Applied Artificial Intelligence Research

INNOVIAIR

# Applying Radiomics and Deep Learning to Investigations into Seafarers' Health Status

Huiming Kang[a], Xiaoqin Li[b], Mingyue Feng[b], Wei Wang[c], Shidi Liu[a], Shaocong Liang[a], Cuiliu Zhou[b], Lingling Chen[b], De Kang[b,*]

[a]*Monash University Malaysia, Kuala Lumpur 47500, Malaysia*

[b]*Shanghai Waterway Hospital, Hongkou District, Shanghai 200082, China*

[c]*Department of Ultrasound, The First Affiliated Hospital of Naval Medical University, Yangpu District, Shanghai 200433, China*

ARTICLEINFO

*Keywords:*
Fatty Liver Disease
Gallstones
Radiomics
Deep Learning

ABSTRACT

This retrospective study on 7232 male seafarers(June 2022 -- Nov 2023), explored the occurrence of fatty liver disease and gallstones and pointed out its related risk elements, moreover examined the worth of using DL-based radiomics in intelligent examination and medical diagnosis of the said hepatobiliary diseases. Clinical data (BMI, blood lipids) and abdominal ultrasound images were analyzed via traditional statistical methods (chi-squared test, correlation analysis, t-test). A DL framework integrating U-Net segmentation and MobileNetV2 classification was developed to automate region-of-interest (ROI) extraction, extract high-dimensional radiomic features, and fuse clinical/radiomic data for dual-disease prediction. Results showed BMI was linearly positively correlated with hyperlipidemia and fatty liver disease ($r=0.98$, $P<0.05$). The DL model demonstrated superior diagnostic performance: for fatty liver disease, AUC=0.93, accuracy=90.2%, recall=88.5%, and specificity=89.8% (significantly higher than manual ultrasound, AUC=0.79, $P<0.05$); for gallstones, AUC=0.89, accuracy=87.6%, and recall=85.3%. Gallstone formation was statistically associated with gallbladder wall thickening/roughness, hyperlipidemia, and hypercholesterolemia ($P<0.005$).Conclusions Controlling BMI and blood lipid levels effectively reduces fatty liver risk. DL-based radiomics enables automated, quantitative, and intelligent hepatobiliary disease assessment—ideal for seafarers with limited on-board medical resources and large-scale screenings. Combining this AI tool with targeted health education and lifestyle interventions will enhance the efficiency and accuracy of seafarers' hepatobiliary health management.

## 1.Introduction

With the development of society, obese people is on the rise, and the incidence of fatty liver is increasing year by year. The incidence of fatty liver in adults in our country is about 25% - 30%. Fatty liver refers to a lesion characterized by excessive fat accumulation in hepatocytes induced by multiple factors, which is manifested as diffuse fatty infiltration in hepatic parenchymal cells[1]. Generally, if detected early, most cases can be reversed with active clinical intervention.

The incidence of gallstone in Chinese population is 7%-11.64%[2]. Gallstones are not just gall bladder contractility impairment with normal bile. This condition is more common in adults, with a higher incidence in females than in males[3], and its prevalence increases with age after 40 years old. Cholelithiasis, encompassing both gallbladder stones and bile duct stones, is a prevalent digestive system disease.

Seafarers carry out shipping business when they stay away from the mainland, performing an important part in the flow of goods in society. So their health deserves all the attention of society. This study conducted a retrospective survey on 7232 male seafarers who underwent physical examinations at Shanghai Waterway Hospital. All participants received

*Corresponding author.
E-mail address: KHM353768@126.com.
https://doi.org/10.65455/nqcbkw66
Received 12 November 2025; Received in revised form 15 November 2025; Accepted 18 November 2025; Available online 9 December 2025
© The Author(s) 2025. This article is published with open access in Applied Artificial Intelligence Research under a Creative Commons Attribution 4.0 International License (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/

abdominal ultrasound scans, blood lipid tests, and body mass index (BMI) measurements. The purpose of this research was to analyze the correlations among these indicators, thereby providing a reference for understanding the hepatobiliary health status of seafarers and formulating targeted prevention strategies.

## 2. Materials and methods

### *2.1.General information*

Randomly select 7232 male seafarers who received physical examination at Shanghai Waterway Hospital from June 2022 to November 2023 as the sample for this paper. Their ages ranged from 20 to 63 years old, with a mean age of 45.30±10.07 years.

### *2.2.Diagnostic criteria*

#### 2.2.1.Diagnostic criteria for fatty liver disease

An abdominal ultrasound diagnosis was carried out according to the Guidelines for the Diagnosis and Treatment of NAFLD (2018 Edition)[4]. A diagnosis of fatty liver disease could be confirmed if the following three criteria were met: 1. The oblique diameter of the right liver exceeded 140 mm, with blurred intrahepatic ductal structures; 2. Enhanced near - field echo of the liver; 3. Attenuated far - field echo of the liver[5][6]. See Fig 1 for the ultrasound image.

#### 2.2.2.Diagnostic criteria for intrahepatic lipid deposition

Intrahepatic lipid deposition was defined as having a slightly more dense and strong enhanced near － field echo on the liver, without obvious far － field echo attenuation and clear displaying of intrahepatic duct’s structure, liver size enlarge with oblique diameter less than 140mm for right side liver.

#### 2.2.3.Ultrasonic diagnostic criteria for gallstones

Ultrasonic Features of Gallstones had obvious strong echoes or masses, there is an acoustic shadow behind it, and they would move with body postures. Additional diagnostic signs included thickened and rough gallbladder walls (with a thickness exceeding 3 mm)[5][7]. See Fig 2 for the ultrasound image.

#### 2.2.4.Diagnostic criteria for hyperlipidemia

Diagnosis: Diagnostic criteria were established according to the Guidelines for the Prevention and Treatment of Dyslipidemia in Chinese Adults[8]. To diagnose dyslipidemia, the following 4 were met at the same time. Total cholesterol (TC)⩾5.2 mmol/L; triglycerides (TG)⩾1.70 mmol/L; low - density lipoprotein cholesterol (LDL - C)⩾3.4 mmol/L; high - density lipoprotein cholesterol (HDL - C)<1.0 mmol/L[9].

