Article

Assessment of Liver Fibrosis Stage Using Integrative Analysis of Hepatic Heterogeneity and Nodularity in Routine MRI with FIB-4 Index as Reference Standard

Tae-Hoon Kim [1,2,]*, Chang-Won Jeong [1,2], Ji Eon Kim [1], Jin Woong Kim [3], Hoon Gil Jo [4], Youe Ree Kim [5], Young Hwan Lee [5] and Kwon-Ha Yoon [5,]*

[1] Medical Convergence Research Center, Wonkwang University, Iksan 54538, Korea; mediblue@wku.ac.kr (C.-W.J.); kakasky112@wku.ac.kr (J.E.K.)
[2] Smart Health IT Center, Wonkwang University Hospital, Iksan 54538, Korea
[3] Department of Radiology, Chosun University College of Medicine, Chosun University Hospital, Gwangju 61452, Korea; jw4249@hanmail.net
[4] Department of Hepatology & Gastroenterology, Wonkwang University Hospital, Iksan 54538, Korea; jojo4206@hanmail.net
[5] Department of Radiology, Wonkwang University School of Medicine, Wonkwang University Hospital, Iksan 54538, Korea; sweetynn@naver.com (Y.R.K.); yjyh@wku.ac.kr (Y.H.L.)
[*] Correspondence: tae_hoonkim@hanmail.net (T.-H.K.); khy1646@wku.ac.kr (K.-H.Y.); Tel.: +82-63-859-1921 (K.-H.Y.)

Abstract: Image-based quantitative methods for liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) provide helpful information for evaluating liver fibrosis; however, their combinations are not fully understood in liver diseases. We developed an integrated software for assessing L_{Het} and L_{Nod} and compared L_{Het} and L_{Nod} according to fibrosis stages in chronic liver disease (CLD). Overall, 111 CLD patients and 16 subjects with suspected liver disease who underwent liver biopsy were enrolled. The procedures for quantifying L_{Het} and L_{Nod} were bias correction, contour detection, liver segmentation, and L_{Het} and L_{Nod} measurements. L_{Het} and L_{Nod} scores among fibrosis stages (F0–F3) were compared using ANOVA with Tukey's test. Diagnostic accuracy was determined by calculating the area under the receiver operating characteristics (AUROC) curve. The mean L_{Het} scores of F0, F1, F2, and F3 were 3.49 ± 0.34, 5.52 ± 0.88, 6.80 ± 0.97, and 7.56 ± 1.79, respectively ($p < 0.001$). The mean L_{Nod} scores of F0, F1, F2, and F3 were 0.84 ± 0.06, 0.91 ± 0.04, 1.09 ± 0.08, and 1.15 ± 0.14, respectively ($p < 0.001$). The combined $L_{Het} \times L_{Nod}$ scores of F0, F1, F2, and F3 were 2.96 ± 0.46, 5.01 ± 0.91, 7.30 ± 0.89, and 8.48 ± 1.34, respectively ($p < 0.001$). The AUROCs of L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ for differentiating F1 vs. F2 and F2 vs. F3 were 0.845, 0.958, and 0.954; and 0.619, 0.689, and 0.761, respectively. The combination of L_{Het} and L_{Nod} scores derived from routine MR images allows better differential diagnosis of fibrosis subgroups in CLD.

Keywords: chronic liver disease (CLD); heterogeneity; nodularity; integrative

1. Introduction

Liver fibrosis is a hallmark of chronic liver disease (CLD) characterized by excessive accumulation of extracellular matrix proteins responsible for fibrogenesis [1,2]. Liver fibrosis may progress to cirrhosis, the end stage, which constitutes the most important risk factor for developing hepatocellular carcinoma (HCC) [3]. Liver biopsy has been regarded as the reference diagnostic method for evaluating the stage of liver fibrosis in CLD [4]. However, this method has well-known weaknesses including sampling errors, low patient acceptance, and complications such as pain, bleeding, infection, and rarely death [5]. Moreover, in the CLD patients with an initial diagnosis of early stage fibrosis (compensated liver) or cirrhosis (decompensated liver as end stage), it is difficult to accurately predict hepatic compensation or decompensation using noninvasive methods [6]. Thus, there is an

unmet need for widely applicable noninvasive methods to diagnose fibrosis and advanced cirrhosis and to predict future risk of hepatic decompensation.

Recently, there have been considerable efforts to develop imaging techniques and quantification programs for diagnosing and staging liver fibrosis. There are several methods including contrast-enhanced imaging, elastography, image-based morphologic analysis, and texture analysis [7–9]. Among them, image-based morphologic analysis includes quantification of parenchymal heterogeneity, contour change by the liver nodules, atrophic or necrotic change, edge blunting, fissural widening, and so on [10]. Several quantification software program have been introduced for assessing the findings of liver fibrosis and cirrhosis on medical images [10–12]. Heterogeneity quantification programs using coefficient of variation (CV) maps can help to assess the severity of fibrosis and cirrhosis in patients with chronic hepatitis B [10,11]. Several studies [10,11] reported that the area under the receiver operating characteristic curves (AUROCs) on the magnetic resonance (MR) CV map scores were 0.875 for discriminating significant fibrosis (≥fibrosis grade 2; F2) in chronic hepatitis B and 0.788 for the presence of HCC in patients with liver cirrhosis (F4). Moreover, liver surface nodularity (LSN) can be useful for differentiating the severity of fibrosis. A study [12] reported that the AUROC was 0.788 for discriminating significant fibrosis (≥F2) in nonalcoholic fatty liver disease (NAFLD). Other studies showed that the AUROCs of computed tomography (CT) LSN scores were 0.902 and 0.959 for discriminating significant fibrosis (≥F2) and cirrhosis (F4), respectively [13,14]. A comparative study using MR LSN and MR elastography demonstrated that the AUROCs for diagnosing significant fibrosis (≥F2) were 0.61 for MR LSN and 0.87 for MR elastography [15]. However, these studies used a single measurement, either liver heterogeneity (L_{Het}) or nodularity (L_{Nod}), for evaluating liver fibrosis; the method was not clearly distinguished among each fibrosis stage. Taking all of these findings into consideration, the computer-aided L_{Het} and L_{Nod} scores can provide important information for differential diagnosis of hepatic fibrosis. The integrative analysis of L_{Het} or L_{Nod} for evaluating liver fibrosis is not yet fully understood.

For this study, we developed an integrated semiautomated quantification software for assessing L_{Het} and L_{Nod} and compared them across fibrosis stages in CLD.

2. Materials and Methods

2.1. Ethics Statment

The study protocol was approved as retrospective research (WKUH-2017-03-026) by the institutional review board (IRB) of University Hospital. Written informed consent was waived by the University Hospital IRB committee due to the use of anonymous archival data including MRI data (radiology_common data model: R_CDM, version 2.0.0) and electronic health records (observational medical outcomes partnership-CDM: OMOP-CDM, version 5.3) for the application of developed software. This study was conducted in accordance with the Helsinki Declaration and Good Clinical Practice guidelines.

2.2. Subject Population

Among the 1654 consecutive patients who underwent radiological examination at our institution from April 2003 to December 2018, patients 20 years or older who underwent abdominal MRI at 3.0 T and who had available serologic tests within five months of MRI were retrospectively identified. Of 121 eligible patients, 10 were excluded due to the absence of MR images for liver protocols and the absence of medical records for CLD (Figure 1). The inclusion criteria of CLD patients were the elevation of liver function enzymes, alanine transaminase (ALT), and aspartate transaminase (AST) and the absence of liver cirrhosis (F4) [14]. The CLD subgroups were divided into three fibrosis groups according to the serum biomarkers of fibrosis-4 index (FIB-4, Equation 1) values as follows: F1, mild fibrosis group < 1.45; F2, significant fibrosis group 1.45–3.25; and F3, advanced fibrosis group > 3.25 (Table 1). Finally, the subgroups consisted of 9 F1 (mean age; 50.3 ± 14.9 years), 57 F2 (60.3 ± 12.1 years), and 45 F3 (mean 64.8 ± 13.6 years). This study included 16 subjects (35.0 ± 15.5 years) with suspected liver disease who underwent the needle biopsy for

comparison (Figure 1). These individuals had symptoms of fatigue and inactivity. They had abnormal liver function tests, and there was no histological evidence for liver fibrosis and advanced cirrhosis (no fibrosis group, F0).

