

# Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images

Michał Byra<sup>1</sup>  · Grzegorz Styczynski<sup>2</sup> · Cezary Szmigielski<sup>2</sup> · Piotr Kalinowski<sup>3</sup> · Łukasz Michałowski<sup>4</sup> · Rafał Paluszakiewicz<sup>3</sup> · Bogna Ziarkiewicz-Wróblewska<sup>4</sup> · Krzysztof Zieniewicz<sup>3</sup> · Piotr Sobieraj<sup>2</sup> · Andrzej Nowicki<sup>1</sup>

Received: 3 February 2018 / Accepted: 31 July 2018 / Published online: 9 August 2018  
© The Author(s) 2018

## Abstract

**Purpose** The nonalcoholic fatty liver disease is the most common liver abnormality. Up to date, liver biopsy is the reference standard for direct liver steatosis quantification in hepatic tissue samples. In this paper we propose a neural network-based approach for nonalcoholic fatty liver disease assessment in ultrasound.

**Methods** We used the Inception-ResNet-v2 deep convolutional neural network pre-trained on the ImageNet dataset to extract high-level features in liver B-mode ultrasound image sequences. The steatosis level of each liver was graded by wedge biopsy. The proposed approach was compared with the hepatorenal index technique and the gray-level co-occurrence matrix algorithm. After the feature extraction, we applied the support vector machine algorithm to classify images containing fatty liver. Based on liver biopsy, the fatty liver was defined to have more than 5% of hepatocytes with steatosis. Next, we used the features and the Lasso regression method to assess the steatosis level.

**Results** The area under the receiver operating characteristics curve obtained using the proposed approach was equal to 0.977, being higher than the one obtained with the hepatorenal index method, 0.959, and much higher than in the case of the gray-level co-occurrence matrix algorithm, 0.893. For regression the Spearman correlation coefficients between the steatosis level and the proposed approach, the hepatorenal index and the gray-level co-occurrence matrix algorithm were equal to 0.78, 0.80 and 0.39, respectively.

**Conclusions** The proposed approach may help the sonographers automatically diagnose the amount of fat in the liver. The presented approach is efficient and in comparison with other methods does not require the sonographers to select the region of interest.

**Keywords** Nonalcoholic fatty liver disease · Ultrasound imaging · Deep learning · Convolutional neural networks · Hepatorenal index · Transfer learning

## Introduction

The nonalcoholic fatty liver disease, diagnosed in a large number of obese patients, is the most common liver abnormality [1]. It is defined as the accumulation of fat in more than 5% of liver cells. This disease is associated with increased risk of hepatic cirrhosis and hepatocellular carcinoma, but it is also influencing higher cardiovascular morbidity and mortality in affected patients [2, 3]. Liver biopsy is the reference standard for direct liver steatosis quantification in hepatic tissue samples [4]. However, biopsy is a costly and invasive procedure that carries a high risk of serious complications, commonly including pain, bleeding and in rare cases, death [4]. Therefore, liver biopsy is not considered to be an easy, optimal way to assess and follow-up the progress of com-

✉ Michał Byra  
mbyra@ippt.pan.pl

<sup>1</sup> Department of Ultrasound, Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland

<sup>2</sup> Department of Internal Medicine, Hypertension and Vascular Diseases, Medical University of Warsaw, Warsaw, Poland

<sup>3</sup> Department of General, Transplant and Liver Surgery, Medical University of Warsaw, Warsaw, Poland

<sup>4</sup> Department of Pathology, Center for Biostructure Research, Medical University of Warsaw, Warsaw, Poland

mon liver diseases. Noninvasive liver imaging methods such as computed tomography, magnetic resonance imaging or ultrasound (US) have been intensively investigated [5]. US may be the preferred modality for screening liver steatosis because of its non-invasiveness, low cost and wide availability.

Up to date various approaches have been proposed to assess the level of steatosis in liver using US [6]. Among them, the hepatorenal sonographic index (HI) is considered to be highly efficient and simple [7, 8]. The HI method is based on comparison of the liver echogenicity to that of the right kidney cortex. Normal liver and renal tissues show similar echogenicity. However, in the presence of steatosis, the liver tissue brightness is higher than the kidney brightness. The ultrasound-based diagnostic results may depend on skills and experience of physicians performing the examination, type of ultrasound machine and even on US image settings [9, 10]. This operator dependence makes the comparison of results difficult and limits wider practical application of this important imaging technique. Another approach to liver steatosis assessment employs texture analysis. According to the review paper on liver image analysis [6], the gray-level co-occurrence matrix (GLCM) algorithm is the most frequently used method for liver disease characterization [11]. GLCMs provide useful information about spatial gray-level dependencies in an image. Texture patterns of US images arise from the interference of backscattered US waves on tissue microstructures. The GLCM-based approaches to liver steatosis classification using US images have been proposed in several papers [12–15].