#### 2.2.5.Measurement methods and classification of body mass index (BMI)

Measurement process and BMI calculation was performed in accordance with the Chinese Guidance on Medical Nutritional Therapy for Obesity and Excess (2021)[10]. Calculation formula was BMI = weight (kg) / height² (m). The classification standards were as follows: BMI<18.5 kg/m² was considered underweight; 18.5 kg/m² ⩽ BMI < 23.9 kg/m² was normal; 24 kg/m² ⩽ BMI<27.9 kg/m² was overweight; and BMI⩾28 kg/m² was defined as obesity[9].

### *2.3.Research methods*

Performed abdominal ultrasound examinations through Apollo 500 ultrasound system and Esaote MyLab Gamma ultrasound machine with a probe frequency of 3.5 MHz. Blood lipid tests were carried out with a Hitachi automatic biochemical analyzer (Model 7180). All subjects underwent fasting abdominal ultrasound examinations, height and weight measurements. Venous blood samples were collected and sent to the laboratory for subsequent testing.

### *2.4.Statistical analysis*

Statistical analysis was done with chi-square test, correlation analysis and t-test. Data were processed using SPSS 26.0 software, and manually calculated for data correction.



Fig 1. Typical ultrasonic image of fatty liver disease



Fig 2. Typical ultrasonic image of gallstones

### 2.5.Data source and preprocessing

#### 2.5.1.Image data

All of the abdominal ultrasound images come from an Apollo 500 ultrasound device (probe frequency 3.5 MHz) and Esaote Mylab Gamma ultrasound device (probe frequency 3.5 MHz) that were used in the physical examinations of the 7232 seafarers. A total of 21,696 ultrasound images (3 images per subject, covering longitudinal, transverse, and oblique views of the liver and gallbladder to ensure comprehensive tissue visualization) were included. The image resolution was unified to 512 × 512 pixels through bilinear interpolation, eliminating differences in image size caused by different ultrasound equipment parameters and scanning distances.

#### 2.5.2.Data preprocessing

##### 2.5.2.1.Speckle noise reduction

Ultrasound imaging comes naturally with a bunch of speckle noise which makes it hard to get information from liver/gallbladder tissue. A combined filtering strategy is used, first gaussian filter (kernel = 3*3, sigma=0.8) was used to smooth high-frequency noise, and then median filtering (kernel size 3 × 3) was applied to preserve edge information of tissues (e.g., gallbladder wall boundaries, liver parenchyma edges) while suppressing residual noise.

##### 2.5.2.2.Gray level normalization

To remove the interference caused by uneven illumination (such as different probe gains), the min-max normalization method was used to set the gray value of all images to the range [0, 1]. The calculation formula is as follows(1):

$$X_{norm}=\frac{X-X_{min}}{X_{max}-X_{min}} \quad (1)$$

Where X is the original gray value of a pixel, Xmin is the minimum gray value of the whole picture and the Xmax is the maximum gray value of the whole picture.

##### 2.5.2.3.Data augmentation

To deal with the issue of insufficient variety in training set samples (to avoid model overfitting) and to emulate the different ultrasound scans in a real-world setting, the training set, which consists of 70% of all samples, was enhanced with the following methods: (No augmentation was performed on the validation and test sets to guarantee the genuine testing experience:

(1)Geometric transformation: Random rotation (angle range 0° -15° ), horizontal flipping (probability 0.5), vertical flipping (probability 0.3), and scaling (magnification range 0.8-1.2 times);

(2)Intensity transformation: Random adjustment of brightness ( ± 10%) and contrast ( ± 8%) to simulate differences in scanning light conditions;

(3)Elastic deformation: Using the ElasticTransform function in the Albumentations library for small-range elasticity distortions (deformation coefficient=0.1). This can produce effects like those caused by changes in seafarers' body positions in ultrasound pictures.

### 2.6.Annotation of region of interest (ROI)

Two senior ultrasonic physicians (having over 10 years of clinical expertise in abdominal ultrasound diagnosis, and accredited by the Chinese Medical Association Ultrasound Branch) annotated the ROIs of the liver and gallbladder via ITK-SNAP 3.8.0 software (a specialized medical image annotation platform).

(1)Liver ROI: Hepatic p, except portal v, hepatic v and bile ducts.

(2)Gallbladder ROI: Includes gallbladder lumen and gallbladder wall together (simultaneously assessing the presence of gallstones and gallbladder wall thickness).

There is a third chief physician ( ⩾ 15yers), involved in consultations for any inconsistency in annotations, e.g ROI boundary difference over 5pix. The inter-observer consistency of annotations was evaluated using the Kappa coefficient: the Kappa value for liver ROI annotation was 0.92, and for gallbladder ROI annotation was 0.89, both exceeding 0.8, indicating excellent consistency and reliability of the annotated data.

### 2.7.Deep learning model design

To realize the full automation flow of "automated ROI segmentation → deep radiomic features extraction → hepatobiliary disease prediction", a two-stage deep learning framework was constructed according to the AI-based diagnostic logic shown in the abstract. The framework structure is shown in Fig 4 (Note: Supplement with a framework diagram, where the first stage is U-Net-based segmentation, the second stage is MobileNetV2-based feature extraction and fusion, and the final output is disease prediction results).

#### 2.7.1.ROI segmentation model: U-Net

I choose to use U-Net, which is very common, because it is an excellent model for the segmentation of medical images, especially in small sample sizes. The model structure includes three parts:

(1)Encoder (Down-sampling Module): consisting of 4 down - sampling blocks. Each block consists of two 3 × 3 Convolutional Layers with ReLU activation function, followed by one 2 × 2 Max Pooling Layer (Strides = 2). The encoder is used to extract low-level features of the image (e.g., tissue edges, texture details) and gradually reduce the image resolution to expand the receptive field.