$$\text{FIB} - 4 = \frac{(\text{Age(year)} \times \text{AST(U/L)})}{(\text{platelet count } (10^9/\text{L}) \times \text{ square root of ALT (U/L)})} \tag{1}$$

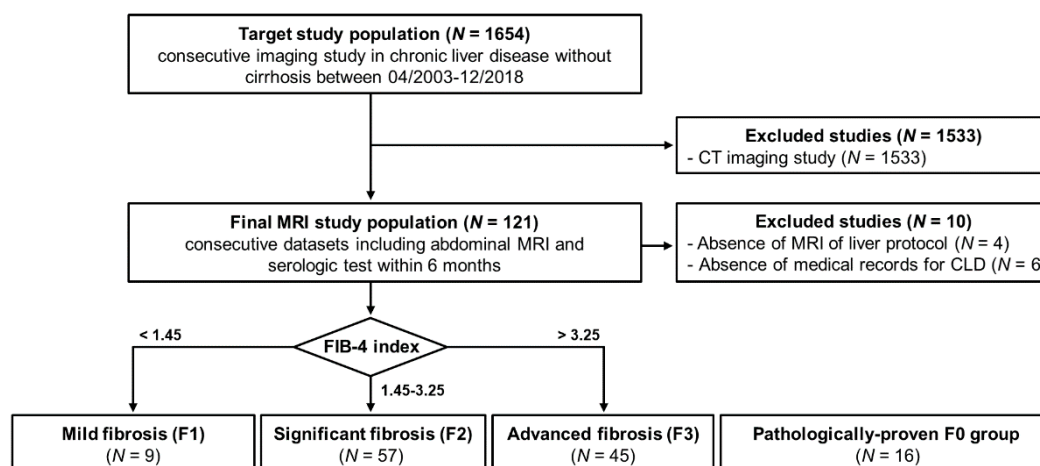The upper limit of normal AST was 35 in this study.

Figure 1. Flowchart for inclusion of chronic liver disease. FIB-4 index = the serum biomarkers of fibrosis-4 index.

Table 1. Etiology of study population and clinical data of serum biochemistry in subgroups of chronic liver disease.

|	F0 (*n* = 16)	F1 (*n* = 9)	F2 (*n* = 57)	F3 (*n* = 45)	*p*-Value * †	Multiple Comparisons		
						[a] F1 vs. F2	[b] F1 vs. F3	[c] F2 vs. F3
Etiology *								
Unknown	16 (100)	-	-	-				
Hepatitis B	-	6 (67)	37 (65)	27 (60)	* 0.570			
Hepatitis C	-	0 (0)	5 (9)	4 (9)				
Coinfection	-	0 (0)	2 (4)	0 (0)				
Alcohol	-	1 (11)	5 (9)	6 (13)				
NAFLD	-	0 (0)	6 (10)	5 (11)				
Autoimmune	-	2 (22)	2 (3)	3 (7)				
Serum biochemistry †								
Age (year)	35.0 ± 15.5	50.3 ± 14.9	60.3 ± 12.1	64.8 ± 13.6	† 0.008 [b]	0.086	0.008	0.191
BMI	24.8 ± 3.2	24.8 ± 1.7	24.6 ± 2.0	25.3 ± 2.9	† 0.211	0.953	0.777	0.184
Albumin (g/dL)	4.36 ± 0.07	3.80 ± 0.31	4.08 ± 0.43	4.30 ± 0.69	† 0.873	0.962	0.888	0.928
ALP (IU/L)	73.9 ± 8.3	279.2 ± 36.7	285.0 ± 17.9	459.8 ± 53.6	† 0.002 [c]	0.998	0.125	0.002
ALT (IU/L)	18.8 ± 2.1	31.7 ± 8.7	37.5 ± 3.8	74.2 ± 25.6	† 0.224	0.989	0.552	0.230
AST (IU/L)	21.0 ± 1.4	26.0 ± 4.4	38.7 ± 3.6	109.1 ± 36.0	† 0.058	0.972	0.311	0.064
Bilirubin (mg/dL)	0.64 ± 0.09	0.73 ± 0.09	0.78 ± 0.09	2.66 ± 0.90	† 0.058	1.000	0.441	0.052
GGT (IU/L)	38.5 ± 10.9	83.4 ± 29.9	63.2 ± 7.2	184.9 ± 30.0	† 0.001 [c]	0.946	0.266	<0.001
Platelet count (10³/µL)	243.5 ± 8.2	223.9 ± 17.4	172.3 ± 7.0	135.4 ± 6.9	<† 0.001 [a b c]	0.014	<0.001	0.001

Groups: F1, mild fibrosis group; F2, significant fibrosis group; and F3, advanced fibrosis group. Abbreviations: ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; GGT: gamma-glutamyl transpeptidase; NAFLD; nonalcoholic fatty liver disease. Etiology data are presented as the number of patients. The value in parenthesis indicates the percentage as the number of patients/total number of patients ×100. Serum biochemistry data are presented as mean ± SEM. Reference ranges for serum biochemistry are as follows: 3.4–5.4 g/dL for albumin, 44–147 IU/L for ALP, 0–35 IU/L for ALT, 0–35 IU/L for AST, 0.1–1.2 mg/dL for total bilirubin, 8–38 IU/L for GGT, and 120–473 × 10³/µL for platelet count. * The difference among fibrosis groups in etiology data was analyzed by Pearson's chi-square test. † The difference among three fibrosis groups was analyzed by one-way ANOVA with Tukey post hoc test as follows: [a] F1 vs. F2, [b] F1 vs. F3, and [c] F2 vs. F3.

2.3. Acquisition of MRI

All MRI examinations were performed on a 3T MRI scanner (Achieva; Philips, Netherlands) with a 32-channel receiver body matrix coil. The T1WI were acquired with three-dimensional T1 high-resolution isotropic volume excitation (THRIVE) pulse sequence: TR/TE = 4.2/1.97 msec; field of view (FOV) = $38 \times 38 \times 14$ cm^3, matrix size = 512×512, number of excitation (NEX) = 2, slice thickness = $0.74 \times 0.74 \times 2.0$ mm^3, and number of slices = 70.