Nowadays new algorithms for image analysis are intensively studied, including deep learning. These machine learning methods let the computers automatically develop useful features for classification. The usefulness of convolutional neural networks (CNNs) has been reported in solving various medical image analysis problems [16, 17]. CNNs transform input images with convolutional filters into a single decision variable as an output that usually indicates the input image label. However, to successfully train a CNN, usually a large amount of input data are required. This issue limits the practical applications of deep models in medical image analysis, since the available medical image datasets are usually small. Therefore, as a solution, various transfer learning techniques have been proposed [18]. Instead of building a completely new model from scratch, it is possible to use a model developed for another problem. The usefulness of a pre-trained model depends on its ability to adjust to images outside the original training dataset. In the case of medical image analysis, the implementation of transfer learning techniques has been reported in several papers [19–22].

The aim of this paper is to develop a deep learning model for steatosis level assessment based on US liver B-mode images and to compare it with the HI and the GLCM

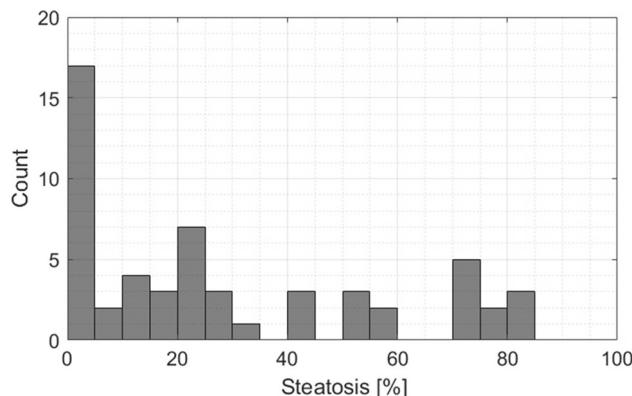
techniques. The US data analyzed in this study were collected from severely obese patients evaluated before bariatric surgery. We used a pre-trained CNN to extract features based on B-mode images. Next, using the neural features, we employed the support vector machine (SVM) algorithm to classify images containing fatty liver. Aside of fatty liver classification, it is clinically relevant to quantify the grade of liver steatosis. For this task, we used the extracted features and the Lasso regression method. In both cases, liver biopsy results served as a reference. The performance of the proposed approach was compared with the HI and the GLCM methods.

This paper is organized in the following way. First, we describe the patient group and the data acquisition routines. It is presented how to calculate the HI- and the GLCM-related features using liver US images. Next, our deep learning solution to fatty liver assessment is described. We show how to apply the transfer learning to extract CNN-based features using B-mode liver images. Next, we employ the CNN- and the GLCM-based features to perform fatty liver disease classification and to assess the level of steatosis. Results are presented and evaluated. Finally, we discuss the advantages and disadvantages of the applied methods.

## Materials and methods

### Clinical dataset

Our study involved 55 severely obese patients (mean age  $40.1 \pm 9.1$ , mean BMI  $45.9 \pm 5.6$ , 20% of males) admitted for bariatric surgery (laparoscopic sleeve gastrectomy). The ultrasound data were acquired in the Department of Internal Medicine, Hypertension and Vascular Diseases, Medical University of Warsaw, Poland, during preoperative cardiac echocardiographic evaluation, 1–2 days before the surgery. The study was approved by the Ethical Committee at the Medical University of Warsaw, and all patients gave informed consent for echocardiography and abdominal ultrasound examination. Each patient underwent a wedge liver biopsy during the bariatric surgery as a part of the routine protocol implemented at the Department of General, Transplant and Liver Surgery, Medical University of Warsaw, Poland [23]. Tissue sample was extracted from the subcapsular part of the left liver lobe. The histopathological assessment was performed by a single pathologist following the recommendations of the Clinical Research Network [24]. The level of steatosis was defined based on the percentage of hepatocytes with fatty infiltration. The fatty liver was defined to have more than 5% hepatocytes with steatosis. The number of patients with fatty liver was equal to 38. The steatosis level distribution across the population of patients is depicted in Fig. 1.



**Fig. 1** Histogram of steatosis level across the population of patients in the study group

The ultrasound data were acquired using the GE Vivid E9 Ultrasound System (GE Healthcare INC, Horten, Norway) equipped with a sector probe operating at 2.5 MHz. The default general abdominal preset with harmonic imaging was used. The resolution of B-mode images was equal to  $434 \times 636$  (pixel size of  $0.373 \text{ mm} \times 0.373 \text{ mm}$ ), see Fig. 2. For each patient, a sequence of B-mode images, corresponding to one heart beat, was acquired and stored on the workstation (EchoPac PC software, GE Healthcare INC, Norway). The image loops were saved in DICOM format for further off-line processing. Due to motion, the speckle patterns and relative position of the liver and the kidney were slightly different across the images in each sequence. Moreover, the number of images in sequences was not constant. It depended, for example, on the number of focal zones and the scanner frame rate. For each sequence, ten consecutive images were used for further processing. Finally, the dataset contained 550 B-mode images. We decided to analyze image sequences rather than single B-mode images selected by the physician. It was a way of data augmentation, which enabled us to provide more diverse data to the models.

The dataset described above can be downloaded via the Zenodo repository (<https://doi.org/10.5281/zenodo.1009146>). The dataset repository includes sequences of B-mode images and the biopsy results. The provided dataset could be useful for researchers interested in fatty liver imaging. It should be noted that during the acquisition of the data with the cardiac probe, we recorded the images with the kidney on the left side of the screen. For convenience of those researchers who are used to kidney on the right side of the image, we provide in Fig. 2 the example images following the standard convention. In the case of the dataset, the images were provided with the left sided kidney arrangement as recorded during the image acquisition.