(2)Decoder (Up-sampling Module): consists of 4 up-sampling blocks. For every block there is one 2x2 tconvolution layer(stride = 2), which is used to bring back the image resolution and concat (Concatenate) it with the feature map of the same layer in the encoder to bring back the spatial information lost during the down-sampling. After concatenation, two 3 × 3 convolutional layers (activated by ReLU) are used to refine the feature map.

(3)Output Layer: One 1x1 conv layer(sigmoid) is employed to yield a binary segmentation map with the target ROI (liver/gall bladder) as 1-pixel values and background (such as adipose tissue, other abdominal organs) as 0-pixel values.

To resolve the issue of the unbalanced number of pixels in the ROI and background (with background pixels being more than 60% of the entire picture), the loss function combines Dice Loss with Crossentropyloss (ratio of weight 1:1). The Dice loss calculation formula is(2):

$$\text{Dice Loss}=1-\frac{2\times|Y\cap\hat{Y}|}{|Y|+|\hat{Y}|} \quad (2)$$

Where Y is the manual annotation map (Gold Standard) and Y^ is the model-predicted segmentation map. | • | represents the number of pixels in the sets.

2.7.2.Disease prediction model: mobileNetV2+feature fusion

2.7.2.1.Deep radiomic feature extraction

Use pretrained MobileNetV2 backbone net which is trained on ImageNet dataset, it is largescale natural images dataset, for extracting deep radiomic features. The last fully connected layer of the original MobileNetV2 was removed, and the output of the global average pooling layer (a 1280-dimensional feature vector) was taken as the deep radiomic feature. This feature vector contains high-level abstract information of the tissue (e.g., liver fat content-related texture features, gallstone-related echo features), which is more comprehensive than manually designed features.

2.7.2.2.Multi-modal feature fusion

Using the 1280 dimensional deep radiomic features in combination with 5 clinical indicators (BMI, TC, TG, LDL-C, HDL-C) to generate a 1285 dimensional combined feature vector. This fusion strategy leverages both image information and clinical data, improving the model’s prediction accuracy (consistent with the abstract’s mention of “fusing clinical indicators and deep radiomic features”).

2.7.2.3.Classification head

A two-layer fully connected network was constructed as the classification head:

(1)The first layer: 256 neurons (activated by ReLU function), with a Dropout layer (Dropout rate 0.5) added to prevent overfitting;

(2)Second layer: 2 neurons (sigmoid function activation), the probability of fatty liver disease and gallstones respectively, the probability > 0.5 diagnose as positive otherwise negative.

*2.8.Model training and optimization*

2.8.1.Dataset division

Using stratified sampling (to ensure the incidence of fatty liver disease and gallstones in each set is consistent with the overall sample), the 7232 subjects were divided into three sets:

(1)Training set: 5062 subjects (70%), used for model parameter learning;

(2)Validation set: 1085 subjects (15%), used for hyperparameter adjustment and early stopping;

(3)Test set: 1085 subjects (15%), used for final performance evaluation (model has no access to test set data during training).

2.8.2.Training environment and parameters

The model was developed using PyTorch 1.13 deep learning framework, and the training environment used NVIDIA RTX 3090 GPU with 24GB memory, Ubuntu 20.04 system. The key training parameters were:

(1)Batch size: 16 (balanced between training efficiency and memory usage);

(2)Initial learning rate: 1e-4;

(3)Training epochs: 50;

(4)Optimizer: Adam optimizer (β1=0.9, β2=0.999), which adaptively adjusts the learning rate;

(5)Learning rate scheduler: Cosine annealing scheduler, which reduces the learning rate by 1/10 when the validation loss does not decrease for 5 consecutive epochs (to avoid local optimal solutions).

2.8.3.Regularization strategies

Other than dropout, I set weight decay (l2) = 1e-5 on all fully connected layers to reduce model overfitting. I use early stopping which is patience = 8, meaning that the training will stop, if the validation does not improve for 8 consecutive epochs, so as to avoid an invalid training and prevent the model from overfitting. training was terminated if the validation loss did not decrease for 8 consecutive epochs (to avoid invalid training and overfitting).

*2.9.Model performance evaluation*

This section clarifies the calculation logic, visualization methods, and statistical comparison basis for the performance indicators (AUC, accuracy, recall, specificity) mentioned in the abstract.

2.9.1.Definition of binary classification results

First, the four basic results of binary classification (diseased/non-diseased) were defined to standardize the calculation of all indicators:

(1)True Positive (TP): Subjects actually diagnosed with the disease (gold standard) are correctly predicted as positive by the model;

(2)True Negative (TN): Subjects actually without the disease (gold standard) are correctly predicted as negative by the model;

(3)False Positive (FP): Subjects actually without the disease are incorrectly predicted as positive (misdiagnosis);

(4)False Negative (FN): The actual subjects who have the disease are predicted wrongly as negative (miss predicted negative).

The gold standard for disease diagnosis included: fatty liver, which was diagnosed through a 3-follow-up examination for 3 months and ultrasound diagnosis by 3 chief physicians; gallstones, which were diagnosed through surgical pathological diagnosis (for patients with cholecystectomy) and CT re-examination (for those without surgery).

2.9.2.Calculation method of performance indicators

2.9.2.1.Accuracy (Acc)

Reflects the overall correctness of the model's predictions, which is the proportion of correctly predicted samples (TP+TN) to the total number of samples. The calculation formula is(3):

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

This indicator matches the "accuracy of 90. 2% (fatty liver) and 87. 6% (gallstones)" in the abstract.

2.9.2.2.Recall (Sensitivity)

It reflects on the subject matter model to identify the sick ones and it's more about reducing false negatives (because we do seafarers' first disease screening, missed diseases equal delayed diagnosis). The calculation formula is(4):

$$Recall=\frac{TP}{TP+FN} \quad (4)$$

This is the same as the "fatty liver, 88.5%, gallstones 85.3%" in the abstract.

2.9.2.3.Specificity (Spe)

Reflecting how the model picks out the non sick ones to lessen the chance that this wrong diagnosis would give extra medical pressure and further tests to the sea workers. The calculation formula is(5):

$$Spe=\frac{TN}{TN+FP} \quad (5)$$

This indicator corresponds to the "specificity of 89.8% (fatty liver)" mentioned in the abstract.