2.4. Software for Quantification of Liver Heterogeneity and Nodularity

L$_{Het}$ and L$_{Nod}$ quantification software (customized software; named WALTS) was coded by MATLAB (MathWorks, Natick, Massachusetts). Wonkwang Abdomen and Liver Total Solution (WALTS) software is a customized semiautomated postprocessing program that operates on Windows platform (client version: XP or higher; Microsoft, Redmond, WA). We used WALTS to process the MR images in the DICOM (Digital Imaging and Communications in Medicine) format to generate the L$_{Het}$ and L$_{Nod}$ scores using a previously described procedure [13,14,16]. Figure 2 shows the GUI of WALTS and a simple flowchart showing the development of an algorithm for qualitative and quantitative analysis. The procedures for quantifying L$_{Het}$ and L$_{Nod}$ scores were as follows: bias correction of field uniformity, liver contour detection for drawing the liver reference line, liver segmentation, and L$_{Het}$ and L$_{Nod}$ measurements.
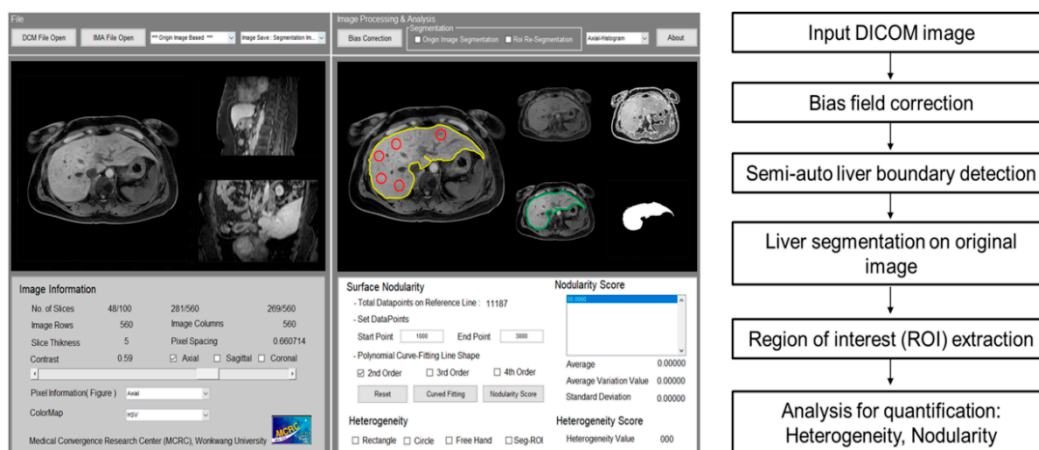
Figure 2. Graphical user interface (GUI) of a tailor-made quantification software for assessing liver heterogeneity (L$_{Het}$) and nodularity (L$_{Nod}$) (**left** side) and flowchart for quantifying liver fibrosis (**right** side).
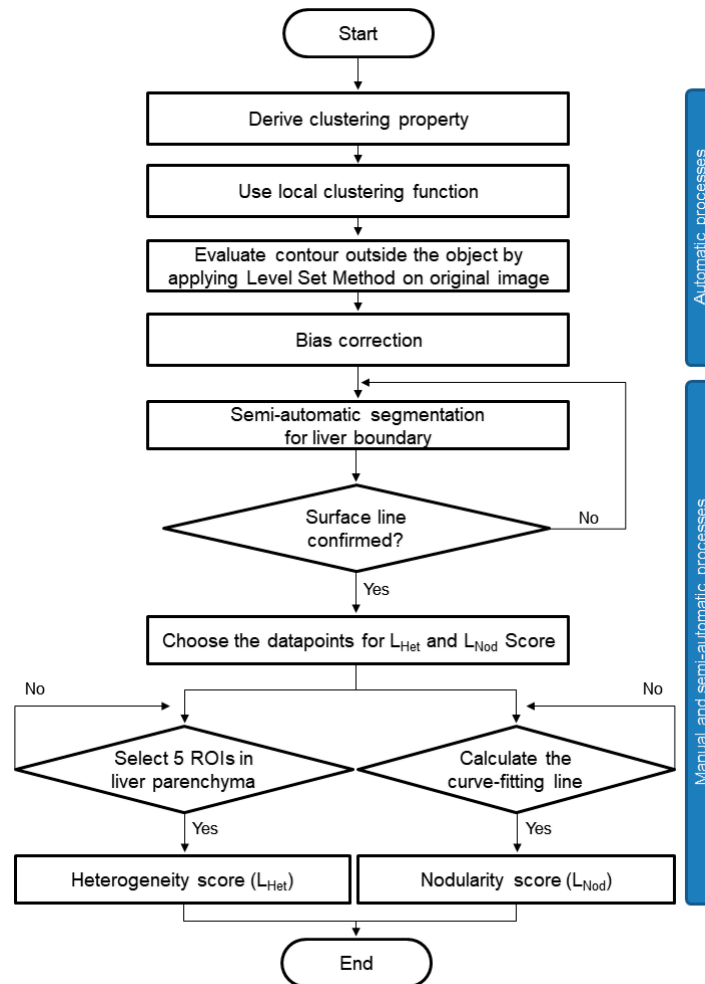
2.5. Data Processing and Quantification of MRI in CLD

Figure 3 shows the overall image postprocessing procedures for hepatic heterogeneity (L$_{Het}$) and nodularity (L$_{Nod}$) quantification using MR images. To automatically detect the liver's contour, we used a novel region-based method for liver segmentation as a level set method, which provided the local clustering criterion function with correction with intensity inhomogeneities (Figure 4B) [17]. The boundary detection and segmentation techniques maximize the local intensity clustering property and minimize the energy formulation to determine and exclude any existing signal outliers caused by generated systematic artifacts as described in previous studies [12,18]. The contour line in the selected slice of the liver was produced after bias correction. Following preprocessing of MRI data, the liver surface line for L$_{Het}$ and L$_{Nod}$ quantification was extracted as a reference line, and the extracted line was confirmed by two abdominal radiologists (with 29 and 8 years of experience in abdominal imaging) (Figure 4C). Five circular regions of interest (ROIs; each 40 pixels) for L$_{Het}$ measurement were drawn on the liver parenchyma. In all subjects, ROIs were placed on the liver parenchymal areas with no overlap over large vessels or focal

lesions. The L_{Het} score and L_{Het} map were calculated using the following Equations (2) and (3):

$$L_{Het} \text{ score} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100 \tag{2}$$

$$L_{Het} \text{ map} = \frac{CV}{\text{Pixel Value}} \times 100 \tag{3}$$

Figure 3. Overall image postprocessing procedures for hepatic heterogeneity (L_{Het}) and nodularity (L_{Nod}) quantification using MR images.

To measure the L_{Nod} score, the liver parenchyma within the confirmed liver boundary line was used for the multipolynomial curve fitting analysis. ROIs for L_{Nod} measurement were selected along the contour of the liver (Figure 4G). The user would insert an L_{Nod} ROI range across the data points of the liver surface line. After input of an L_{Nod} ROI range, a smooth curve-fitting line (polynomial line shape) was generated on a selected ROI dataset (Figure 4H). Finally, the difference between the liver surface line and the new polynomial curve-fitting line (one of second-, third-, and fourth-order line shape) was evaluated on a pixel-by-pixel basis. The difference value was squared; then it was used to calculate the mean, variation, and standard deviation (SD). The final L_{Nod} score in an individual subject was calculated as the mean L_{Nod} obtained from the measurements on ROIs. In addition, a combined score derived from L_{Het} and L_{Nod} was calculated as the multiplication of both scores (= $L_{Het} \times L_{Nod}$).
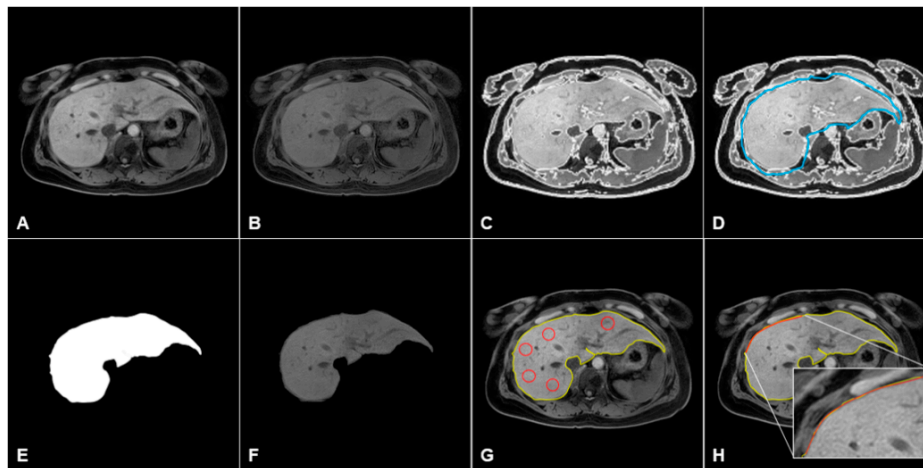
Figure 4. Representative postprocessing images for liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) quantification in a CLD patient: (**A**) raw MRI data, (**B**) bias field correction, (**C**) liver contour detection, (**D**) drawn region of interest (ROI, blue line), (**E**) segmented liver binary image, (**F**) liver ROI extraction, (**G**) L_{Het} and L_{Nod} quantification (L_{Het} ROIs = 5 red circles; liver surface line for L_{Nod} = yellow line), and (**H**) curve-fitting line (red line) for L_{Nod} quantification along the liver surface line.