## Hepatorenal index

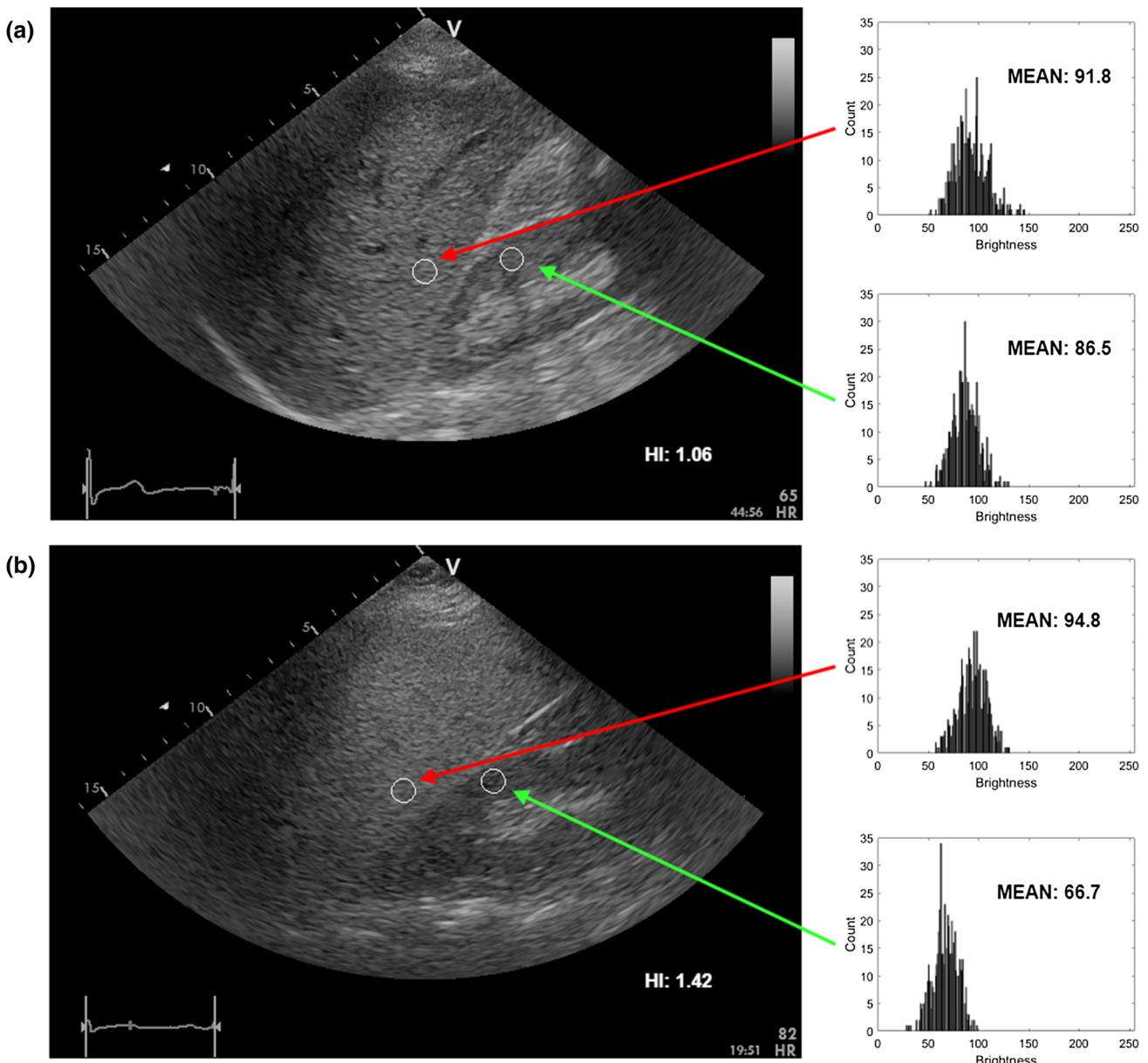
The HI is defined as the ratio of average brightness level of the liver and the kidney cortex. Generally, the HI is expected to increase with the steatosis level. In our study, the HI was determined by a physician with experience in ultrasonography and echocardiography research acquisition [25]. The physician was blind to biopsy results. In the first step, a single scan frame from the B-mode sequence was selected by the physician. Next, two regions of interest (ROIs) corresponding to the liver and the kidney cortex were specified. The ROI selection is illustrated in Fig. 2. Care was taken to select liver and kidney ROI in the middle part of the image sector, side by side at the same depth. If infeasible due to suboptimal image quality, liver ROI was selected above kidney ROI with the shortest distance possible. The ROI was determined by using circular method with the radius of the circle equal to 5 mm. In each case, the ROI was as uniform as possible. Regions of non-uniform speckle pattern, vessels or ducts were omitted during the ROI selection procedure. The ratio between the average brightness levels in the ROIs was determined with Matlab software (MathWorks INC, USA) using histogram analysis, see Fig. 2.