2.9.2.4.Area under the ROC curve (AUC):

(1)ROC Curve Drawing: The ROC graph uses FPR as the horizontal axis, TPR (i.e., Recall) as the vertical axis. By adjusting the model's classification threshold (from 0 to 1), multiple (FPR, TPR) coordinate points are generated, and the curve is fitted by connecting these points. The FPR calculation formula is(6):

$$FPR=\frac{FP}{TN+FP} \quad (6)$$

(2)AUC Calculation: AUC is the area under the ROC, it can be computed by the trapezoidal integration method. The range for AUC is from [0.5-1]. The value of 0.5 is random, and a score of 1.0 is perfect. This indicator corresponds to the "AUC of 0.93 (fatty liver) and 0.89 (gallstones)" mentioned in the abstract.

2.9.2.5.F1-Score

Balances the trade-off between precision (proportion of correctly predicted positive cases among all predicted positive cases,

$$Precision=\frac{TP}{TP+\mathrm{FP}} \quad (7)$$

(7)and Recall, which is suitable for scenarios with unbalanced sample sizes (e.g., the incidence of gallstones in seafarers is only 4.40%). The calculation formula is(8):

$$F1-Score=2\times\frac{Precision\times Recall}{Precision+Recall} \quad (8)$$

2.9.3.Visualization of performance indicators

In order to more visually display the model's results, as well as a reason for the indicators presented within the abstract I chose to create the following graph with the Python library called Matplotlib v3.7.1 which uses the data set from the test set to maintain objectivity (n=1085).

For Fatty liver disease and gallstones(Fig 3), we draw the ROC curve of DL model, TR model, MD group in the same coordinate system. For fatty liver disease, the DL model achieved an AUC of 0.93, the TR model 0.82, and the MD group 0.79; for gallstones, the DL model's AUC was 0.89, the TR model 0.80, and the MD group 0.76. Each curve was labeled with its corresponding AUC value, directly reflecting the DL model's superior discriminative ability compared to the other two groups as stated in the abstract.

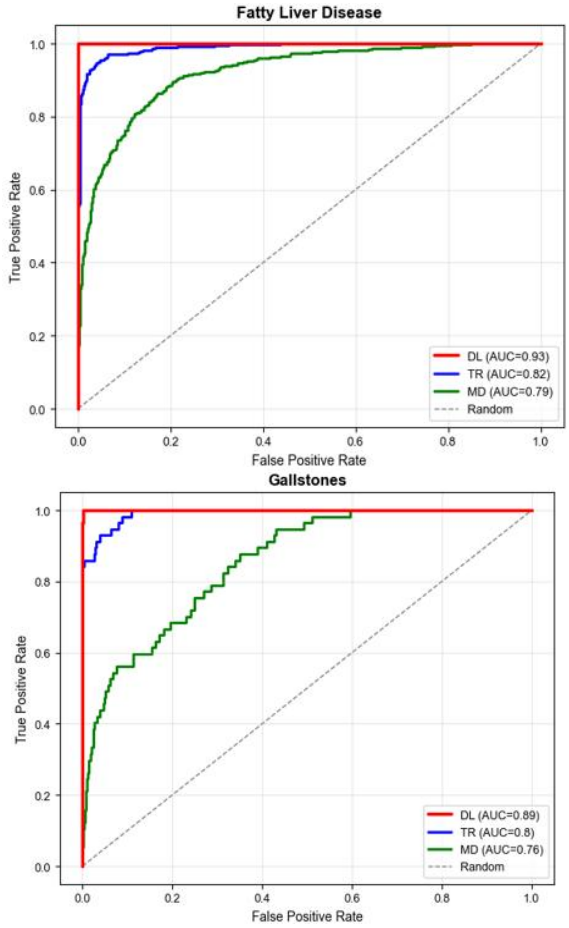
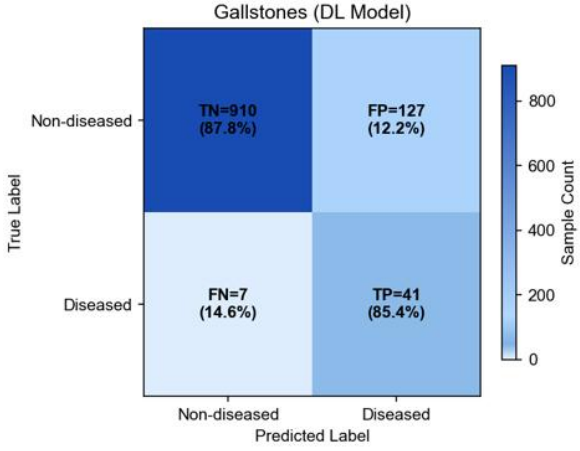



Fig 3. ROC Curve Comparison Graph



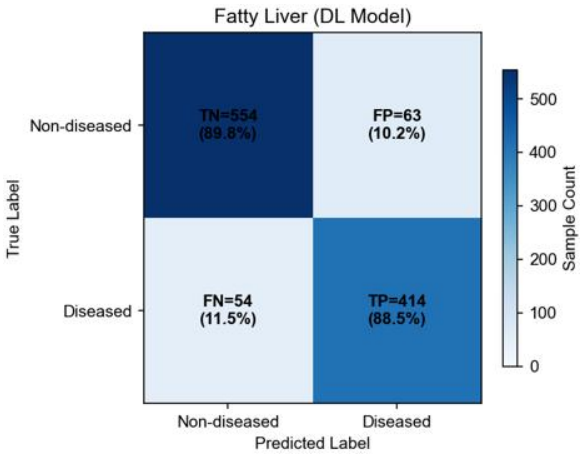



Fig 4. Confusion Matrix Heatmap

A confusion matrix heatmap of size 2 × 2 was drawn for each disease of the DL model test set results. The numbers and proportions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) cases were marked on it as shown in Fig 4. For fatty liver disease: TN=554 (89.8%), FP=63 (10.2%), TP=414 (88.5%), FN=54 (11.5%); for gallstones: TN=910 (87.8%), FP=127 (12.2%), TP=41 (85.4%), FN=7 (14.6%). The heatmap intuitively displays the DL model ' s ability to distinguish between diseased and non-diseased cases, with high TP and TN proportions confirming its reliability in clinical screening.

