7. Receiver operating characteristics curves of the MLR and RF models from 10-fold cross-validation to differentiate the presence and absence of liver fibrosis in the (A) training set and (B) test set. MLR: multiple linear regression, RF: random forest.

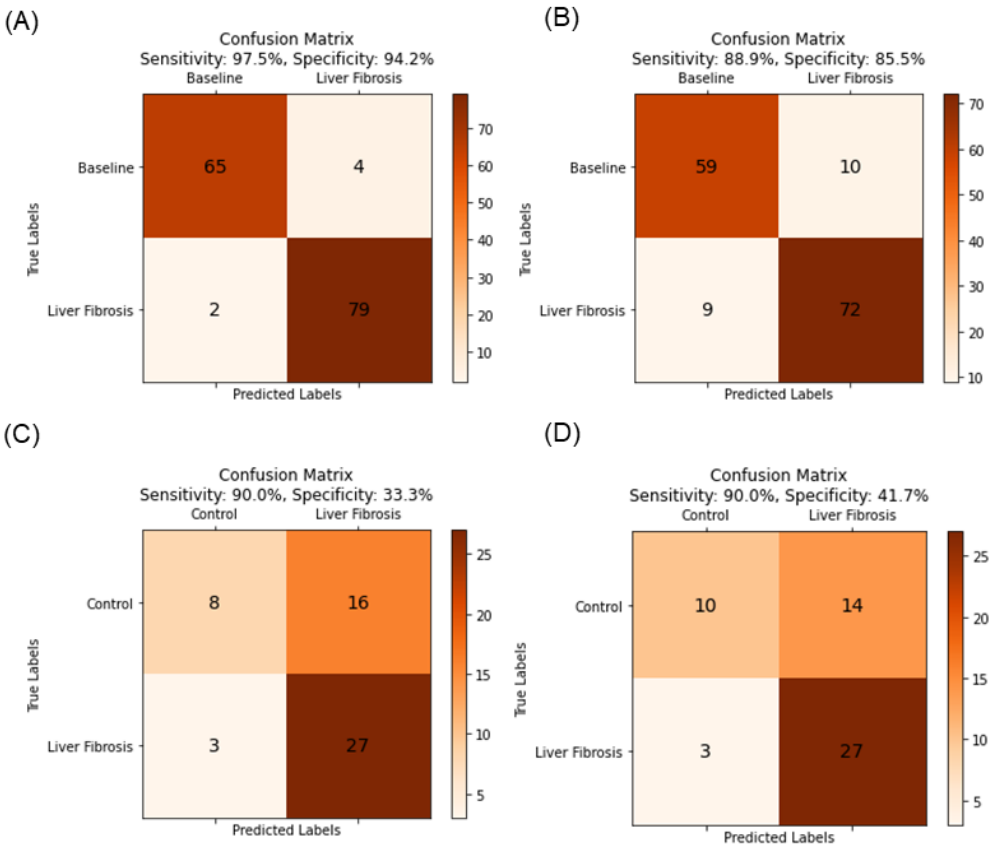


Figure 8. Sensitivity and specificity of the (A) MLR model and (B) RF model in the training set and sensitivity and specificity of the (C) MLR model and (D) RF model in test set.

MLR: multiple linear regression, RF: random forest.

3.4. Radiomics model and pathologic correlation of liver fibrosis

The radiomics features selected as the best subsets in the MLR and RF models showed correlations with liver fibrosis grades in the training set (Table 1).

Table 1. Correlation between radiomics features selected in the MLR and RF Models and fibrosis grades of mice liver pathology in the training set

Model	Correlation analysis	Variables	Coefficient (tau)	P-value
MLR model	Kendall analysis for fibrosis grades	wavelet-HLH_glrlm_ShortRunEmphasis	-0.689	<0.001
		wavelet-HLL_firstorder_Mean	-0.338	<0.001
		wavelet-LLL_firstorder_Minimum	0.374	<0.001
RF model	Kendall analysis for fibrosis grades	wavelet-HHH_glcm_DifferenceEntropy	0.474	<0.001
		wavelet-HLH_glcm_ClusterProminence	0.669	<0.001
		wavelet-HLH_glcm_ClusterTendency	0.669	<0.001
		wavelet-HLH_glrlm_ShortRunEmphasis	-0.689	<0.001
		wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis	-0.672	<0.001
		wavelet-HLL_firstorder_Mean	-0.338	<0.001
		wavelet-LLL_firstorder_Minimum	-0.374	<0.001
		wavelet-LLL_glcm_Correlation	-0.235	0.001

MLR: multiple linear regression, RF: random forest

The radiomics features selected as the best subset in the MLR and RF models correlated with varying degrees of liver fibrosis, including the CPA and liver fibrosis grade in the test set (Table 2).

Table 2. Correlations between radiomics features selected in the MLR and RF Models and the CPA and fibrosis grades of mice liver pathology in the test set

Model	Correlation analysis	Variables	Coefficient (rho)	Coefficient (tau)	P-value
MLR model	Spearman analysis for CPA	wavelet-HLH_glrlm_ShortRunEmphasis	0.505		<0.001
		wavelet-HLH_glszm_HighGrayLevelZoneEmphasis	0.385		0.004
		wavelet-LLH_firstorder_Entropy	0.348		0.010
		wavelet-LLL_firstorder_Minimum	0.387		0.004
	Kendall analysis for fibrosis grades	wavelet-HLH_glrlm_ShortRunEmphasis		0.462	<0.001
		wavelet-HLH_glszm_HighGrayLevelZoneEmphasis		0.339	0.002
		wavelet-LLH_firstorder_Entropy		0.218	0.046
		wavelet-LLL_firstorder_Minimum		0.382	<0.001
RF model	Spearman analysis for CPA	wavelet-HLH_glcm_ClusterProminence	0.497		<0.001
		wavelet-HLH_glcm_ClusterTendency	0.535		<0.001
		wavelet-HLH_glrlm_ShortRunEmphasis	0.505		<0.001
		wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphas	-0.312		0.022

	is			
	wavelet-LLL_firstorder_Minimum	0.387		0.004
Kendall analysis for fibrosis grades	wavelet-HLH_glcm_ClusterProminence		0.448	<0.001
	wavelet-HLH_glcm_ClusterTendency		0.471	<0.001
	wavelet-HLH_glrlm_ShortRunEmphasis		0.462	<0.001
	wavelet-LLL_firstorder_Minimum		0.382	<0.001

MLR: multiple linear regression, RF: random forest, CPA: collagen proportional area

3.4.1 Correlation between MLR model variables and mice liver fibrosis grade in the training set

When performing a Kendall analysis to assess correlations between fibrosis grades and the variables in the best subset of the MLR model, wavelet-HLH_glrlm_ShortRunEmphasis, wavelet-HLL_firstorder_Mean, and wavelet-LLL_firstorder_Minimum, ($P < 0.001$ for all) showed significant correlation (Table 1 and Figure 9).