All MR studies reviewed standard picture archiving and communication system (PACS) stations and software with standard window settings. The liver MR images in each CLD patient were assessed by two abdominal radiologists, who were blinded to clinical outcome, using WALTS software. After opening DICOM images on the software, they selected image slices at the level of the hepatic hilum. Then, bias correction and segmentation were performed on the selected images. For L_{Het} measurement in each subject, five circular ROIs were manually drawn in the liver parenchyma (Figure 5); these areas contained the liver parenchyma, avoiding the perceivable bile duct, major intrahepatic vessels, subcapsular area, and focal lesions such as cysts or benign and malignant tumors [11]. The final L_{Het} and L_{Nod} scores in each fibrosis group were calculated as an averaged score obtained by reporting scores of each observer (observer A: YRK, observer B: YHL) for AUROC differential diagnosis according to fibrosis stages.
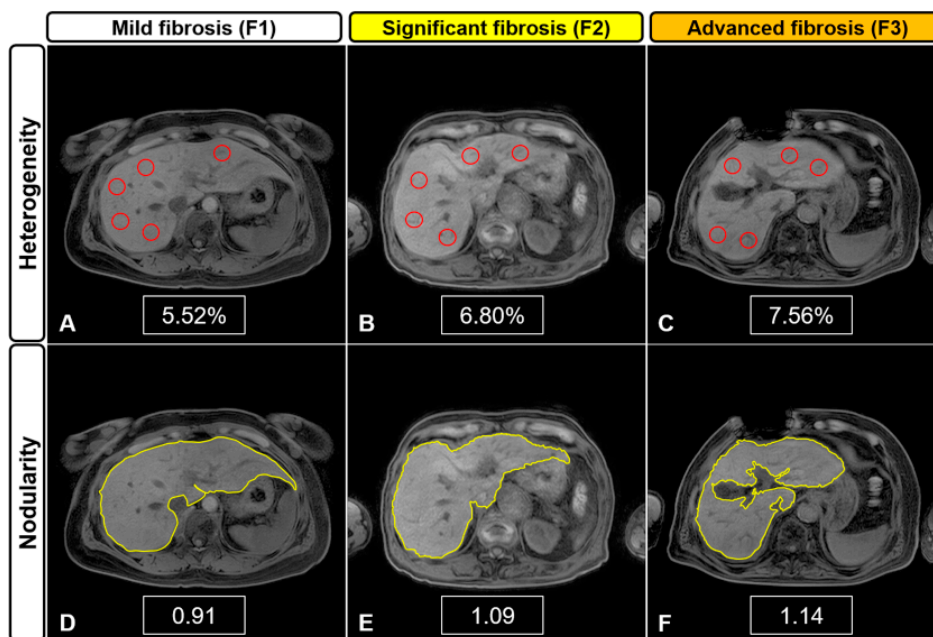
Figure 5. Representative MR images for liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) quantification according to Figure 1. F1 (**A**,**D**), F2 (**B**,**E**), and F3 (**C**,**F**). Higher heterogeneity and nodularity scores are seen with increased severity in the fibrotic liver.

All the measurements on selected MR images (the level of the hepatic hilum) were repeated two weeks after the first measurement was obtained to evaluate intraobserver agreement. Furthermore, to determine interobserver agreement, both radiologists independently measured L_{Het} and L_{Nod} scores on selected images. The intra- and interobserver variability in the L_{Het} and L_{Nod} measurements was assessed. The overall scores of L_{Het} and L_{Nod} were calculated as the mean score of the three measurements taken for each patient. The WALTS program for L_{Het} and L_{Nod} quantification used MR images in DICOM format to generate the L_{Het} and L_{Nod} scores. The technical details for obtaining the L_{Het} and L_{Nod} measurements were described in recent papers [10–13]. Measurements of at least three and/or four ROIs were performed for each subject. Final L_{Het} and L_{Nod} scores were calculated by the program as a mean value of the individual measurements, with a higher L_{Het} (or L_{Nod}) score indicating a higher degree of parenchymal heterogeneity (or nodularity). Figure 4 shows the representative images in L_{Het} and L_{Nod} measurements in an axial MR image.

2.6. Statistical Analysis

The L_{Het} and L_{Nod} scores among three different stages of fibrosis in CLD were compared using the SPSS version 20.0 program (SPSS Inc., Chicago, IL, USA). The variation in L_{Het} and L_{Nod} scores was analyzed using analysis of variance (ANOVA) with Tukey's post hoc test. The difference between CLD patients and the control group was analyzed using the independent two-sample *t*-test. Intraobserver agreement (between measurements from the same observer) was calculated as the mean coefficient of variance (%) for the variability of L_{Het} and L_{Nod} scores taken by the same single observer [19]. Also, the variation between the scores of both observers was analyzed with paired *t*-test. Intraobserver agreement was performed on the basis of the intraclass correlation coefficient (ICC) between the L_{Het} and L_{Nod} scores. The ICCs were indicated based on the levels of reliability as follows [20]: poor (<0.4), moderate (0.4–0.6), good (0.6–0.8), and excellent (0.8–1.0).

The diagnostic performance of L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores according to fibrosis stages was evaluated with ROC curve analysis including the AUROC, sensitivity, and specificity. Two-sided *p*-values less than 0.05 were considered to denote statistical significance in all tests.

3. Results

3.1. Patient Characteristics in Fibrosis Stages

Figure 1 shows the inclusion flowchart for the study population. The etiology of liver fibrosis in CLD and the average enzyme levels according to fibrosis stages are listed in Table 1. The serum biochemistry showed significant difference among three groups in the levels of alkaline phosphatase (ALP, $p = 0.002$), glutamyl transpeptidase (GGT, $p = 0.001$), and platelet count ($p < 0.001$). However, there was no significant difference among the fibrosis groups as follows: albumin ($p = 0.873$), alanine aminotransferase (ALT, $p = 0.224$), aspartate aminotransferase (AST, $p = 0.058$), and bilirubin ($p = 0.058$).

3.2. Liver Heterogeneity and Nodularity Measurements in CLD

MRI data for 111 CLD patients and 16 subjects with suspected liver disease were analyzed with developed WALTS software. Figure 3 shows the detailed image processing procedures. Figure 4 shows a representative MR image of a patient with CLD (Figure 4A), bias-corrected MR image (Figure 4B), liver contour detection (Figure 4C), region of interest (ROI) drawing on MR image (Figure 4D), binary image of ROI for liver segmentation (Figure 4E), segmented liver MRI (Figure 4F), final reference line (Figure 4G), and curve-fitting lines (Figure 4H) for L_{Het} and L_{Nod} measurements.

Figure 5 shows the representative quantification images in each fibrosis stage (F1–F3), and the results are summarized in Table 2. Mean L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores in CLD were higher than those in the F0 group ($p < 0.001$). Mean L_{Het}, L_{Nod}, and combined $L_{Het} \times L_{Nod}$ scores were significantly different between fibrosis stages (F1–F3) (ANOVA;

$p < 0.001$). In multiple comparisons, L_{Het} scores were different from each other (F1 vs. F2, $p = 0.032$; F1 vs. F3, $p < 0.001$; and F2 vs. F3, $p = 0.008$). Additionally, L_{Nod} scores were different in F1 vs. F2, $p < 0.001$; F1 vs. F3, $p < 0.001$; and F2 vs. F3, $p = 0.022$. The combined scores were significantly different in the contrasts of F1 vs. F2, F1 vs. F3, and F2 vs. F3 ($p < 0.001$). There were significant differences among fibrosis stages based on quantitative L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores.

Table 2. Comparison of liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) scores according to fibrosis stages (F).

| | F0 ($n = 16$) | F1 ($n = 9$) | F2 ($n = 57$) | F3 ($n = 45$) | p-Value * | Multiple Comparisons | | |
						[a] F1 vs. F2	[b] F1 vs. F3	[c] F2 vs. F3
Heterogeneity (L_{Het}, %)	3.49 ± 0.34	5.52 ± 0.88	6.74 ± 0.89	7.56 ± 1.79	<0.001 [a b c]	0.032	<0.001	0.008
Nodularity (L_{Nod})	0.84 ± 0.06	0.91 ± 0.04	1.09 ± 0.08	1.14 ± 0.14	<0.001 [a b c]	<0.001	<0.001	0.022
$L_{Het} \times L_{Nod}$ score	2.96 ± 0.46	5.01 ± 0.91	7.30 ± 0.89	8.48 ± 1.34	<0.001 [a b c]	<0.001	<0.001	<0.001

Data are presented as mean $\pm$ SD. The final L_{Het} and L_{Nod} scores in each fibrosis group were calculated as an averaged score obtained by reporting scores of two observers (observer A: Y.R.K., observer B: Y.H.L.) for AUROC differential diagnosis according to fibrosis stages. The average scores in F0 group were used as the reference ranges. * The difference among the three fibrosis groups was analyzed by one-way ANOVA with Tukey post hoc test as follows: [a] F1 vs. F2, [b] F1 vs. F3, and [c] F2 vs. F3.