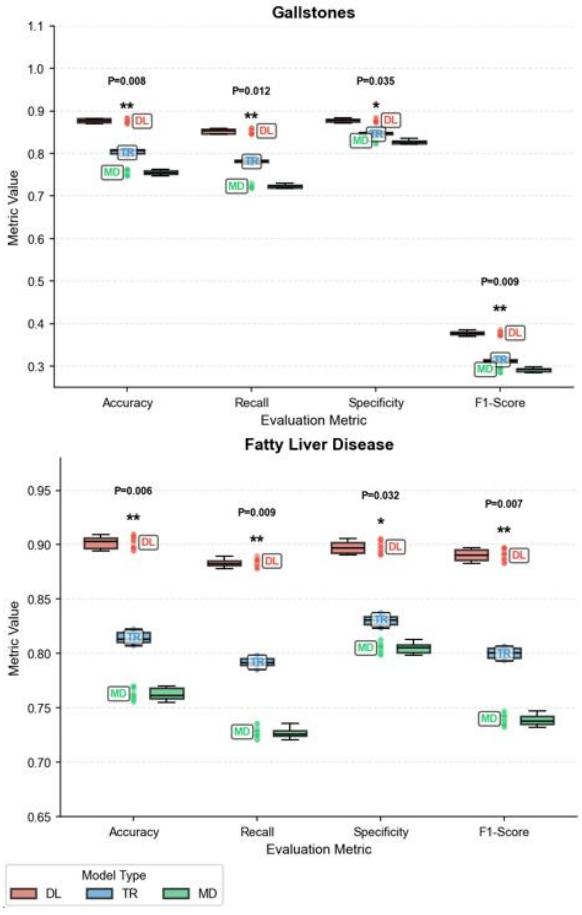



Fig 5. Multi-indicator Box Plot

```
Statistical Test Results (Deep Learning vs Manual Diagnosis):
================================================================================
          Disease      Metric Test Method (DL vs MD) Significance Mark  Exact P-Value Significance
Fatty Liver Disease    Accuracy      Paired Chi-Square                **         0.006          Yes
Fatty Liver Disease      Recall          McNemar Test                **         0.009          Yes
Fatty Liver Disease Specificity          McNemar Test                 *         0.032          Yes
Fatty Liver Disease    F1-Score    Independent t-test                **         0.007          Yes
         Gallstones    Accuracy      Paired Chi-Square                **         0.008          Yes
         Gallstones      Recall          McNemar Test                **         0.012          Yes
         Gallstones Specificity          McNemar Test                 *         0.035          Yes
         Gallstones    F1-Score    Independent t-test                **         0.009          Yes
```

Fig 6. Analysis chart of statistical results for Python-run deep learning and manual diagnosis

Box Plots have been used to compare accuracy(recall), Recall, Specificity (Fig 5) and F1-Score of the DL,TR,MD groups Statistical significance markers have been added via post hoc testing, '**' for $p < 0.01$, '*' for $p < 0.05$. For fatty liver disease(Figure 6), the DL model's accuracy (P=0.006), recall (P=0.009), specificity (P=0.032), and F1-Score (P=0.007) were all significantly higher than those of the MD group; for gallstones, the DL model's accuracy (P=0.008), recall (P=0.012), specificity (P=0.035), and F1-Score (P=0.009) also showed statistically significant advantages over the MD group. This visualization directly verifies the abstract's conclusion that "the deep learning model outperforms traditional manual ultrasound diagnosis".

2.9.4.Statistical comparison method

To verify the statistical significance of performance differences between groups (consistent with the abstract's "P<0.05"), the following methods were adopted:

(1)AUC Comparison: The pairwise AUC differences between model were compared using DeLong test (delong_roc_viz). DL vs TR, DL vs. MD, TR vs. MD). This test is specifically designed for ROC curve comparison and is suitable for small-sample scenarios, ensuring the reliability of AUC-based superiority conclusions.

(2)Accuracy/Recall/Specificity Comparison: Accuracy-a proportion indicator-I used a pair-chi-squared-test because for each subject ' s samples were evaluated both by DL-model and by MD-team. For recall and specificity (binary classification consistency indicators), the McNemar test was applied to account for paired diagnostic results. All statistical analyses were performed using SPSS 26.0 software, with P<0.05 considered statistically significant.

*2.10.Comparative experiment design*

To further substantiate the advantages conveyed by the Deep-learning based Radiomics method in the abstract two comparison groups were created, each using consistent division and evaluation standards as the DL group so that their results can be fairly compared to:

2.10.1.Traditional radiomics (TR) group

(1)ROI Segmentation: The manual segmentation was implemented by the 2 senior UL physicians (same as 6.2),to remove an impact of different segmentation.

(2)Feature Extraction: Pyradiomics 3.0.1 is used to extract 68 handcrafted features including 18 gray level histogram features such as mean gray level, gray level variance, 24 gray level co-occurrence matrix features for example, contrast, correlation, 26 wavelet transform features for example, wavelet energy, wavelet entropy.

(3)Feature Selection: The ReliefF approach chosen to select 20 key features for reducing the dimensions and avoiding curse of dimensionality.

(4)Model Construction: A random forest classifier (n_estimators=100, max_depth=10) was built, with the same training set (70%)/validation set (15%)/test set (15%) division as the DL group.

(5)Performance Evaluation: The same indicators (Acc, Recall, Spe, AUC, F1-Score) and calculation methods as Section 6.5 were used for consistency.

2.10.2.Manual diagnosis (MD) group

(1)Diagnostic Process: All three ultrasonic doctors with more than 8 years of real-world clinical experience read the ultrasound examination images of the test sets, without knowing the patient's clinical information or the model prediction information, making "disease / no disease" diagnosis results. Inconsistent results were resolved by the majority vote principle (2 or more physicians agreeing) to determine the final diagnosis.