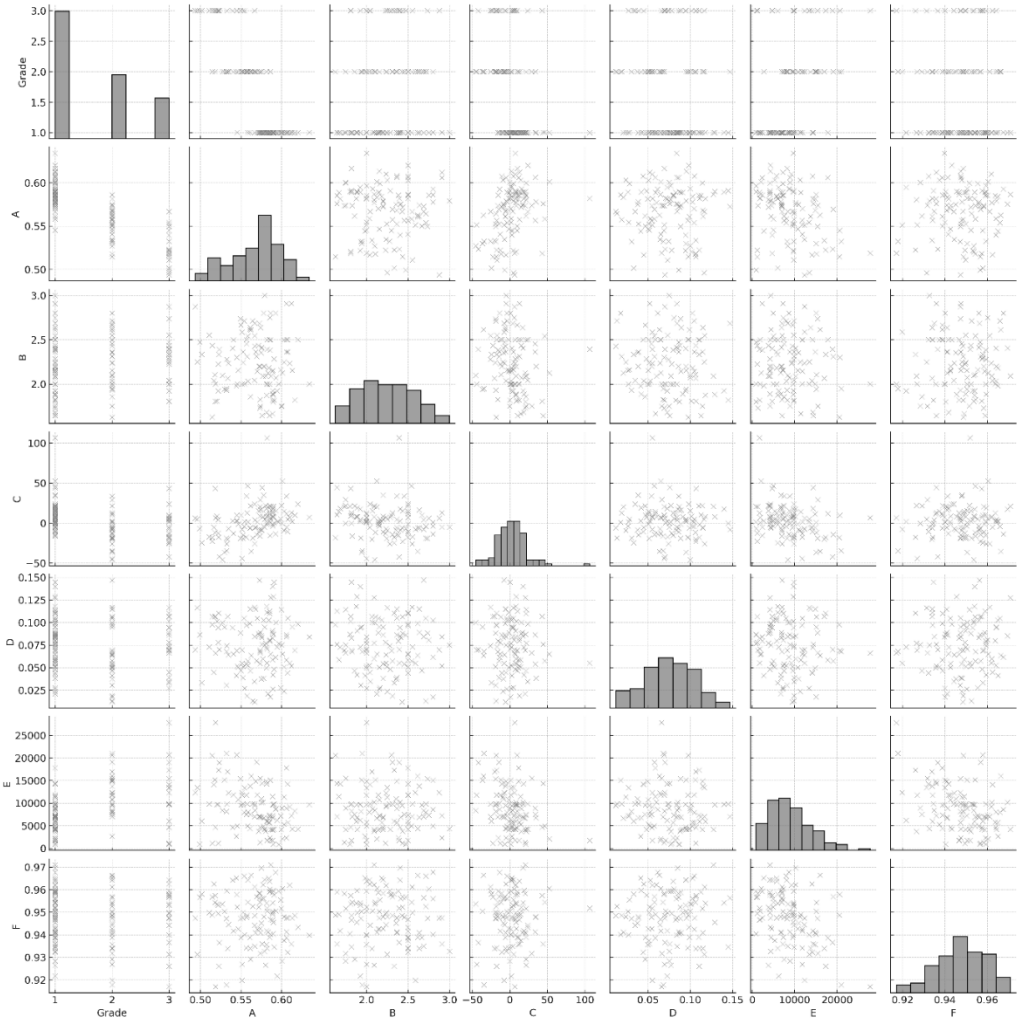


Figure 9. Kendall correlation analysis between liver fibrosis grades and variables in the best subset of the MLR model in the training set.

MLR: multiple linear regression. A: wavelet-HLH_glrlm_ShortRunEmphasis, B: wavelet-HLH_glszm_HighGrayLevelZoneEmphasis, C: wavelet-HLL_firstorder_Mean, D: wavelet-LLH_firstorder_Entropy, E: wavelet-LLL_firstorder_Minimum, F: wavelet-LLL_glcm_Idn

3.4.2. Correlation between RF model variables and mice liver fibrosis grade in the training set

When performing a Kendall correlation analysis between fibrosis grades and the variables in the best subset of the RF model, all eight important variables showed significant correlations, including wavelet-HHH_glcm_DifferenceEntropy ($P < 0.001$), wavelet-HLH_glcm_ClusterProminence ($P < 0.001$), wavelet-HLH_glcm_ClusterTendency ($P < 0.001$), wavelet-HLH_glrlm_ShortRunEmphasis ($P < 0.001$), wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis ($P < 0.001$), wavelet-HLL_firstorder_Mean ($P < 0.001$), wavelet-LLL_firstorder_Minimum ($P < 0.001$), and wavelet-LLL_glcm_Correlation ($P = 0.001$) (Table1 and Figure 10).