3.3. ROC Analysis for Differential Diagnosis According to Fibrosis Stages

Figure 6 shows AUROC curves of L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores for the discrimination of fibrosis stages, and the results are summarized in Table 3. The AUROCs of L_{Het} scores were 0.845 for F1 vs. F2 (95%CI, 0.722–0.968; $p = 0.001$), 0.619 for F2 vs. F3 (95%CI, 0.509–0.730; $p = 0.040$), 0.944 for F0–1 vs. F2 (95%CI, 0.890–0.997; $p < 0.001$), and 0.963 for F0–1 vs. F3 (95%CI, 0.918–1.000; $p < 0.001$). The diagnostic accuracy of F1 vs. F2 had 0.714 sensitivity and 0.778 specificity at a cut-off L_{Het} score of 6.22; F2 vs. F3 had 0.533 sensitivity and 0.536 specificity at a cut-off L_{Het} score of 7.02; F0–1 vs. F2 had 0.875 sensitivity and 0.880 specificity at a cut-off L_{Het} score of 5.68; and F0–1 vs. F3 had 0.867 sensitivity and 0.880 specificity at a cut-off L_{Het} score of 6.17.

Table 3. Receiver operator curve analysis for diagnosing fibrosis stage using liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) scores.

Comparison	Threshold Value	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	DA (%)	AUROC	p-Value
L_{Het} score								
F1 vs. F2	6.22	71.4 (41/57)	77.8 (7/9)	95.3 (41/43)	30.4 (7/23)	72.7 (48/66)	0.845	0.001
F2 vs. F3	7.02	53.3 (24/45)	53.6 (30/57)	47.1 (24/51)	58.8 (30/51)	52.9 (54/102)	0.619	0.040
F0–1 vs. F2	5.68	87.5 (50/57)	88.0 (22/25)	94.3 (50/53)	75.9 (22/29)	87.8 (72/82)	0.944	<0.001
F0–1 vs. F3	6.17	86.7 (39/45)	88.0 (22/25)	92.9 (39/42)	78.6 (22/28)	87.1 (61/70)	0.963	<0.001
L_{Nod} score								
F1 vs. F2	0.97	94.6 (54/57)	100.0 (9/9)	100.0 (54/54)	75.0 (9/12)	95.4 (63/66)	0.958	<0.001
F2 vs. F3	1.12	64.4 (29/45)	64.9 (37/57)	59.2 (29/49)	69.8 (37/53)	64.7 (66/102)	0.689	0.001
F0–1 vs. F2	0.95	94.6 (54/57)	96.0 (24/25)	98.2 (54/55)	88.9 (24/27)	95.1 (78/82)	0.969	<0.001
F0–1 vs. F3	0.94	91.1 (41/45)	92.0 (23/25)	95.3 (41/43)	85.2 (23/27)	91.4 (64/70)	0.936	<0.001
$L_{Het} \times L_{Nod}$ score								
F1 vs. F2	6.10	89.5 (51/57)	88.9 (8/9)	98.1 (51/52)	57.1 (8/14)	89.4 (59/66)	0.954	<0.001
F2 vs. F3	7.85	66.7 (30/45)	68.4 (39/57)	62.5 (30/48)	72.2 (39/54)	67.6 (69/102)	0.761	<0.001
F0–1 vs. F2	5.98	91.1 (52/57)	92.0 (23/25)	96.3 (52/54)	82.1 (23/28)	91.5 (75/82)	0.984	<0.001
F0–1 vs. F3	6.70	97.8 (44/45)	100.0 (25/25)	100.0 (44/44)	96.2 (25/26)	98.6 (69/70)	0.999	<0.001

Note—Data in parentheses are raw data used to calculate percentages. AUROC: area under the receiver operator curve; DA: diagnostic accuracy = (TP + TN)/(TP + FP + TN + FN); F: fibrosis stages; FN: false negative; FP: false positive; NPV: negative predictive value; PPV: positive predictive value; TN: true negative; TP: true positive.

The AUROCs of L_{Nod} scores were 0.958 for F1 vs. F2 (95%CI, 0.890–1.000; $p < 0.001$) and 0.689 for F2 vs. F3 (95%CI, 0.577–0.801; $p = 0.001$), 0.969 for F0–1 vs. F2 (95%CI, 0.931–1.000; $p < 0.001$) and 0.936 for F0–1 vs. F3 (95%CI, 0.872–0.999; $p < 0.001$). The diagnostic accuracy of F1 vs. F2 had 0.947 sensitivity and 1.000 specificity at a cut-off L_{Nod} score of 0.97; F2 vs. F3 had 0.644 sensitivity and 0.649 specificity at a cut-off L_{Nod} score of

1.12; F0–1 vs. F2 had 0.946 sensitivity and 0.960 specificity at a cut-off L_{Het} score of 0.95; and F0–1 vs. F3 had 0.911 sensitivity and 0.920 specificity at a cut-off L_{Het} score of 0.94.
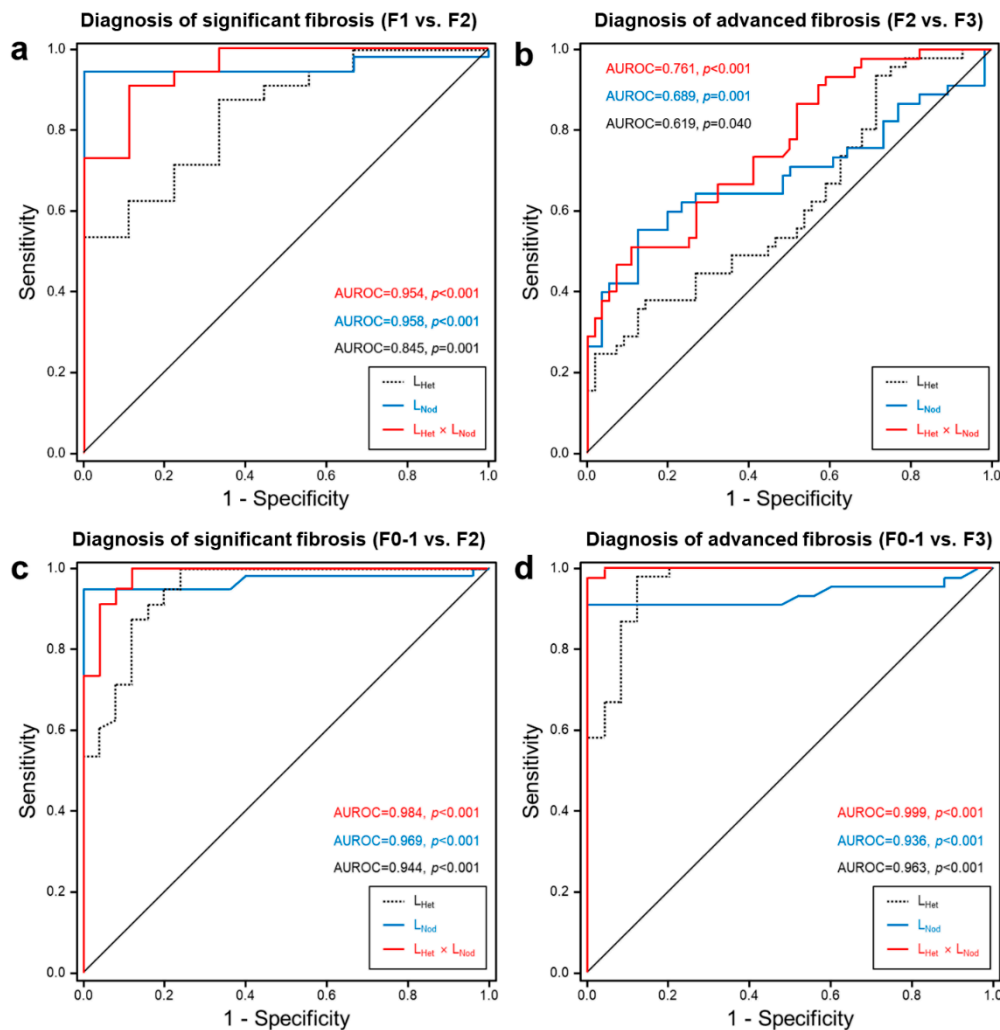
Figure 6. AUROC curves of liver heterogeneity (L_{Het}, dotted line), nodularity (L_{Nod}, blue line), and combined $L_{Het} \times L_{Nod}$ (red line) scores for the differentiation of fibrosis stages: (**a**) diagnosis of significant fibrosis (F1 vs. F2), (**b**) diagnosis of advanced fibrosis (F2 vs. F3), (**c**) diagnosis of significant fibrosis (combined F0–1 vs. F2), and (**d**) diagnosis of advanced fibrosis (combined F0–1 vs. F3). AUROC = area under the receiver operating characteristic curve.