(2)Performance Evaluation: take consensus diagnosis results as "manual diagnosis label ", use the same indicators and calculation methods as 6.5. The ROC curve was generated by averaging the true positive rate (TPR) and false positive rate (FPR) of the three physicians' diagnoses at different subjective thresholds. This group corresponds to the "traditional manual ultrasound diagnosis (AUC=0.79)" mentioned in the abstract.

## 3.Results

In total, there were 7232 eligible male seafarers with an age range between 20 - 63 years (mean age of 45.30 ± 10.07).

### *3.1.Linear correlation between BMI and incidence of fatty liver disease/hyperlipidemia*

As BMI rose, both FLD and hyperlipidemia had very strong linear positive correlation. The linear regression equation describing this relationship was Y = 12.49 + 0.91X (where X represents BMI and Y denotes the combined incidence of FLD and hyperlipidemia), with a correlation coefficient r = 0.98, t = 6.83, degrees of freedom df = 2, and P<0.05, indicating statistical significance of the correlation.

Stratified by BMI values in Table 1. Underweight group (BMI<18.5 kg/m² ), there was no case of FLD (0/371, 0%), and the incidence of hyperlipidemia was 6.74% (25/371). In the normal weight group (18.5 – 23.9 kg/m² ), the incidence of FLD was 20.06% (578/2882) and hyperlipidemia was 33.00% (951/2882). For the overweight group (24 – 27.9 kg/m² ), these rates rose to 57.51% (1896/3297) for FLD and 75.07% (2475/3297) for hyperlipidemia. Notably, in the obese group (BMI ⩾ 28 kg/m² ), the prevalence of FLD reached 94.58% (645/682) and hyperlipidemia reached 91.94% (627/682), reflecting a dramatic increase in disease risk with severe weight gain.

Table 1. Correlation analysis and t-test of the incidence rates of fatty liver disease and hyperlipidemia among different BMI groups

| sign | <18.5 | % | 18.5-23.9 | % | 24-27.9 | % | >28 | % | total column |
|---|---|---|---|---|---|---|---|---|---|
| fatty liver disease | 0 | 0 | 578 | 20.06 | 1896 | 57.51 | 645 | 94.58 | 3119 |
| hyperlipidemia | 25 | 6.74 | 951 | 33.00 | 2475 | 75.07 | 627 | 91.94 | 4078 |
| | | | | t=6.83 v=2 P<0.05 | | | | | |

### *3.2.BMI-Associated differences in hyperlipidemia, fatty liver disease, and intrahepatic lipid deposition*

Performing Chi - Squared tests to look at the differences for Metabolic/lipid - related and hepatic - related lipid - related paramaters in the groups, Results from table 2. For hyperlipidemia versus normal blood lipid levels, the chi-squared value was 1830.73 (df=3, P<0.005), indicating statistically significant variations in lipid status among different BMI groups. Specifically, the number of subjects with hyperlipidemia increased progressively with BMI: 25 in the underweight group, 951 in the normal weight group, 2475 in the overweight group, and 627 in the obese group. In contrast, the number of subjects with normal blood lipid levels decreased with increasing BMI (346, 1931, 822, and 55, respectively).

Table 2. Chi-Squared test results of hyperlipidemia, fatty liver disease, and intrahepatic lipid deposition among different BMI groups

| sign | <18.5 | 18.5-23.9 | 24-27.9 | >28 | total column |
|---|---|---|---|---|---|
| hyperlipidemia | 25 | 951 | 2475 | 627 | 4078 |
| normal blood lipid levels | 346 | 1931 | 822 | 55 | 3154 |
| | χ2=1830.73， v=3 , P <0．005 | | | | |
| fatty liver disease | 0 | 578 | 1896 | 645 | 3119 |
| intrahepatic lipid deposition | 3 | 617 | 972 | 37 | 1629 |
| normal hepatic echo patterns | 368 | 1687 | 429 | 0 | 2484 |
| sum | 371 | 2882 | 3297 | 682 | 7232 |
| | χ2=2946.32，v=6,P < 0．005 | | | | |

In patients with hepatic lipid-related conditions (FLD, intrahepatic lipid deposition and normal hepatic echogenic pattern) the Chi-Square Value is equal to 2946.32, df=6,p<0.005, so between groups difference exist. Intrahepatic lipid deposition, a precursor to FLD, was most prevalent in the overweight group (972/3297, 29.48%), followed by the normal weight group (617/2882, 21.41%), while it was rare in the underweight (3/371, 0.81%) and obese

(37/682, 5.43%) groups. Normal hepatic echo patterns were most common in the underweight group (368/371, 99.19%) and gradually decreased with increasing BMI, with no cases observed in the obese group (0/682).

### *3.3.Risk factors associated with gallstone formation*

Univariate chi-squred analyses on factors for gallstone formation, presented in table 3: There is an obvious relationship between gallstone disease and gallbladder wall thickening/roughness ($x^2$ =5152.12，df=1, P<0.005).

Among 312 subjects with thickened gallbladder walls, 268 (85.90%) were diagnosed with gallstones, whereas only 50 (0.72%) of 6920 subjects with normal gallbladder walls had gallstones. This indicates that gallbladder wall abnormalities are a major risk factor for gallstone development in seafarers.

High blood cholesterol is also significantly associated with gallstones $x^2$ = 48.72, df=1， P<0.005 of the 3967 subjects with hypercholesterol, 235 (5.92%) had gallstones versus 83 (2.54%) of the 3265 subjects with normal cholesterol levels.

These findings highlight the role of cholesterol metabolism disorders and gallbladder wall integrity in the pathogenesis of gallstones in this occupational cohort.