## GLCM-based features

GLCM-based features were extracted following a similar approach proposed in [12–14]. The same liver ROIs were employed for analysis as in the case of the HI method. However, instead of the circular regions, we used square regions with side length of 10 mm. For each ROI nine different GLCMs were calculated considering angles between 0, 45 and 135, and path distances of 1, 2 and 3 [12]. Next, for each GLCM the following texture features were extracted: maximum probability, uniformity, entropy, dissimilarity, contrast, inverse difference, inverse difference moment and correlation [26].

## CNN-based features

CNN features were extracted using the Inception-ResNet-v2 CNN implemented in Keras [27, 28]. Calculations were performed in Python. The model was pre-trained on the ImageNet dataset [29]. This CNN includes a mixture of residual inception modules followed by grid reduction modules and has achieved state-of-the-art accuracy on the ImageNet dataset that contains 1.2 millions of labeled images [28]. Sample labels include animals, fruits and daily necessities. This dataset was successfully used for transfer learning in several medical imaging applications [16]. In our case, the CNN features were extracted using entire US images. Minimal pre-processing was applied to liver B-mode images, and the non-relevant data, such as frame number, were removed.



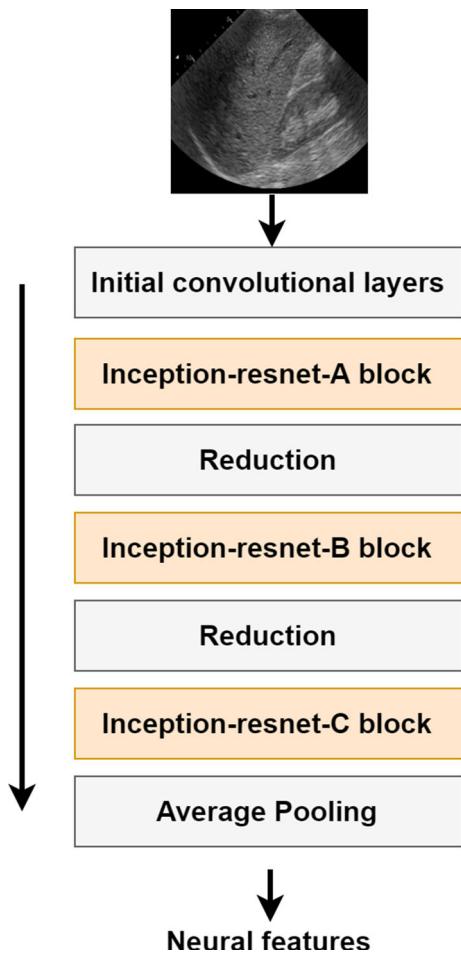
**Fig. 2** Liver B-mode images and the ROIs selected for HI calculation, **a** steatosis level of 3% and **b** 25%, respectively

Images were resized using the bi-cubic interpolation algorithm to the resolution originally designed for the network. Each liver image was given as the network input, and the corresponding neural features were extracted from the average pooling layer. Feature extraction procedure is depicted in Fig. 3. Next, zero-variance features were removed.

### Classification and evaluation

We utilized the SVM algorithm to perform the classification of fatty liver images [30] using the GLCM- or the CNN-based features. Methods that exclude outliers were used to normalize the features. The validation scheme is presented

in Fig. 4. Patient-specific leave-one-out cross-validation (LOOCV) was applied to evaluate the classification. In each case, the test set consisted of 10 images from the same patient and the training set contained 540 images from the remaining 54 patients. For each training set, fivefold cross-validation and grid search were applied to indicate the optimal SVM classifier hyperparameters and the best kernel. To address the problem of class imbalance, the SVM hyperparameter  $C$  of each class was adjusted inversely proportional to that class frequency in the training set. Label 1 indicated the image containing a fatty liver and label –1 otherwise. After the training phase, the a posteriori probabilities were calculated for each image in the test set and the results were averaged to



**Fig. 3** Illustration of feature extraction using the Inception-ResNet-v2 model [28]

obtain the final a posteriori probability related to the examined liver. Next, these probabilities were used to calculate the receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) was used for evaluation of the classification performance. We applied the Delong statistical

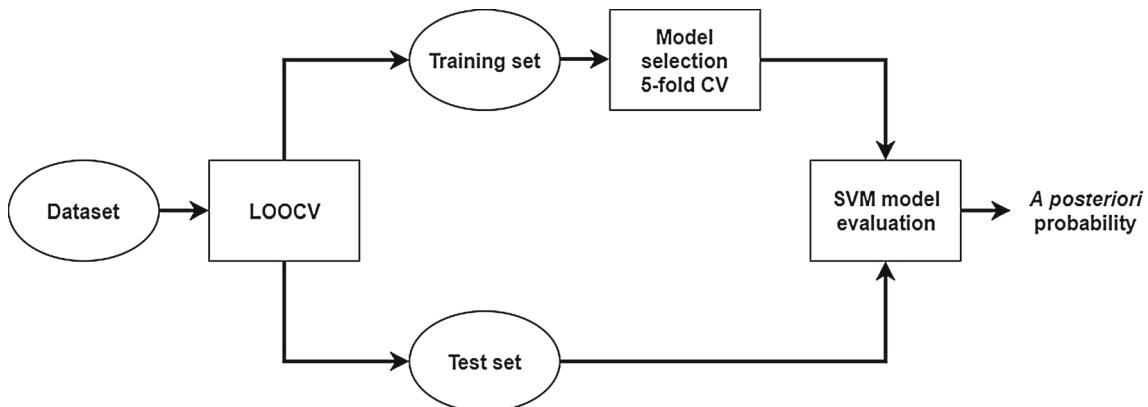
test implemented in the pROC package in R to compare the AUC values obtained for all methods [31, 32].