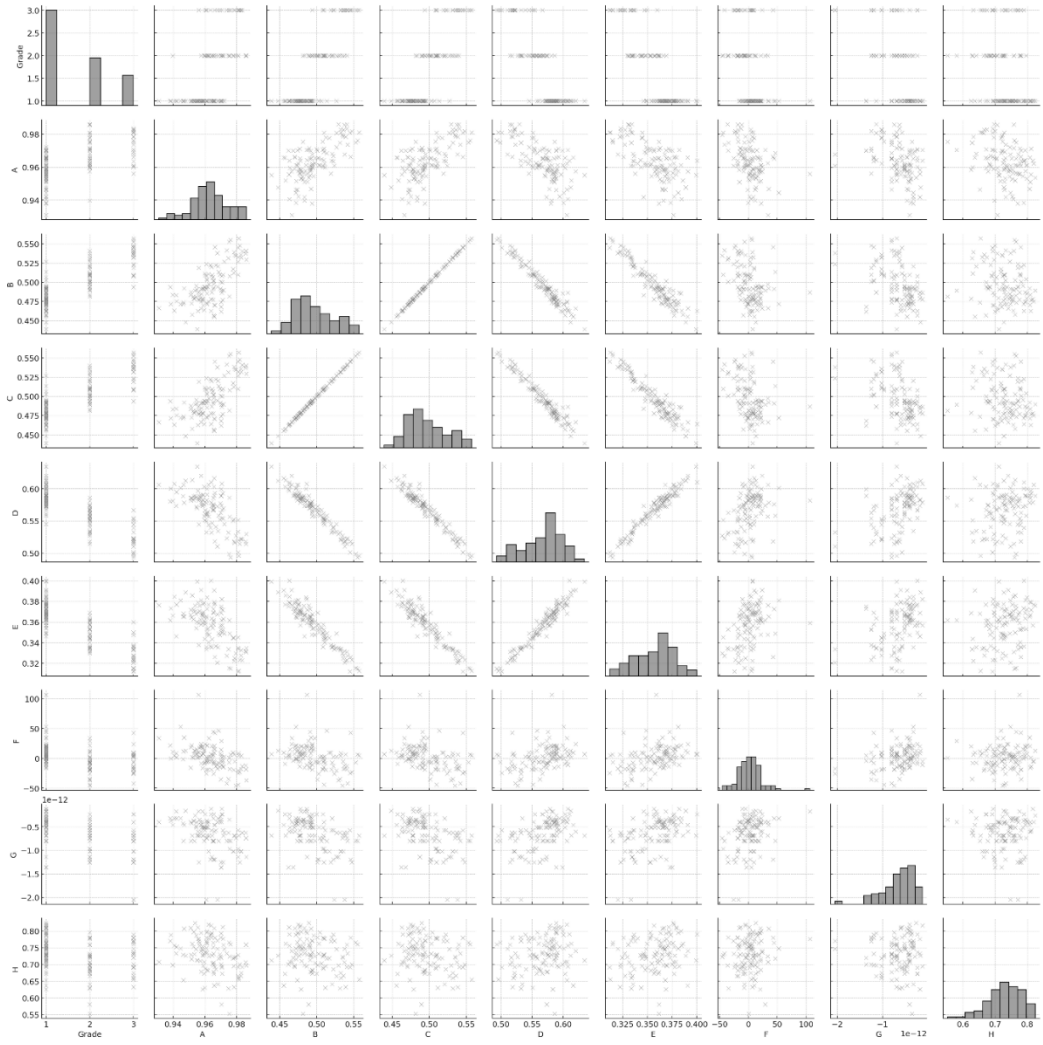


Figure 10. Kendall correlation analysis between liver fibrosis grades and variables in the best subset of the RF model in the training set.

RF: random forest. A: wavelet-HHH_glcm_DifferenceEntropy, B: wavelet-HLH_glcm_ClusterProminence, C: wavelet-HLH_glcm_ClusterTendency, D: wavelet-HLH_glrlm_ShortRunEmphasis, E: wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis, F: wavelet-HLL_firstorder_Mean, G: wavelet-LLL_firstorder_Minimum, H: wavelet-LLL_glcm_Correlation

3.4.3. Correlation between MLR model variables and the CPA of mice liver pathology and liver fibrosis grades in the test set

When performing a Spearman analysis to assess correlations between the CPA of mice liver pathology and the variables in the best subset of the MLR model, wavelet-HLH_glrlm_ShortRunEmphasis (P < 0.001), wavelet-HLH_glszm_HighGrayLevelZoneEmphasis (P = 0.004), wavelet-LLH_firstorder_Entropy (P = 0.010), and wavelet-LLL_firstorder_Minimum (P = 0.004) showed significant correlation (Table 2). The Kendall correlation analysis between fibrosis grades and the variables in the best subset of the MLR model showed correlations in wavelet-HLH_glrlm_ShortRunEmphasis (P < 0.001), wavelet-HLH_glszm_HighGrayLevelZoneEmphasis (P = 0.002), wavelet-LLH_firstorder_Entropy (P =0.046), and wavelet-LLL_firstorder_Minimum (P < 0.001) (Table 2, Figure11).