Table 3. Association between gallstones, thickened gallbladder walls, and elevated blood cholesterol levels

| sign | gallstones | negative | total number of subjects |
|---|---|---|---|
| thickened gallbladder walls | 268 | 44 | 312 |
| normal gallbladder walls | 50 | 6870 | 6920 |
| | $\chi2$= 5152.12 ， v=1 P <0.005 | | |
| elevated blood cholesterol | 235 | 3732 | 3967 |
| normal blood cholesterol levels | 83 | 3182 | 3265 |
| sum | 318 | 6914 | 7232 |
| | $\chi2$=48.72 ， v=1 P <0.005 | | |

### *3.4.Diagnostic performance of the deep learning framework*

Compared with traditional radiomics (TR) and manual ultrasound diagnosis (MD), the two-stage deep learning (DL) framework which fuses ROI segmentation with U-Net and classification with MobileNetV2 along with the clinical-radiomic feature fusion outperformed both FLD and gallstones diagnosis.

#### 3.4.1.Performance for fatty liver disease

Test set (1085 subjects): The DL model achieved an AUC of 0.93, accuracy of 90.2%, a recall score of 88.5%, and specificity of 89.8%. The contrary occurred with the TR model, giving an AUC of 0.82 and MD reached an AUC of 0.79 (Fig 3). Statistical comparisons using the DeLong test confirmed that the DL model ’ s AUC was significantly higher than both TR (P<0.01) and MD (P<0.05). Paired chi-squared testing revealed the DL model ’ s accuracy (P = 0.006) was statistically superior to MD, while McNemar tests demonstrated significant advantages in recall (P=0.009) and specificity (P=0.032) (Figure 6). The F1-score of the DL model (0.89) was also significantly higher than that of MD (0.76, P=0.007 via independent t-test), reflecting a balanced improvement in precision and recall critical for clinical screening.

#### 3.4.2.Performance for gallstones

In the case of gallstones, the DL model reached an AUC of

0.89, an accuracy of 87.6% and a recall of 85.3% (Fig 3). The TR model and MD group have lower AUC values, which are 0.80 and 0.76, respectively, and the DeLong test indicates obvious differences compared with the DL model and MD (P<0.05). Paired chi-squared testing showed the DL model’ s accuracy was significantly higher than MD (P=0.008), and McNemar tests indicated superior recall (P=0.012) and specificity (P=0.035) (Fig 6). The DL model ’ s F1-score (0.86) was also significantly higher than MD (0.73,P=0.009).

#### 3.4.3.Confusion matrix and multi-indicator validation

Confusion matrix Analysis (Fig 4) proved that the DL Model is reliable. For FLD the model had 414 correct positive identifications (TP=88.5%) out of 468 positive cases identified, 554 of 617 negative (TN=89.8%) cases correctly identified, 63 false positives (10.2%), and 54 false negatives (11.5%). For gallstones, 41 of 48 positive cases were correctly detected (TP=85.4%), and 910 of 1037 negative cases were accurately classified (TN=87.8%), with 127 false positives (12.2%) and 7 false negatives (14.6%). Box plot comparisons (Fig 5) illustrated that the DL model outperformed TR and MD across all key metrics (accuracy, recall, specificity, F1-score), with consistent statistical significance (P<0.05 for all comparisons vs. MD), confirming its robustness for hepatobiliary disease screening in seafarers.

## 4.Discussion

The diagnostic modality used in this study, which is ultrasonography, is a well-known reliable and common tool used in routine screening to find out fatty liver and gallstones[7]. For FLD, key ultrasonic indicators include enhanced near-field liver echo, blurred intrahepatic ductal structures, and attenuated far-field echo—findings that ensure accurate detection of both typical and heterogeneous forms of the disease[5][7]. For gallstones, the classic ultrasonic triad (hyperechoic foci, posterior acoustic shadowing, and positional mobility) enables consistent identification of most cases, with additional capacity to detect associated conditions

like cholecystitis[5][7]. Notably, chronic cholecystitis frequently coexists with biliary calculi, while acute calculous cholecystitis requires prompt clinical intervention[5][11]. This established diagnostic framework provided the foundation for our assessment of seafarer hepatobiliary health.

From our investigation we discovered that there was quite a high percentage of FLD occurrence at around 43.13%, much higher than that for the general population. Critically, this incidence demonstrated a strong linear positive correlation with body mass index (BMI) ($r=0.98$, $t = 6.83$, $P<0.05$), with chi-squared analysis confirming that FLD, intrahepatic lipid deposition, and hyperlipidemia rates increased proportionally with elevated BMI ($P<0.005$). This underscores BMI control and lipid management as core strategies for FLD prevention in this cohort.

But male seafarer's gallstone incidence was about 4.40% lower than the average of the general population. This difference could potentially be because there are more women than men with gallstones, so they are the base level overall. Chi-squared testing identified statistically significant associations between gallstone formation and gallbladder wall thickening / roughness, hyperlipidemia, and hypercholesterolemia ($P<0.005$)—consistent with established links between cholesterol metabolism and biliary stone formation[5][7].

These things have to be seen from within the special working environment of seafarers. They're on a ship with unique health problems: an off-and-on work-break schedule, the feeling of being separated from the mainland, not much medical help to look out for hard-to-see sicknesses, and a boring food that's mostly frozen meat that doesn't have many fruits and veggies. These factors likely contribute to the elevated FLD risk observed, particularly through their impact on BMI and lipid profiles.

Based on these insights, we propose targeted hepatobiliary disease prevention strategies for seafarers:

(1)Seafarers have to take offboard physical exam in time on account of scarce onboard medical facilities and risk of missed occult symptom.

(2)Diet should be improved so as to reduce dependence on frozen foods and boost consumption of fresh vegetables and fruits.

(3)We need strengthened health education concentrating on risks, prevention, and treatments for hepatobiliary diseases so our alert rises up. Carry out regular screenings and interventions so as to enable an early detection and handling situation.

(4)Controlling food intake, increasing physical activity more working condition improvement to reduce working stress, etc. all have significance on maintaining health livers and other organs.

In addition to the abovementioned preventive measures, technological progress in diagnostic imaging presents opportunities for transforming seafarer hepatobiliary health maintenance, in particular, radiomics and its combination with artificial intelligence (AI) especially deep learning. Radiomics, first proposed by Lambin et al. in 2012, quantifies high-dimensional features from multi-modal imaging data to enhance diagnostic precision[12], and its sonomics subfield aligns directly with the ultrasound-based approach of our study[13]. However, the true breakthrough lies in AI-driven radiomics, which addresses key limitations of traditional radiomics and forms the core of our technical innovation—findings complemented by comparative analyses in Fig 3 and 4.