To assess the level of steatosis, we employed the Lasso regression method. The same validation scheme was applied as in the case of the classification, but the steatosis level was estimated instead of the a posteriori class probability. Spearman correlation coefficients (SCCs) were calculated to assess the relation between the steatosis level, the models' outputs and the HI parameter. Moreover, the SCCs between the models' outputs and the HI parameter were determined. Next, the linear regression algorithm was used to relate the steatosis level and the HI parameter. All regression models were compared using the Meng test implemented in the cocor package in R [33, 34].

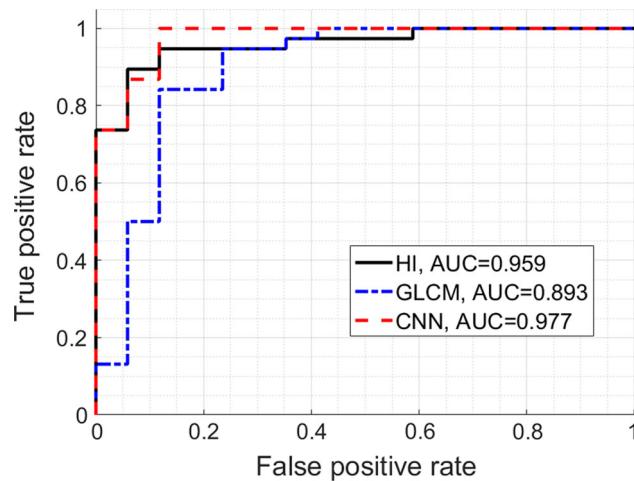
## Results

The classification performance related to the HI parameter and the SVM classifiers is presented in Fig. 5. All methods achieved good classification performance. The highest AUC value, equal to 0.977, was obtained for the CNN-based classification. The performance of the HI-based approach was slightly lower, with AUC value equal to 0.959. However, the Delong test indicated that this difference in AUC values was not statistically significant ( $p$  value > 0.05). The AUC value obtained for the GLCM-based approach was equal to 0.892 and was significantly lower than that for the CNN-based method (Delong test  $p$  values < 0.05). The performance summary is presented in Table 1. Sensitivity, specificity and accuracy were calculated using the threshold corresponding to the ROC curve point, which was the closest to the upper left corner of the ROC plot, point (0, 1) [35].

Figure 6 shows the usefulness of the Lasso regression method and the HI parameter in steatosis level assessment. The SCCs obtained for the Lasso algorithms utilizing the CNN- and the GLCM-based features were equal to 0.78 ( $p$



**Fig. 4** The validation pipeline



**Fig. 5** The ROC curves for the HI method ( $AUC=0.959$ ), the GLCM algorithm ( $AUC=0.893$ ) and the classifier developed using CNN features ( $AUC=0.977$ )

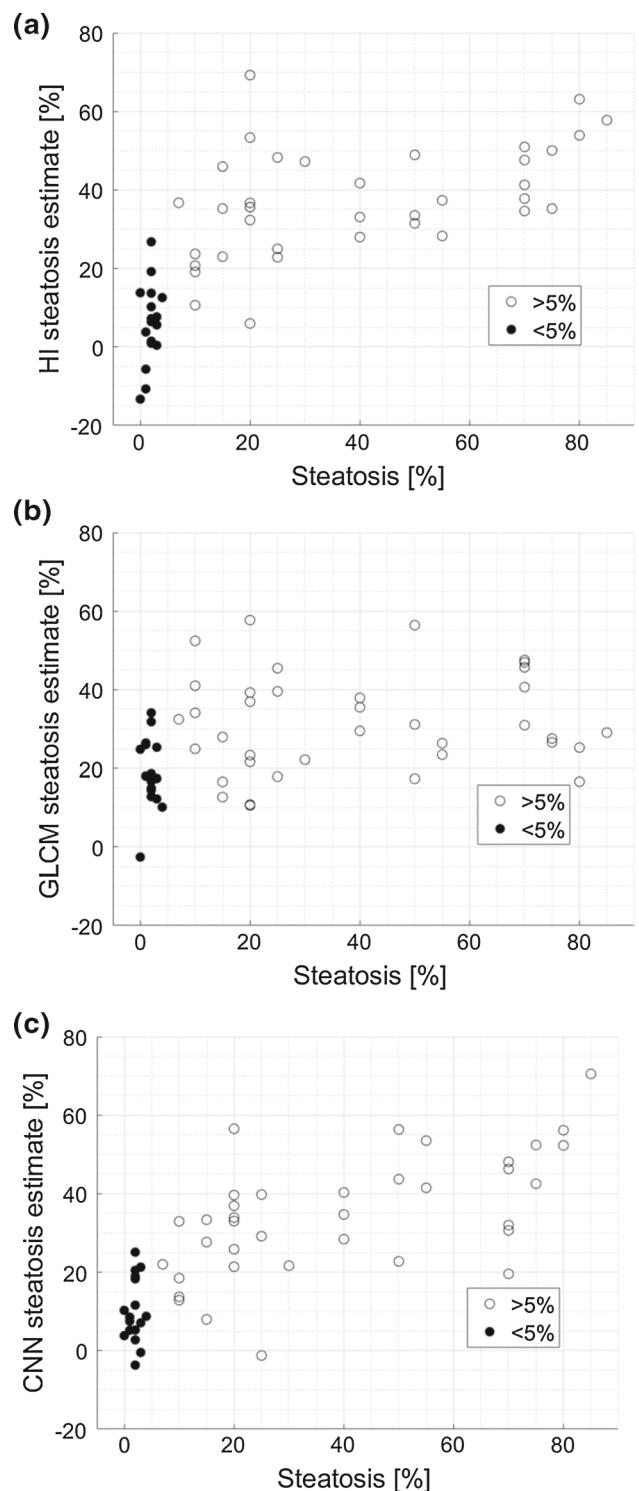
**Table 1** Classification performance summary