The AUROCs of $L_{Het} \times L_{Nod}$ scores were 0.954 for F1 vs. F2 (95%CI, 0.884–1.000; $p < 0.001$) and 0.761 for F2 vs. F3 (95%CI, 0.669–0.853; $p < 0.001$), 0.984 for F0–1 vs. F2 (95%CI, 0.960–1.000; $p < 0.001$) and 0.999 for F0–1 vs. F3 (95%CI, 0.996–1.000; $p < 0.001$). The diagnostic accuracy of F1 vs. F2 had 0.895 sensitivity and 0.889 specificity at a cut-off $L_{Het} \times L_{Nod}$ score of 6.10; F2 vs. F3 had 0.667 sensitivity and 0.684 specificity at a cut-off $L_{Het} \times L_{Nod}$ score of 7.85; F0–1 vs. F2 had 0.911 sensitivity and 0.920 specificity at a cut-off $L_{Het} \times L_{Nod}$ score of 5.98; and F0–1 vs. F3 had 0.978 sensitivity and 1.000 specificity at a cut-off $L_{Het} \times L_{Nod}$ score of 6.70.

Diagnostic accuracy (%, number of subject) of L_{Het} scores in discriminating the FIB-4 stages was 72.7% (48/66) for F1 vs. F2, 52.9% (54/102) for F2 vs. F3, 87.8% (72/82) for F0–1 vs. F2 and 87.1% (61/70) for F0–1 vs. F3. Diagnostic accuracy of L_{Nod} scores was 95.4% (63/66) for F1 vs. F2, 64.7% (66/102) for F2 vs. F3, 95.1% (78/82) for F0–1 vs. F2, and 91.4% (64/70) for F0–1 vs. F3. Diagnostic accuracy of $L_{Het} \times L_{Nod}$ scores was 89.4% (59/66) for F1 vs. F2, 67.6% (69/102) for F2 vs. F3, 91.5% (75/82) for F0–1 vs. F2, and 98.6% (69/70) for F0–1 vs. F3.

3.4. Intraobserver and Interobserver Agreement

The intra- and interobserver variabilities of L_{Het} and L_{Nod} scores from two observers are summarized in Table 4. In intraobserver variability, the mean coefficient of variation within the same observer was in the range of 13–28% for L_{Het} measurements and in the range of 4–14% for L_{Nod} measurements. Also, there was no significant difference between the averaged L_{Het} and L_{Nod} values of the two observers ($p > 0.05$). In interobserver variability, ICCs were higher than 0.6, indicating good reliability. The ICCs (range: 0.601–0.852) were 0.718 for L_{Het} measurements and 0.832 for L_{Nod} measurements. The overall L_{Het} and L_{Nod} measurements of both observers showed good agreement ($p < 0.05$).

Table 4. Intra- and interobserver variability in liver heterogeneity (L_{Het}) and nodularity (L_{Nod}) measurements according to fibrosis stages.

	Observer A	Observer B	*p*-Value *	Intrarater Reliability (ICC) †	95% CI		*p*-Value †
					Lower Bound	Upper Bound	
L_{Het}score							
Overall (*n* = 111)	7.06 ± 1.68 (24)	6.94 ± 1.64 (24)	0.140	0.718	0.589	0.806	<0.001
F1 (*n* = 9)	5.45 ± 1.08 (20)	5.58 ± 0.85 (15)	0640	0.782	0.035	0.951	0.023
F2 (*n* = 57)	6.85 ± 1.15 (17)	6.75 ± 1.12 (13)	0.535	0.624	0.362	0.779	<0.001
F3 (*n* = 45)	7.65 ± 2.05 (27)	7.47 ± 2.08 (28)	0.558	0.679	0.417	0.824	<0.001
L_{Nod}score							
Overall (*n* = 111)	1.10 ± 0.12 (11)	1.10 ± 0.14 (13)	0.459	0.832	0.755	0.885	<0.001
F1 (*n* = 9)	0.91 ± 0.04 (4)	0.90 ± 0.06 (7)	0.869	0.768	−0.029	0.948	0.027
F2 (*n* = 57)	1.09 ± 0.09 (8)	1.09 ± 0.10 (9)	0.905	0.601	0.322	0.765	<0.001
F3 (*n* = 45)	1.14 ± 0.13 (11)	1.15 ± 0.16 (14)	0.596	0.852	0.731	0.919	<0.001

Abbreviations: ICC: intraclass correlation coefficient; CI: confidence interval. L_{Het} and L_{Nod} scores of each observer are presented as means ± SD (mean coefficient of variance, %). * The differences between both observers in L_{Het} and L_{Nod} scores were assessed by the paired *t*-test. † The intrarater reliability between both observers was assessed by the intraclass correlation (ICC) test.

4. Discussion

This study developed an integrated system (semiautomated WALTS software) for evaluating L_{Het} and L_{Nod} in liver diseases and compared the subgroups of fibrosis stages in CLD patients obtained from retrospective routine MRI datasets with serologic laboratory tests. In this study, liver MR images with three-dimensional THRIVE pulse sequence (routine T1 MR images) demonstrated acceptable accuracy in diagnosing fibrosis stages of CLD patients. L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores in CLD patients were higher than those in the F0 group. The AUROC-based differentiation in comparison of F1 vs. F2 fibrosis was significant as L_{Het} 0.845, L_{Nod} 0.958, and $L_{Het} \times L_{Nod}$ 0.954. Moreover, the AUROC in F2 vs. F3 was significant as L_{Het} 0.619, L_{Nod} 0.689, and $L_{Het} \times L_{Nod}$ 0.761. Smith et al. [13] and Pickhardt et al. [14] reported that the L_{Nod} diagnostic accuracy using CT images is excellent for predicting fibrosis (≥F2) or cirrhosis (F4) (0.910 and 0.959 AUROC, respectively). Furthermore, Lee et al. [10] reported that the mean L_{Het} values showed good discrimination for staging of significant fibrosis (≥F2) in chronic hepatitis B (aspartate aminotransferase to platelet ratio index: APRI 0.875 and FIB-4 0.831 AUROC). In the present study, the L_{Het} and L_{Nod} scores in the CLD patients are in accordance with these previous results [10,11,13,14], confirming patients with significant fibrosis (≥F2) and/or precirrhotic hepatic fibrosis.

This study investigated the potential variation in L_{Het} and L_{Nod} measurements and interobserver assessment. To successfully detect signals from the liver parenchyma and surface, all T1-weighted MRI data were performed for bias correction of field homogeneity before the liver contour detection. Quantitative L_{Het} and L_{Nod} scores showed reliable measurements as an averaged CV value <25%. The L_{Het} and L_{Nod} scores measured from two observers showed good interobserver agreement (>0.6), indicating reproducibility. Thus, the WALTS software-based L_{Het} and L_{Nod} measurements can be reproducible in clinical MR images. However, the most accurate test for assessing liver fibrosis is currently MR elastography (MRE), which has ICC >95% and accuracy >90% in liver stiffness measurements [21]. In our study, the L_{Het} and L_{Nod} have ICC of 0.72 and 0.83, respectively,

which is quite low compared to MRE. Thus, further study is needed for a more accurate quantification method in the L_{Het} and L_{Nod} measurements.