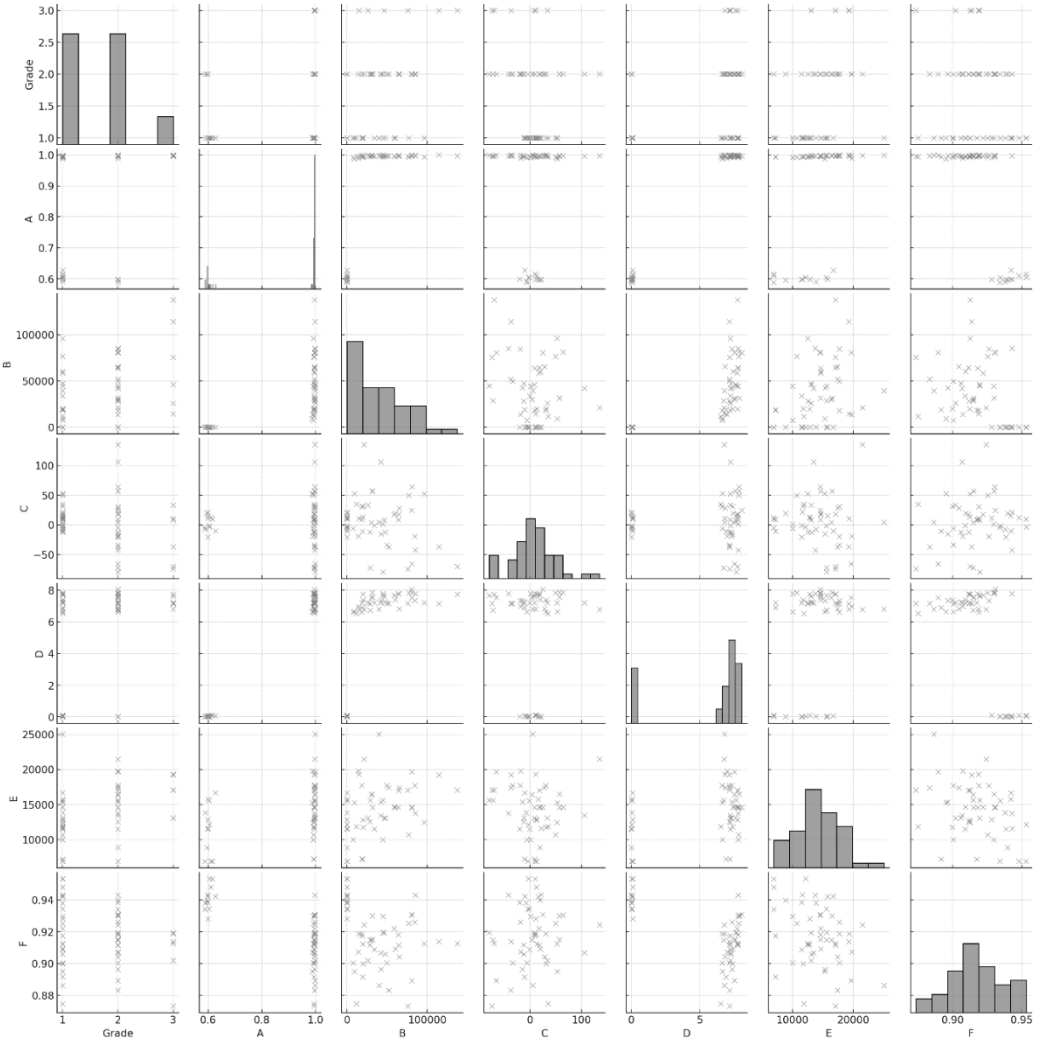


Figure 11. Kendall correlation analysis between liver fibrosis grades and variables in the best subset of the MLR model in the test set.

MLR: multiple linear regression. A: wavelet-HLH_glrlm_ShortRunEmphasis, B: wavelet-HLH_glszm_HighGrayLevelZoneEmphasis, C: wavelet-HLL_firstorder_Mean, D: wavelet-LLH_firstorder_Entropy, E: wavelet-LLL_firstorder_Minimum, F: wavelet-LLL_glcm_Idn

3.4.4. Correlations between RF model variables and the CPA of mice liver pathology and liver fibrosis grades in the test set

When performing a Spearman analysis to assess correlations between the CPA of mice liver pathology and the variables in the best subset of the RF model, wavelet-HLH_glcm_ClusterProminence ($P < 0.001$), wavelet-HLH_glcm_ClusterTendency ($P < 0.001$), wavelet-HLH_glrlm_ShortRunEmphasis ($P < 0.001$), wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis ($P = 0.022$), and wavelet-LLL_firstorder_Minimum ($P = 0.004$) showed significant correlation (Table 2). Additionally, the Kendall correlation analysis between fibrosis grades and four of the important variables in the best subset of the RF model showed significant correlations including wavelet-HLH_glcm_ClusterProminence ($P < 0.001$), wavelet-HLH_glcm_ClusterTendency ($P < 0.001$), wavelet-HLH_glrlm_ShortRunEmphasis ($P < 0.001$), and wavelet-LLL_firstorder_Minimum ($P < 0.001$) (Table 2, Figure 12).