Method	AUC	Sensitivity	Specificity	Accuracy
Hepatorenal index	$0.959 \pm 0.044$	0.895	0.941	0.909
GLCM	$0.893 \pm 0.059$	0.842	0.882	0.854
CNN	$0.977 \pm 0.021$	1	0.882	0.963

value  $<0.001$ ) and 0.39 ( $p$  value  $<0.05$ ), respectively. The SCC for the HI parameter was equal to 0.80 ( $p$  value  $<0.05$ ). The SCC between the CNN-based approach and the HI parameter was equal to 0.78 ( $p$  value  $<0.05$ ). The difference between the Lasso algorithm and the HI method correlation coefficients was statistically insignificant ( $p$  value  $>0.05$ ). Figure 7 illustrates the agreement between these two methods.

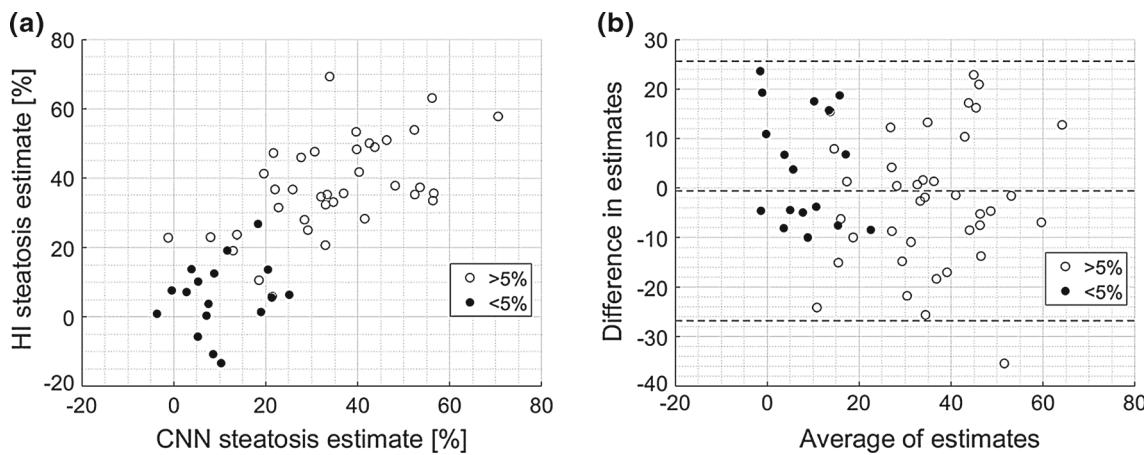
## Discussion

Ultrasound imaging is the most commonly applied imaging modality. Our study confirms that the HI parameter is a good predictor of steatosis level in liver. It is simple to calculate and efficient. Our results are in a good agreement with other studies reporting the usefulness of the HI parameter. We obtained high values of the AUC and the SCC parameters, which were equal to 0.959 and 0.80, respectively. The AUC values reported for the HI method ranges from 0.76 [36] to 0.99 [8]. However, the papers commonly report different ranges of the HI parameter and different optimal cutoffs for the fatty liver classification. [7, 8, 36–39]. This issue illustrates the ambiguity related to the HI-based fat assessment. The performance of the GLCM-based approach was worse with the AUC and the SCC equal to 0.893 and 0.39, respec-



**Fig. 6** The usefulness of **a** the HI parameter ( $SCC=0.80$ ), **b** GLCM-based features ( $SCC=0.39$ ) and **c** the CNN-based features ( $SCC=0.78$ ) in steatosis level assessment

tively. Low value of the SCC parameter suggests that the GLCM-based features are not efficient for the steatosis level assessment. The obtained AUC value is in agreement with



**Fig. 7** **a** The relation between the HI parameter and the Lasso regression ( $\text{SCC} = 0.78$ ) and **b** the corresponding Bland–Altman plot

the results reported in the previous studies that employed GLCM-based features [12, 14, 15]. In [12, 15] the authors reported AUC values of around 0.8. In [14] the accuracy of around 0.8 was reported. In [13] the authors achieved high AUC value of 0.96. However, in this study the cross-validation was not applied and the authors used the same dataset to develop and evaluate the classifiers what could result in overfitting.