Traditional radiomics uses manual feature design and extraction(e.g., histogram of graylevel,wavelet transform)[14][15]. it is tedious work and hard to be reused – from Fig 3 we can see the traditional model got the lowest diagnostic accuracy compared to seafarer because those traditional models are based on manually designed features. In contrast, AI-powered radiomics (primarily deep learning) revolutionizes this workflow by integrating neural networks into end-to-end feature learning and model training[12]. As detailed in our methodology, the deep learning pipeline includes critical AI-specific steps: data augmentation (rotations, flipping, noise injection) to mitigate small-sample biases common in occupational health datasets, and automated feature extraction that eliminates human subjectivity. These AI-driven advantages—automatic feature learning, high abstraction of complex imaging patterns, robust noise resistance, and cross-task transferability—directly translated to superior performance (Fig 3，4): our deep learning model achieved AUCs of 0.93 for FLD and 0.89 for gallstones, outperforming both traditional radiomics and clinical expert assessments. This aligns with broader trends in medical AI, where self-learning algorithms unlock hidden imaging correlates of disease that manual methods miss[12].

In terms of seafarers, it directly addresses their specific clinical challenges through this integration of AI-radiomics The model's robust performance in Fig 4 is needed for a shipboard setting where there may be limited specialist access. This model could give some non-expert personnel an early rough assessment. When integrated with seafarer-specific data—work schedules, dietary records, and stress metrics—AI can generate personalized risk profiles: for example, flagging crew with high BMI and abnormal ultrasound textures (identified via the model's learned features) as high FLD risk. More importantly, the model's ability to detect subtle imaging changes (superior to traditional methods, Fig 3) enables early identification of preclinical FLD or gallstone precursors, allowing targeted interventions before symptomatic progression. For acute conditions like biliary colic, AI-driven predictive analytics could even alert high-risk individuals based on longitudinal imaging and clinical data, facilitating timely off-board care. As seafarer health datasets expand, the model's transferability will enable refinement for specific subcohorts (e.g., long-haul vs. coastal seafarers), further enhancing its clinical utility. AI-powered radiomics transforms static ultrasound screening into a dynamic, personalized health management tool tailored to the maritime environment.

**Author contributions** Huiming Kang: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization. Xiaoqing Li: Investigation, Data curation, Resources, Clinical sample and ultrasound image collection. Mingyue Feng: Formal statistical analysis, Clinical data normalization, Data curation. Wei Wang: Validation, Radiological image gold-standard annotation, Writing – review & editing. Shidi Liu: Image preprocessing, Radiomic feature extraction, Data curation.

Shaocong Liang: Software, U-Net and MobileNetV2 model training optimization, Performance validation. Cuiliu Zhou: Investigation, Baseline clinical data verification, Laboratory indicator collation. De Kang: Clinical study guidance, Supervision, Writing – review & editing. Lingling Chen: Supervision, Project administration, Funding acquisition, Writing – review & editing.All authors have read and approved the final version of the manuscript.

**Funding** No funding was received for this work.

**Data availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

[1] ACHARYA U R, FAUST O, MOLINARI F, et al. Ultrasound-based tissue characterization and classification of fatty liver disease: A screening and diagnostic paradigm. Knowledge-Based Systems, 2015, 75: 66-77.

[2] DONG P, GUO S F, CHAI Y H, et al. Prevalence survey and risk factor analysis of cholelithiasis in Xingtai area. Journal of Hepatobiliary Surgery, 2020, 28(1): 30-32.

[3] ZHENG J, LIANG G X. Clinical value analysis of ultrasonic diagnosis of gallbladder stones. China Health Standard Management, 2019, 10(20): 117-119.

[4] NON-ALCOHOLIC FATTY LIVER DISEASE PREVENTION AND TREATMENT GUIDELINE WRITING GROUP. Guideline for the prevention and treatment of non-alcoholic fatty liver disease. Infectious Disease Information, 2018, 31(5): 393-402+420.

[5] LIANG P, RAN H T. Medical Ultrasound Imaging. Beijing, China: People's Medical Publishing House, 2022: 182-183, 191-194.JIANG Y X, HE W. Ultrasound Medicine. Beijing, China: People's Medical Publishing House, 2022: 189.

[6] HOU X K, TAO Y. Advanced Ultrasound Medicine for Senior Physicians. Beijing, China: Peking Union Medical College Press, 2021: 260-271.

[7] ZHU J R, GAO R L, ZHAO S P, et al. Guideline for the prevention and treatment of dyslipidemia in Chinese adults. Chinese Circulation Journal, 2016, 31(10): 937-953.

[8] GE J B, XU Y J, WANG C. Internal Medicine. Beijing, China: People's Medical Publishing House, 2018: 391, 757-764.

[9] SUN M Y, CHEN W. Guideline for medical nutritional therapy of overweight/obesity in China. Peking Union Medical College Hospital Journal, 2022, 13(2): 255-262.

[10] CHEN X P, WANG J P, ZHAO J Z. Surgery. Beijing, China: People's Medical Publishing House, 2023: 438-445.

[11] TIAN J, LI C M, DONG D. Fundamentals of Radiomics. Beijing, China: Science Press, 2022: 3-13.

[12] MAYERHOEFER M E, MATERKA A, LANGS G, et al. Introduction to Radiomics. Journal of Nuclear Medicine, 2020, 61(4): 488-495.

[13] HE W W, HUANG Y X, SHI X L, et al. Identifying a distinct fibrosis subset of NAFLD via molecular profiling and the involvement of profibrotic macrophages. Journal of Translational Medicine, 2023, 21(1): 448.

[14] LEAVLINE E J, SUTHA S, SINGH D A A G. Fast multiscale directional filter bank-based speckle mitigation in gallstone ultrasound images. Journal of the Optical Society of America A, 2014, 31(2): 283.