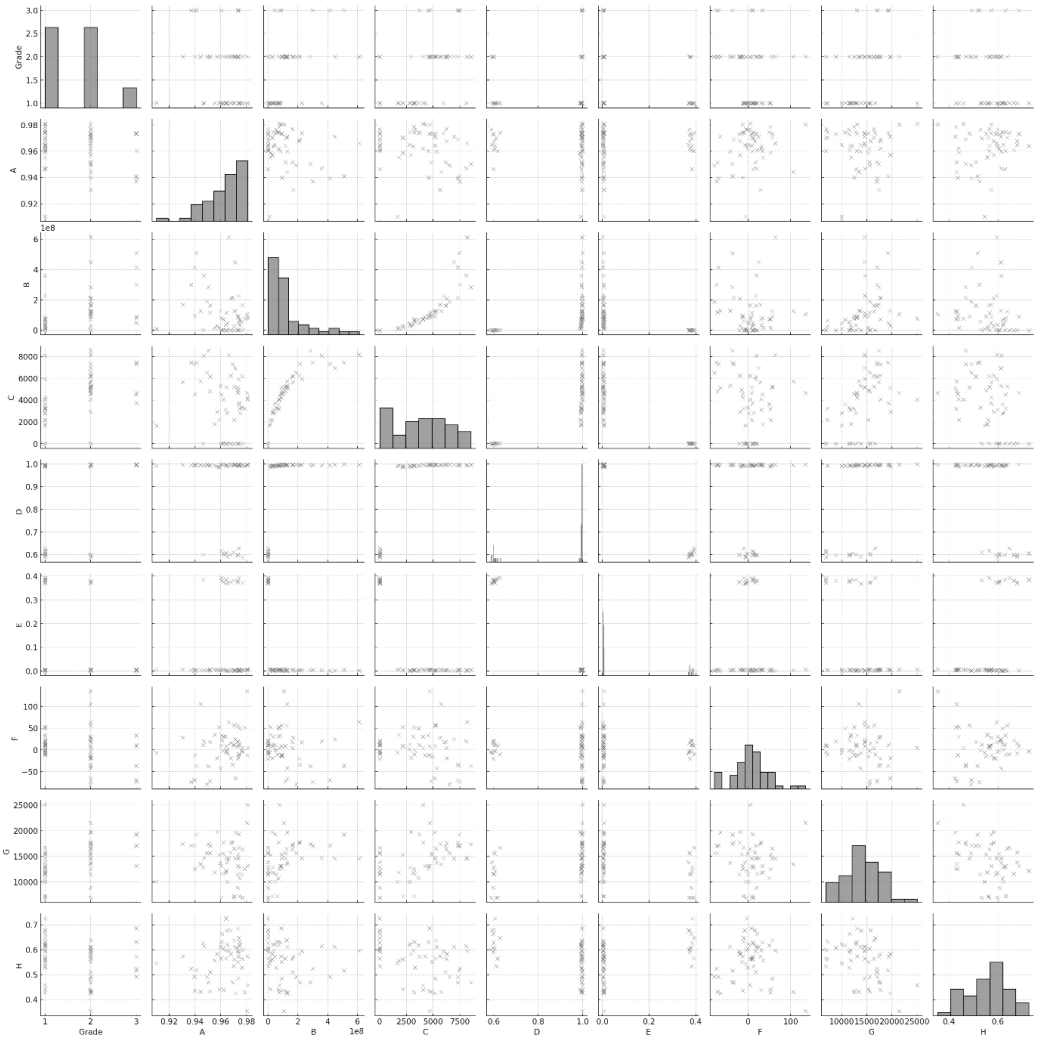


Figure 12. Kendall correlation analysis between liver fibrosis grades and variables in the best subset of the RF model in test set.

RF: random forest. A: wavelet-HHH_glcm_DifferenceEntropy, B: wavelet-HLH_glcm_ClusterProminence, C: wavelet-HLH_glcm_ClusterTendency, D: wavelet-HLH_glrlm_ShortRunEmphasis, E: wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis, F: wavelet-HLL_firstorder_Mean, G: wavelet-LLL_firstorder_Minimum, H: wavelet-LLL_glcm_Correlation

3.5. Intraclass correlation analysis of radiomics features for the right, left, and whole liver ROIs

In the MLR model, the average ICC of the important radiomics features was fair (0.56). HLH_glrlm_ShortRunEmphasis showed excellent ICC (0.88), LLH _firstorder_Entropy and LLL _firstorder_Minimum showed good ICC (0.73 and 0.66, respectively), LLL_glcm _Idn showed fair ICC (0.50), and HLL _firstorder_Mean and HLH_glszm _HighGrayLevelZoneEmphasis showed poor ICC (0.29 and 0.28, respectively).

In the RF model, the average ICC of the important radiomics features was good (0.72). Wavelet-HLH_glcm_ClusterProminence and wavelet-HLH_glcm_ClusterTendency showed excellent ICC (0.90), wavelet-HLH_glrlm_ShortRunEmphasis and wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis showed excellent ICC (0.88), wavelet-HHH_glcm_DifferenceEntropy and wavelet-LLH_firstorder_Minimum showed good ICC (0.65), wavelet-LLL_glcm_Correlation showed fair ICC (0.60), and wavelet-HLL_firstorder_Mean showed poor ICC (0.29).

4. Discussion

In our study, we developed and tested machine learning models, including the MLR and RF models, based on radiomics features extracted from T2-weighted images of mice to diagnose cholestatic liver fibrosis. Both models demonstrated excellent performance in differentiating the presence and absence of liver fibrosis in the training set. For the test set, only the RF model showed good diagnostic performance for liver fibrosis. Additionally, key radiomics features from the MLR and RF models showed significant correlations with liver fibrosis grades in the training set, and both the CPA and liver fibrosis grades in the test set.

We selected 6-week-old male C57BL/6 mice as our experimental subjects, as their small liver size and rapid breathing rate[24] closely reflect those of neonates. The DDC diet was chosen to induce liver fibrosis because prolonged DDC feeding in mice replicates key histopathological features of human cholestatic liver fibrosis, including the bile duct remodeling that leads to ductular reactions, periductular fibrosis, and infiltration of inflammatory cells.[25] This DDC model effectively mirrors the liver fibrosis observed in cholestatic conditions, making it valuable for studying fibrosis in a non-invasive manner.

Although the pathophysiology of mice may not exactly match that of neonatal liver disease such as biliary atresia, we believed our mouse model could appropriately represent neonatal cholestatic liver fibrosis and any findings from this model could be extended to research for the pediatric population.

Radiomics involves post-image processing to extract features from radiological images and transform the data using artificial intelligence techniques, such as machine learning, to enhance predictive accuracy.[11,26] It is based on the premise that pathophysiological tissues and organs hold substantial information that can be quantified and distinguished from normal tissues and organs. Radiomics considers texture as the spatial arrangement of predefined voxels, enabling the interpretation of complex image features.[27] Mathematical analysis of these characteristic arrangements helps distinguish normal tissues from abnormal tissues. The heterogeneity of the selected features mirrors the heterogeneity of histopathological changes.

Our study suggests that MLR and RF models developed from MRI radiomics features may have the potential to diagnose liver fibrosis in humans. This finding aligns with that of previously published papers showing that radiomics and machine learning based on MRI images can diagnose liver fibrosis in older patients.[8] In adult studies, radiomics and machine learning approaches that incorporate T2-weighted images[28], T1-weighted images[29], R2* mapping[30], and gadoxetic acid-enhanced MRI[14,29,31] have been explored to diagnose liver fibrosis. In adolescent studies, machine learning models that incorporate radiomics data from T2-weighted images, along with liver volume and liver chemical shift-encoded fat fraction, have shown promise when diagnosing liver fibrosis.[13] Moreover, advances in artificial intelligence have enabled liver segmentation using a convolutional neural network pre-trained on the ImageNet archive and deep transfer learning to detect liver cirrhosis with expert-level accuracy in adult patients.[32] However, studies that focus on infants and young children are rare and far between. One of the few studies published indicated that a radiomic-based nomogram derived from T1-weighted and T2-weighted images could aid in diagnosing liver fibrosis in young children, including infants.[18] One animal study compared the performance of individual radiomic models based on multiparametric MRI sequence and showed good performance for predicting early liver fibrosis.[33] Our study used T2-weighted images as it is routinely performed in young patients and is less susceptible to motion artifacts due to its higher sensitivity to magnetic field changes.[19,34] The radiomics features gained from T2-weighted images could differentiate liver fibrosis based on fine-grained details. ……