Our study shows the feasibility of using deep learning for the liver steatosis assessment. Although we used a small dataset containing only 550 images from 55 patients, these data were sufficient to develop a well performing classifier with transfer learning. The AUC value in the case of the fatty liver classification was equal to 0.977. According to Table 1, the obtained performance was higher than in the case of the HI method. Moreover, the CNN-based approach achieved significantly better results than the GLCM-based approach. The CNN features were useful and enabled efficient training of the classification and regression models. Good performance of the CNN-based approach was expected. In our study, we did not train the network from scratch, instead the pre-trained CNN was used for feature extraction. This model was developed using the ImageNet dataset containing 1.2 million labeled images of various objects. The HI calculation includes two convolutional operations (spatial averaging), which should be supposedly learned by the CNN to perform well on the ImageNet dataset. These two operations have to be conducted in the liver and the kidney, so the network has to detect these tissues first. The appearance of the liver with respect to surrounding tissues is important for efficient steatosis assessment.

In the case of the liver steatosis assessment, the obtained SCC, equal to 0.78, was slightly lower than the SCC calculated for the HI parameter, which was equal to 0.80. However, this difference was not statistically significant. Both regression models performed well, except for the patients with

severe steatosis. In this case, the estimated values of steatosis were slightly too small. This may be due to the dataset, which was too small to build an accurate regression model. Moreover, the transfer learning in this case may not be efficient enough to capture the dependence between the input images and the liver steatosis level. Nevertheless, the proposed approach should be considered to be good, especially since the results were obtained in an automated process. Figure 7a illustrates the relation between the Lasso regression method and the HI parameter. In this case, the SCC was equal to 0.78, indicating high degree of correlation. According to the Bland–Altman plot in Fig. 7b, the average bias in estimates is low.

Although the performance of the proposed method was only slightly better than the performance obtained using the HI parameter, the proposed approach has several advantages that illustrate its clinical value. First, our method can be considered as an integrated computer-aided diagnosis system. It is operator independent and does not require ROI selection in comparison with the HI method and the GLCM-based approach. Next, the proposed method efficiently utilizes sequences of US images to assess the level of steatosis, while the approaches proposed in the literature commonly employs only one US image to conduct classification [6]. However, there are several issues related to our work. First of all, the ROI selection is operator dependent and has impact on calculation of the HI parameter and the GLCM-based features. For proper estimation of the HI parameter, the physician has to select ROIs in the liver and the kidney. These ROIs have to be as uniform as possible to omit the regions of blood vessels, ducts or other structures in the organs. In our study we focused on machine learning and did not examine observer variability, the ROIs were determined by a single physician. The obtained results may differ between observers [9, 10]. Second, all employed methods are to some extent scanner dependent. B-mode image intensities

can be modified by using different image reconstruction and processing algorithms, what may affect the feature extraction and consequently the classification. This is a general issue encountered in studies that aim to develop US-based computer-aided diagnosis systems. Image quality (speckle patterns and boundary visibility) depends on scanner settings. The Inception-ResNet-v2 network utilized in our study was trained using the ImageNet dataset that contains images recorded under slightly different lighting conditions. Therefore, we believe that the impact of image reconstruction algorithms implemented in the US scanners should be lower for the proposed approach than in the case of the HI- and the GLCM-based methods. We would like to investigate this problem in the future in two ways. First, it would be interesting to acquire raw ultrasound data and investigate how the image reconstruction algorithms impact the feature extraction from the CNN [40]. Second, we are going to acquire B-mode images of the same liver using different scanner settings and investigate whether the model can learn features for classification that are independent of scanner settings. To make the assessment scanner independent, it would be interesting to employ the quantitative US techniques. These methods are used to estimate various physical properties of the tissue, such as the attenuation or scattering characteristics [41]. Quantitative US techniques can be used to create parametric maps that serve as an additional source of information on investigated tissue in comparison with standard B-mode images [42, 43]. Those maps may serve as a more proper input to the CNN than regular B-mode images. The usefulness of quantitative US techniques in liver steatosis assessment has been reported in several studies [13, 44, 45]. In the future, we plan also to acquire more data and investigate various approaches to model development.

## Conclusions

In this paper we proposed a CNN-based approach to steatosis level assessment utilizing B-mode ultrasound images. The model was developed using data acquired in obese patients undergoing wedge liver biopsy during bariatric surgery. Our approach is efficient and operator independent. Moreover, it outperforms the HI- and the GLCM-based classification.