With regards to the grading of liver fibrosis, the APRI and FIB-4 serologic indices are well known [22,23]. In a meta-analysis study [22], the pooled ROCs of the FIB-4 index were 0.74–0.84 in the patients with chronic hepatitis B virus infection. The summary ROC (SROC) values of FIB-4 were higher than those of APRI for advanced fibrosis and cirrhosis. Two systematic reviews [23,24] reported that the SROC values for the accuracy of APRI in patients with hepatitis C (HCV) or coinfection of HCV/human immunodeficiency virus (HIV) were 0.76–0.77 for significant fibrosis, 0.80 for advanced fibrosis, and 0.82–0.83 for cirrhosis. Based on these findings, the FIB-4 index for diagnosing liver fibrosis and cirrhosis has similar or superior diagnostic accuracy to that of APRI. For this study, we used the FIB-4 scoring system using serum ALT, AST, and platelet levels. The interesting features in this study are that the L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores are significantly different among fibrosis stages. The mean L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores in severe fibrosis stages F2 and F3 were significantly higher than those in mild fibrosis F1 (as shown in Table 2). Thus, it is notable that quantified L_{Het}, L_{Nod}, and $L_{Het} \times L_{Nod}$ scores can provide information for diagnosing hepatic fibrosis. In previous studies, several imaging methods were reported for differentiating hepatic fibrosis and cirrhosis. A study [25] compared the diagnostic accuracy between gadoxetic acid-enhanced MR imaging, transient elastography, and ultrasound shear wave elastography point quantification (ElastPQ). The gadoxetic acid-enhancement index showed similar diagnostic accuracy for significant fibrosis ($\geq$ F2) or cirrhosis (F4) when using transient elastography (AUROC 0.866 and 0.884) or ElastPQ (AUROC 0.751 and 0.786), respectively. A comparative study [26] using hepato-biliary phase imaging (relative enhancement), susceptibility-weighted imaging (SWI; liver-to-muscle ratio), and diffusion-weighted imaging (DWI; apparent diffusion coefficient: ADC value) reported that the AUROC of SWI showed higher value for diagnosing cirrhosis (F4) than the hepato-billiary phase image and DWI (0.92 vs. 0.80 and 0.79), and the AUROC of the combination of all of these showed the highest value for diagnosing cirrhosis (0.93). A recent study [27] of gadoxetic acid-enhanced MR imaging using a radiomics model based on texture analysis reported that the AUROCs of the radiomic fibrosis index were 0.90, 0.89, and 0.91 for significant fibrosis, advanced fibrosis, and cirrhosis, respectively. In the present study, L_{Het} and L_{Nod} scores have similar, excellent diagnostic accuracy for significant fibrosis ($\geq$ F2). Thus, the L_{Het} and L_{Nod} quantification can be a noninvasive technique capable of detecting fibrotic changes within the liver parenchyma in CLD. The major strengths of the integrated WALTS program include the ability to evaluate previously obtained liver MR or CT images (useful for retrospective large-scale population studies), wide availability of MR and CT imaging, no requirement for intravenous contrast media injection, and no additional hardware requirements for image acquisition procedures. Moreover, the L_{Het} and L_{Nod} quantification program may help predict cirrhosis, liver compensation, and death [16]. Therefore, this MRI-compatible WALTS software may be useful for clinical application to various liver diseases including CLD.

Diagnostic accuracy, reproducibility, and repeatability in L_{Het} and L_{Nod} measurements are crucial for assessing the diagnostic performance of an imaging technique [28]. The L_{Het} and L_{Nod} scores derived from routine liver MR images showed good reproducibility between two different observers in diagnosing hepatic fibrosis. WALTS software can quantify axial 3D-THRIVE MR images in less than 5 min. The applicability of WALTS to retrospective clinical studies has great merits since it allows us to predict liver fibrosis and compare disease progression during prospective long-term follow-up studies.

This study included several limitations. First, this study is a retrospective study with relatively small population size and dealt with CLD patients with heterogeneous underlying disease causes as given in Table 1. The patient cohort used largely included hepatitis B and C. However, there was no consideration of the heterogeneous disease causes in enrolled subjects. Thus, the imaging findings and the predictive power in a larger cohort might be diverse in the patients without viral hepatitis. Also, the distribution

skewness (the F0 group is 16/127 patients (12.6% in study population) and F1 group is only 9/127 patients (<10%)) might cause the potentially skewed results; it can lead to spectrum bias in the study population. Future study is needed for a validation study for strengthening the translational impacts in another larger cohort with even subgroups. Although we included a pathologically confirmed F0 group for comparison, this study used the only FIB-4 index to stage liver fibrosis in CLD as a standard for comparison. Further research would be useful to directly compare our method to FibroScan, which currently represents the most utilized technology in the evaluation of fibrosis. FIB-4 index is good for distinguishing cirrhosis from lower fibrosis stages and even then has a modest accuracy. However, this index might potentially lead to false-positive or true-negative findings due to moderate discrimination accuracy and its own limitations. Second, this study did not consider the relationship between L_{Het} and L_{Nod} scores and complications of liver fibrosis. A Lee et al. study [10] reported that quantified L_{Het} scores are correlated with serologic indices, reflecting liver functional status. Smith et al. [13] reported that a single L_{Nod} score allows the prediction of decompensated cirrhosis and death. Sartoris et al. [29] reported that portal hypertension can be detected using a CT-based L_{Nod} score with a high degree of reliability. Considering these findings, future studies are needed to investigate the correlation between L_{Het} and L_{Nod} scores and complications of liver fibrosis. Third, this study performed the reproducibility test in the L_{Het} and L_{Nod} measurements at a single center. Although the findings showed good reproducibility, the L_{Het} score could be influenced by image noise. Further studies are needed for external or cross-validation of diverse datasets with a large-scale cohort across modality, vendors, study protocols, and external validation at multiple centers. Fourth, this study was only focused on patients with hepatic fibrosis (F0–F3). Therefore, further study is needed to clarify the finding in which patients have liver fibrosis including liver cirrhosis (F4) for actual clinical settings and practices. Also, the future development of a prospective study with a larger population size which incorporates cirrhotic patients with advanced fibrosis stage (F4) may offer further insights on how this new methodology may expand the current standard of care.

5. Conclusions

This study developed an integrated semiautomatic software for the quantification of hepatic heterogeneity and nodularity, and the measurements of L_{Het} and L_{Nod} scores are reproducible in assessing fibrosis stage in CLD. The combination of quantitative L_{Het} and L_{Nod} scores may be more useful for differentially diagnosing the fibrosis stage in CLD using routine MR images.

Author Contributions: Conceived and designed the experiments, T.-H.K., Y.H.L., and K.-H.Y.; software development, T.-H.K., J.E.K., and C.-W.J.; performed the experiments: T.-H.K., J.E.K., and Y.R.K.; analyzed the data, T.-H.K., C.-W.J., J.E.K., J.W.K., Y.R.K., and Y.H.L.; interpreted the data, T.-H.K., C.-W.J., J.W.K., H.G.J., Y.H.L., and K.-H.Y. wrote the paper, T.-H.K., J.W.K., and K.-H.Y. All authors agree with the content of the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study protocol was approved as retrospective research (WKUH-2017-03-026) by the institutional review board (IRB) of Wonkwang University Hospital.

Informed Consent Statement: Written informed consent was waived by the University Hospital IRB committee due to the use of anonymous archival data including MRI data (radiology_common data model: R_CDM, version 2.0.0) and electronic health records (observational medical outcomes partnership-CDM: OMOP-CDM, version 5.3) for the application of developed software.