The most important features selected from the RF model were wavelet-transformed features, which effectively capture multi-scale texture and intensity patterns by separating high-frequency and low-frequency components of images.[10,27,35] The HLH and HHH components primarily focus on fine textures and high-frequency details, essential for identifying subtle patterns and intricate textures. Meanwhile, the LLL components highlight broader, low-frequency structures crucial for understanding the overall composition and intensity distribution within an image. The features selected in both MLR and RF models—wavelet-HLH_glrlm_ShortRunEmphasis, wavelet-LLL_firstorder_Minimum, and wavelet-HLH_glcm_ClusterProminence—indicate that both models rely on capturing fine texture details and broader structural characteristics. Wavelet-HLH_glrlm_ShortRunEmphasis and wavelet-HLH_glcm_ClusterProminence suggest sensitivity to detailed textures and clustering patterns, while wavelet-LLL_firstorder_Minimum reflects the importance of low-intensity regions in the overall image structure. The significance of these features makes intuitive sense because the normal liver shows relatively homogenous and hypointense signal intensity, and the liver starts to show heterogenous and hyperintense signal intensity as fibrosis progresses on T2-weighted images. The model's reliance on these diverse features suggests it is well-equipped to capture even small differences in data, leading to more accurate evaluation of liver fibrosis. In this study, the RF model outperformed the MLR model. This is probably because the RF model consists of multiple decision trees trained on different subsets of data or feature subspaces, with each tree built using independently sampled random vectors.[36] By aggregating the predictions of these diverse trees, the RF model enhances generalization and reduces overfitting, leading to improved performance.

Our findings show that key radiomics features from MLR and RF models correlate with varying degrees of liver fibrosis and seem promising as diagnostic factors. The key features that were significantly correlated to the CPA of mice liver pathology and fibrosis grades were all wavelet features. Wavelet filtering methods involve decomposing the original image to offer benefits such as adjustable spatial resolution for optimal texture representation, enhanced texture visibility, and flexibility in choosing wavelet functions.[27] Previous research has also shown that texture-based classification of liver fibrosis using MRI is feasible.[28] Recent studies showed that incorporating clinical and radiomic features led to better diagnostic performance for liver stiffness than radiomics features alone.[9,13,37] Further studies that refine models by combining radiomics and clinical data to diagnose liver fibrosis grade could possibly decrease the need for MRE which needs additional exam

time and specialized equipment.

This study has several limitations. First, the small sample size used to develop the machine learning model limits its generalizability. Due to ethical and practical constraints in animal experiments, we used a minimum number of subjects and analyzed three ROIs per mouse to maximize the available data. However, this approach may cause overfitting and statistical dependence. To mitigate these issues, we applied methods such as mRMR, R2 difference, decorrelation, and 10-fold cross-validation. Second, a single radiologist performed ROI segmentation, which can raise concerns about reproducibility. To address this, we analyzed ROIs from multiple liver regions (right, left, and whole liver), and the ICC of key radiomics features from MLR and RF models showed fair to good correlation. Furthermore, our models did not directly diagnose the CPA or liver fibrosis grade but instead explored correlations between radiomics features and these parameters, which limits the direct clinical applicability of our results. The small number of liver pathology samples also restricted the development and validation of a predictive model. We evaluated the performance of both the MLR and RF models, but we did not analyze the statistical significance of the difference in their performance. Nevertheless, our findings suggest that T2-weighted MRI-based radiomics features from MLR and RF models have the potential to differentiate fibrosis severity. Future studies with larger datasets and methods that facilitate direct application to human patients including neonates, are needed to validate our findings and explore the actual clinical utility of the two machine learning models.

5. Conclusion

In conclusion, machine learning models based on radiomics features from T2-weighted images were trained and tested in a mouse liver fibrosis model. The RF model showed excellent diagnostic performance for diagnosing liver fibrosis in both training and test sets of our animal study. Key radiomics features showed significant correlations with the CPA and liver fibrosis grade. Therefore, our RF model shows potential for diagnosing liver fibrosis in pediatric patients, albeit with further research.

———

References

1. Zheng T, Qu Y, Chen J, Yang J, Yan H, Jiang H, et al. Noninvasive diagnosis of liver cirrhosis: qualitative and quantitative imaging biomarkers. Abdom Radiol (NY) 2024;49:2098-115.
2. Zerunian M, Pucciarelli F, Masci B, Siciliano F, Polici M, Bracci B, et al. Updates on Quantitative MRI of Diffuse Liver Disease: A Narrative Review. Biomed Res Int 2022;2022:1147111.
3. Chen BR, Pan CQ. Non-invasive assessment of fibrosis and steatosis in pediatric non-alcoholic fatty liver disease. Clin Res Hepatol Gastroenterol 2022;46:101755.
4. Binkovitz LA, El-Youssef M, Glaser KJ, Yin M, Binkovitz AK, Ehman RL. Pediatric MR elastography of hepatic fibrosis: principles, technique and early clinical experience. Pediatr Radiol 2012;42:402-9.
5. Schwimmer JB, Behling C, Angeles JE, Paiz M, Durelle J, Africa J, et al. Magnetic resonance elastography measured shear stiffness as a biomarker of fibrosis in pediatric nonalcoholic fatty liver disease. Hepatology 2017;66:1474-85.
6. Sawh MC, Newton KP, Goyal NP, Angeles JE, Harlow K, Bross C, et al. Normal range for MR elastography measured liver stiffness in children without liver disease. J Magn Reson Imaging 2020;51:919-27.
7. Trout AT, Anupindi SA, Gee MS, Khanna G, Xanthakos SA, Serai SD, et al. Normal Liver Stiffness Measured with MR Elastography in Children. Radiology 2020;297:663-9.
8. Wang XM, Zhang XJ. Role of radiomics in staging liver fibrosis: a meta-analysis. BMC Med Imaging 2024;24:87.
9. Xiao L, Zhao H, Liu S, Dong W, Gao Y, Wang L, et al. Staging liver fibrosis: comparison of radiomics model and fusion model based on multiparametric MRI in patients with chronic liver disease. Abdom Radiol (NY) 2024;49:1165-74.
10. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104-e7.

11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-62.
12. Wei H, Shao Z, Fu F, Yu X, Wu Y, Bai Y, et al. Value of multimodal MRI radiomics and machine learning in predicting staging liver fibrosis and grading inflammatory activity. Br J Radiol 2023;96:20220512.
13. He L, Li H, Dudley JA, Maloney TC, Brady SL, Somasundaram E, et al. Machine Learning Prediction of Liver Stiffness Using Clinical and T2-Weighted MRI Radiomic Data. AJR Am J Roentgenol 2019;213:592-601.
14. Park HJ, Lee SS, Park B, Yun J, Sung YS, Shim WH, et al. Radiomics Analysis of Gadoxetic Acid-enhanced MRI for Staging Liver Fibrosis. Radiology 2019;290:380-7.
15. Nitsch J, Sack J, Halle MW, Moltz JH, Wall A, Rutherford AE, et al. MRI-based radiomic feature analysis of end-stage liver disease for severity stratification. Int J Comput Assist Radiol Surg 2021;16:457-66.
16. Popa SL, Ismaiel A, Abenavoli L, Padureanu AM, Dita MO, Bolchis R, et al. Diagnosis of Liver Fibrosis Using Artificial Intelligence: A Systematic Review. Medicina (Kaunas) 2023;59.
17. Lee MJ, Kim MJ, Yoon CS, Han SJ, Park YN. Evaluation of liver fibrosis with T2 relaxation time in infants with cholestasis: comparison with normal controls. Pediatr Radiol 2011;41:350-4.
18. Yang Y, Zhang X, Zhao L, Mao H, Cai TN, Guo WL. Development of an MRI-Based Radiomics-Clinical Model to Diagnose Liver Fibrosis Secondary to Pancreaticobiliary Maljunction in Children. J Magn Reson Imaging 2023;58:605-17.
19. Donato H, França M, Candelária I, Caseiro-Alves F. Liver MRI: From basic protocol to advanced techniques. Eur J Radiol 2017;93:30-9.
20. Pose E, Sancho-Bru P, Coll M. 3,5-Diethoxycarbonyl-1,4-Dihydrocollidine Diet: A Rodent Model in Cholestasis Research. Methods Mol Biol 2019;1981:249-57.
21. Tsochatzis E, Bruno S, Isgro G, Hall A, Theocharidou E, Manousou P, et al. Collagen proportionate area is superior to other histological methods for sub-classifying cirrhosis and determining prognosis. J Hepatol 2014;60:948-54.

22. Buzzetti E, Hall A, Ekstedt M, Manuguerra R, Guerrero Misas M, Covelli C, et al. Collagen proportionate area is an independent predictor of long-term outcome in patients with non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2019;49:1214-22.
23. Lv J, Xu Y, Xu L, Nie L. Quantitative Functional Evaluation of Liver Fibrosis in Mice with Dynamic Contrast-enhanced Photoacoustic Imaging. Radiology 2021;300:89-97.
24. Balogh A, Habre W. Neonatal Ventilation Strategies and Their Practical Application. In: Lerman J, editor. Neonatal Anesthesia. Cham: Springer International Publishing; 2023. p.213-26.
25. Fickert P, Stöger U, Fuchsbichler A, Moustafa T, Marschall HU, Weiglein AH, et al. A new xenobiotic-induced mouse model of sclerosing cholangitis and biliary fibrosis. Am J Pathol 2007;171:525-36.
26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
27. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. Radiol Med 2021;126:1296-311.
28. House MJ, Bangma SJ, Thomas M, Gan EK, Ayonrinde OT, Adams LA, et al. Texture-based classification of liver fibrosis using MRI. J Magn Reson Imaging 2015;41:322-8.
29. Yang W, Kim JE, Choi HC, Park MJ, Choi HY, Shin HS, et al. T2 mapping in gadoxetic acid-enhanced MRI: utility for predicting decompensation and death in cirrhosis. Eur Radiol 2021;31:8376-87.
30. Zhang D, Cao Y, Sun Y, Zhao X, Peng C, Zhao J, et al. Radiomics nomograms based on R2* mapping and clinical biomarkers for staging of liver fibrosis in patients with chronic hepatitis B: a single-center retrospective study. Eur Radiol 2023;33:1653-67.
31. Sack J, Nitsch J, Meine H, Kikinis R, Halle M, Rutherford A. Quantitative Analysis of Liver Disease Using MRI-Based Radiomic Features of the Liver and Spleen. J Imaging 2022;8.
32. Nowak S, Mesropyan N, Faron A, Block W, Reuter M, Attenberger UI, et al. Detection of liver cirrhosis in standard T2-weighted MRI using deep transfer learning. Eur Radiol 2021;31:8807-15.
33. Mai XF, Zhang H, Wang Y, Zhong WX, Zou LQ. Multiparametric MRI-based whole-liver radiomics for predicting early-stage liver fibrosis in rabbits. Br J Radiol 2024;97:964-70.

34. Chavhan GB, Babyn PS, Vasanawala SS. Abdominal MR Imaging in Children: Motion Compensation, Sequence Optimization, and Protocol Organization. RadioGraphics 2013;33:703-19.
35. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295:328-38.
36. Breiman L. Random Forests. Machine Learning 2001;45:5-32.
37. Zhao R, Zhao H, Ge Y-Q, Zhou F-F, Wang L-S, Yu H-Z, et al. Usefulness of Noncontrast MRI-Based Radiomics Combined Clinic Biomarkers in Stratification of Liver Fibrosis. Canadian Journal of Gastroenterology and Hepatology 2022;2022:2249447.
38. Trout AT, Sheridan RM, Serai SD, Xanthakos SA, Su W, Zhang B, et al. Diagnostic Performance of MR Elastography for Liver Fibrosis in Children and Young Adults with a Spectrum of Liver Diseases. Radiology 2018;287:824-32.
39. Zou L, Zhang H, Wang Q, Zhong W, Du Y, Liu H, et al. Simultaneous liver steatosis, fibrosis and iron deposition quantification with mDixon quant based on radiomics analysis in a rabbit model. Magn Reson Imaging 2022;94:36-42.