Data Availability Statement: All anonymized data sources described in this study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Karsdal, M.A.; Manon-Jensen, T.; Genovese, F.; Kristensen, J.H.; Nielsen, M.J.; Sand, J.M.; Hansen, N.U.; Bay-Jensen, A.C.; Bager, C.L.; Krag, A.; et al. Novel insights into the function and dynamics of extracellular matrix in liver fibrosis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2015**, *308*, G807–G830. [CrossRef] [PubMed]
2. Ebel, N.H.; Horslen, S.P. Complications and management of chronic liver disease. In *Diseases of the Liver and Biliary System in Children*, 4th ed.; Kelly, D.A., Ed.; John Wiley & Sons Ltd.: New York, NY, USA, 2017.
3. Embade, N.; Millet, O. Molecular Determinants of Chronic Liver Disease as Studied by NMR-Metabolomics. *Curr. Top. Med. Chem.* **2017**, *17*, 2752–2766. [CrossRef] [PubMed]
4. Bedossa, P.; Carrat, F. Liver biopsy: The best, not the gold standard. *J. Hepatol.* **2009**, *50*, 1–3. [CrossRef] [PubMed]
5. Zhou, I.Y.; Catalano, O.A.; Caravan, P. Advances in functional and molecular MRI technologies in chronic liver diseases. *J. Hepatol.* **2020**, *73*, 1241–1254. [CrossRef] [PubMed]
6. Martínez, S.M.; Crespo, G.; Navasa, M.; Forns, X. Noninvasive assessment of liver fibrosis. *Hepatology* **2011**, *53*, 325–335. [CrossRef] [PubMed]
7. Smith, A.D.; Porter, K.K.; Elkassem, A.A.; Sanyal, R.; Lockhart, M.E. Current imaging techniques for noninvasive staging of hepatic fibrosis. *AJR Am. J. Roentgenol.* **2019**, *213*, 77–89. [CrossRef] [PubMed]
8. Mathew, R.P.; Venkatesh, S.K. Imaging of Hepatic Fibrosis. *Curr. Gastroenterol. Rep.* **2018**, *20*, 45. [CrossRef] [PubMed]
9. Petitclerc, L.; Sebastiani, G.; Gilbert, G.; Cloutier, G.; Tang, A. Liver fibrosis: Review of current imaging and MRI quantification techniques. *J. Magn. Reson. Imaging* **2017**, *45*, 1276–1295. [CrossRef]
10. Lee, G.M.; Kim, Y.R.; Ryu, J.H.; Kim, T.H.; Cho, E.Y.; Lee, Y.H.; Yoon, K.H. Quantitative measurement of hepatic fibrosis with gadoxetic acid-enhanced magnetic resonance imaging in patients with chronic hepatitis B Infection: A comparative study on aspartate aminotransferase to platelet ratio index and fibrosis-4 index. *Korean J. Radiol.* **2017**, *18*, 444–451. [CrossRef] [PubMed]
11. Lee, J.H.; Kim, Y.R.; Lee, G.M.; Ryu, J.H.; Cho, E.Y.; Lee, Y.H.; Yoon, K.H. Coefficient of variation on Gd-EOB MR imaging: Correlation with the presence of early-stage hepatocellular carcinoma in patients with chronic hepatitis B. *Eur. J. Radiol.* **2018**, *102*, 95–101. [CrossRef]
12. Kim, T.H.; Kim, J.E.; Ryu, J.H.; Jeong, C.W. Development of liver surface nodularity quantification program and its clinical application in nonalcoholic fatty liver disease. *Sci. Rep.* **2019**, *9*, 9994. [CrossRef]
13. Smith, A.D.; Branch, C.R.; Zand, K.; Subramony, C.; Zhang, H.; Thaggard, K.; Hosch, R.; Bryan, J.; Vasanji, A.; Griswold, M.; et al. Liver surface nodularity quantification from routine CT images as a biomarker for detection and evaluation of cirrhosis. *Radiology* **2016**, *280*, 771–781. [CrossRef] [PubMed]
14. Pickhardt, P.J.; Malecki, K.; Kloke, J.; Lubner, M.G. Accuracy of liver surface nodularity quantification on MDCT as a noninvasive biomarker for staging hepatic fibrosis. *AJR Am. J. Roentgenol.* **2016**, *207*, 1194–1199. [CrossRef] [PubMed]
15. Besa, C.; Wagner, M.; Lo, G.; Gordic, S.; Chatterji, M.; Kennedy, P.; Stueck, A.; Thung, S.; Babb, J.; Smith, A.; et al. Detection of liver fibrosis using qualitative and quantitative MR elastography compared to liver surface nodularity measurement, gadoxetic acid uptake, and serum markers. *J. Magn. Reson. Imaging* **2018**, *47*, 1552–1561. [CrossRef] [PubMed]
16. Smith, A.D.; Zand, K.A.; Florez, E.; Sirous, R.; Shlapak, D.; Souza, F.; Roda, M.; Bryan, J.; Vasanji, A.; Griswold, M.; et al. Liver surface nodularity score allows prediction of cirrhosis decompensation and death. *Radiology* **2017**, *283*, 711–722. [CrossRef]
17. Huang, G.; Ji, H.; Zhang, W. A fast level set method for inhomogeneous image segmentation with adaptive scale parameter. *Magn. Reson. Imaging* **2018**, *52*, 33–45. [CrossRef] [PubMed]
18. Rozario, L.J.; Moazzam, M.G.; Begum, M.; Naowar, F.F. An Improved Method of Segmentation of Medical Images thatc Incorporates Bias Correction. *Int. J. Eng. Sci.* **2014**, *3*, 1–8.
19. Kim, T.H.; Jun, H.Y.; Kim, K.J.; Lee, Y.H.; Lee, M.S.; Choi, K.H.; Yun, K.J.; Jeong, Y.Y.; Jun, C.H.; Cho, E.Y.; et al. Hepatic alanine differentiates nonalcoholic steatohepatitis from simple steatosis in humans and mice: A proton MR spectroscopy study with long echo time. *J. Magn. Reson. Imaging* **2017**, *46*, 1298–1310. [CrossRef] [PubMed]
20. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]
21. Zhou, Y.; Xu, J.; Liu, Q.; Li, C.; Liu, Z.; Wang, M.; Zheng, H.; Wang, S. A Radiomics Approach With CNN for Shear-Wave Elastography Breast Tumor Classification. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1935–1942. [CrossRef] [PubMed]
22. Xiao, G.; Yang, J.; Yan, L. Comparison of diagnostic accuracy of aspartate aminotransferase to platelet ratio index and fibrosis-4 index for detecting liver fibrosis in adult patients with chronic hepatitis B virus infection: A systemic review and meta-analysis. *Hepatology* **2015**, *61*, 292–302. [CrossRef]
23. de Oliveira, A.C.; El-Bacha, I.; Vianna, M.V.; Parise, E.R. Utility and limitations of APRI and FIB4 to predict staging in a cohort of nonselected outpatients with hepatitis C. *Ann. Hepatol.* **2016**, *15*, 326–332. [CrossRef] [PubMed]
24. Yen, Y.-H.; Kuo, F.-Y.; Kee, K.-M.; Chang, K.-C.; Tsai, M.-C.; Hu, T.-H.; Lu, S.-N.; Wang, J.-H.; Hung, C.-H.; Chen, C.-H. APRI and FIB-4 in the evaluation of liver fibrosis in chronic hepatitis C patients stratified by AST level. *PLoS ONE* **2018**, *13*, e0199760. [CrossRef]

25. Jang, H.J.; Min, J.H.; Lee, J.E.; Shin, K.S.; Kim, K.H.; Choi, S.Y. Assessment of liver fibrosis with gadoxetic acid-enhanced MRI: Comparisons with transient elastography, ElastPQ, and serologic fibrosis markers. *Abdom. Imaging* **2019**, *44*, 2769–2780. [CrossRef] [PubMed]

26. Feier, D.; Balassy, C.; Bastati, N.; Fragner, R.; Wrba, F.; Ba-Ssalamah, A. The diagnostic efficacy of quantitative liver MR imaging with diffusion-weighted, SWI, and hepato-specific contrast-enhanced sequences in staging liver fibrosis–a multiparametric approach. *Eur. Radiol.* **2016**, *26*, 539–546. [CrossRef] [PubMed]

27. Park, H.J.; Lee, S.S.; Park, B.; Yun, J.; Sung, Y.S.; Shim, W.H.; Shin, Y.M.; Kim, S.Y.; Lee, S.J.; Lee, M.G. Radiomics analysis of gadoxetic acid-enhanced MRI for staging liver fibrosis. *Radiology* **2019**, *290*, 380–387. [CrossRef] [PubMed]

28. Raunig, D.L.; McShane, L.M.; Pennello, G.; Gatsonis, C.; Carson, P.L.; Voyvodic, J.T.; Wahl, R.L.; Kurland, B.F.; Schwarz, A.J.; Gonen, M.; et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat. Methods Med. Res.* **2015**, *24*, 27–67. [CrossRef]

29. Sartoris, R.; Rautou, P.E.; Elkrief, L.; Pollorsi, G.; Durand, F.; Valla, D.; Spahr, L.; Terraz, S.; Soubrane, O.; Cauchy, F.; et al. Quantification of liver surface nodularity at CT: Utility for detection of portal hypertension. *Radiology* **2018**, *289*, 698–707. [CrossRef] [PubMed]