Appendices

1. Radiomics Feature Abbreviations

Table 3. Radiomics feature abbreviations

Abbreviation	Full Term
wavelet- HLH_ glrlm_ShortRunEmphasis	wavelet- High-Low-High_ Gray Level Run Length Matrix_ShortRunEmphasis
wavelet- LLL_firstorder_Minimum	wavelet- Low-Low-Low_firstorder_Minimum
wavelet- LLH_firstorder_Entropy	wavelet- Low-Low-High_firstorder_Entropy
wavelet-HLH_ glszm_HighGrayLevelZoneEmphasis	wavelet-High-Low-High_ Gray Level Size Zone Matrix_HighGrayLevelZoneEmphasis
wavelet-HLL_firstorder_Mean	wavelet-High-Low-Low_firstorder_Mean
wavelet-LLL_ glcm_Idn	wavelet-Low-Low-Low_ Gray Level Co-occurrence Matrix_Inverse Difference Normalized
wavelet-LHL_glcm_MCC	wavelet-Low-High-Low_Gray Level Co-occurrence Matrix_Maximal Correlation Coefficient
wavelet-HHH_glszm_SmallAreaHighGrayLevel	wavelet-High-High-High_Gray Level Size Zone

Emphasis	Matrix_SmallAreaHighGrayLevelEmphasis
wavelet-HLH_glcm_ClusterProminence	wavelet-High-Low-High_Gray Level Co-occurrence Matrix_ClusterProminence
wavelet-LHH_glszm_GrayLevelNonUniformityNormalized	wavelet-Low-High-High_Gray Level Size Zone Matrix_GrayLevelNonUniformityNormalized
wavelet-HLH_glcm_ClusterTendency	wavelet-High-Low-High_Gray Level Co-occurrence Matrix_ClusterTendency
wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis	wavelet-High-Low-High_Gray Level Run Length Matrix_ShortRunLowGrayLevelEmphasis
wavelet-HHH_glcm_DifferenceEntropy	wavelet-High-High-High_Gray Level Co-occurrence Matrix_DifferenceEntropy
wavelet-LLL_glcm_Correlation	wavelet-Low-Low-Low_Gray Level Co-occurrence Matrix_Correlation

Abstract in Korean

마우스 간 섬유화를 T2-weighted MRI Radiomics기반으로 한 머신러닝 모델로 진단하는 방법에 대한 연구

본 연구에서 T2-강조 영상에서 얻은 방사선 유래 특징(radiomics feature)들을 대상으로 개발된 머신러닝 모델이 간 손상 마우스 모델에서 다양한 정도의 섬유증을 평가할 수 있는지 연구하고자 했습니다.

본 연구에서, 6주령의 수컷 C57BL/6 마우스에게 3,5-디에톡시카보닐-1,4-디히드로콜리딘(DDC)이 포함된 특수 사료를 먹여 간의 섬유화를 유발하였습니다. T2-강조영상은 9.4T MRI에서 촬영하였습니다. 훈련세트는 DDC 사료를 1주에서 11주간 급여받은 39마리 마우스의 간 이미지가 포함되었고, DDC 사료 급여 전과 후에 영상은 촬영하였습니다. 테스트 세트는 일반 사료를 급여받은 10마리와 DDC 사료를 2주간 급여받은 10마리, 총 20마리 마우스로 구성되었습니다. 모든 MRI 사진은 방사선 유래 특징들로 분석되었습니다. 방사선 병리학적 상관관계를 확인하고자, 마우스들은 MRI촬영 후 희생하여 간조직 샘플을 얻었습니다. Sirus red 염색을 통해 콜라겐 면적 비율을 측정하여 간 섬유증을 정량화하였습니다. 섬유증 등급은 콜라겐 면적 비율에 따라 없음, 경도, 중증도로 분류되었습니다.

방사선 유래 특징 분석 및 머신러닝 모델 개발에는 상용 소프트웨어 (Syngo.via Frontier, version 1.3.0; Siemens Healthineers, Munich, Germany)를 사용했으며, 다중 선형 회귀 (Multiple linear regression MLR)과 랜덤 포레스트 (Random Forest, RF) 모델이 개발되었습니다. 간 섬유증 유무 평가에 대한 모델 성능은 수신자 조작 특성 곡선 (Receiver Operating Characteristic, ROC)을 통하여 평가되었습니다. 다양한 섬유증 정도는 콜라겐 면적비율 또는 섬유증 등급으로 표현하여, radiomics 변수들과의 상관관계를 Spearman 및 Kendall 방법으로 분석했습니다.

훈련 세트에는 총 50개의 MRI (기저 데이터: 섬유증 = 22:28)이 포함되었으며,

테스트 세트에는 총 18개의 MRI (대조군: 섬유증: 8:10)이 포함되었고, 총 845개의 방사선 유래 특징이 추출되었습니다. 이 중 가장 관련성 높은 10개의 특징이 선택되었습니다. 간 섬유증 진단에 대한 곡선 아래 면적(Area under the cureve, AUC)은 MLR 모델에서 0.991(민감도 97.5%, 특이도 94.2%)였으며, RF 모델에서 10-겹 교차 검증 평균 AUC는 0.945(민감도 88.9%, 특이도 85.5%)였습니다. 테스트 세트에 적용했을 때, MLR 모델의 AUC는 0.645(민감도 90%, 특이도 33.3%)였으며, RF 모델의 AUC는 0.817(민감도 90%, 특이도 41.7%)로 나타났습니다. MLR 및 RF 모델에서의 주요 방사선 유래 특징은 훈련 세트에서는 섬유증 등급과 상관관계를 보였고, 테스트세트에서는 간 CPA 및 섬유증 등급과 상관관계를 보였으며, 특히 RF 모델의 모든 변수가 섬유증 등급과 유의미한 상관관계를 보였습니다(모두 $P \leq 0.001$).

T2-가중 MRI의 방사선 유래 특징으로부터 개발된 RF 모델은 마우스 모델에서 간 섬유증을 진단하고 분류하는데 유용할 수 있을 것입니다.

핵심되는 말 : 간 섬유화, 자기공명영상, 방사선유래특징, 마우스