## Compliance with ethical standards

**Conflict of interest** The authors do not have any conflict of interest.

**Ethical statements** All procedures performed in studies involving human participants were in accordance with the Ethical Standards of the Medical University of Warsaw and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Beeman SC, Garbow JR (2018) Imaging and metabolism. Springer, New York
2. Adams LA, Scott Harmsen JL, Charatcharoenwitthaya P, Enders FB, Therneau T, Angulo P (2010) Nonalcoholic fatty liver disease increases risk of death among patients with diabetes: a community-based cohort study. *Am J Gastroenterol* 105:1567–1573
3. Adams LA, Anstee QM, Tilg H, Targher G (2017) Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gut* 66:1138–1153
4. Tapper EB, Lok ASF (2017) Use of liver imaging and biopsy in clinical practice. *N Engl J Med* 377:756–768
5. Luapuadat AM, Jianu IR, Ungureanu BS, Florescu LM, Gheonea DI, Sovaila S, Gheonea IA (2017) Non-invasive imaging techniques in assessing non-alcoholic fatty liver disease: a current status of available methods. *J Med Life* 10:19–26
6. Bharti P, Mittal D, Ananthasivan R (2017) Computer-aided characterization and diagnosis of diffuse liver diseases based on ultrasound imaging: a review. *Ultrason Imaging* 39:33–61
7. Marshall RH, Eissa M, Bluth EI, Gulotta PM, Davis NK (2012) Hepatorenal index as an accurate, simple, and effective tool in screening for steatosis. *Am J Roentgenol* 199:997–1002
8. Webb M, Yeshua H, Zelber-Sagi S, Santo E, Brazowski E, Halpern Z, Oren R (2009) Diagnostic value of a computerized hepatorenal index for sonographic quantification of liver steatosis. *Am J Roentgenol* 192:909–914
9. Cengiz M, Senturk S, Cetin B, Bayrak AH, Bilek SU (2014) Sonographic assessment of fatty liver: intraobserver and interobserver variability. *Int J Clin Exp Med* 7:5453–5460
10. Strauss S, Gavish E, Gottlieb P, Katsnelson L (2007) Interobserver and intraobserver variability in the sonographic assessment of fatty liver. *Am J Roentgenol* 189:320–323
11. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* 6:610–621
12. Andrade A, Silva JS, Santos J, Belo-Soares P (2012) Classifier approaches for liver steatosis using ultrasound images. *Procedia Technol* 5:763–770
13. Gaitini D, Baruch Y, Ghersin E, Veitsman E, Kerner H, Shalem B, Yaniv G, Sarfaty C, Azhari H (2004) Feasibility study of ultrasonic fatty liver biopsy: texture vs. attenuation and backscatter. *Ultrasound Med Biol* 30:1321–1327
14. Pavlopoulos S, Kyriacou E, Koutsouris D, Blekas K, Stafylopatis A, Zoumpoulis P (2000) Fuzzy neural network-based texture analysis of ultrasonic images. *IEEE Eng Med Biol Mag* 19:39–47
15. Rivas EC, Moreno F, Benitez A, Morocho V, Vanegas P, Medina R (2015) Hepatic Steatosis detection using the co-occurrence matrix in tomography and ultrasound images. In: 20th symposium on signal processing, Images and Computer Vision (STSIVA), Bogota, pp 1–7. <https://doi.org/10.1109/STSIVA.2015.7330417>
16. Litjens GJS, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sanchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
17. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for

- computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35:1285–1298
- 18. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. *J Big Data* 3:9
  - 19. Byra M (2018) Discriminant analysis of neural style representations for breast lesion classification in ultrasound. *Biocybern Biomed Eng* 38:684–690
  - 20. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep* 6:244–254
  - 21. Huynh BQ, Li H, Giger ML (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging* 3:034501
  - 22. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35:1299–1312
  - 23. Kalinowski P, Paluszakiewicz R, Ziarkiewicz-Wroblewska B, Wroblewska T, Remiszewski P, Grodzicki M, Krawczyk M (2017) Liver function in patients with nonalcoholic fatty liver disease randomized to roux-en-y gastric bypass versus sleeve gastrectomy: a secondary analysis of a randomized clinical trial. *Ann Surg* 266:738–745
  - 24. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, Ferrell LD, Liu YC, Torbenson MS, Unalp-Arida A, Yeh M, McCullough AJ, Sanya AJ (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 41:1313–1321
  - 25. Styczynski G, Milewska A, Marczevska M, Sobieraj P, Sobczynska M, Dabrowski M, Kuch-Wocial A, Szmagelski C (2016) Echocardiographic correlates of abnormal liver tests in patients with exacerbation of chronic heart failure. *J Am Soc Echocardiogr* 29:132–139
  - 26. Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can J Remote Sens* 28:45–62
  - 27. Chollet F (2015) Keras. <https://github.com/fchollet/keras>
  - 28. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the thirty-first AAAI conference on artificial intelligence, vol 4
  - 29. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 248–255
  - 30. Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology* 27:1–27
  - 31. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
  - 32. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M (2011) proc: an open-source package for r and s+to analyze and compare roc curves. *BMC Bioinform* 12:17
  - 33. Diedenhofen B, Musch J (2015) cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE* 10:e0131499
  - 34. Meng XL, Rosenthal R, Rubin DB (1992) Comparing correlated correlation coefficients. *Psychol Bull* 111:172–175
  - 35. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874
  - 36. Wang JH, Hung CH, Kuo FY, Eng HL, Chen CH, Lee CM, Lu SN, Hu TH (2013) Ultrasonographic quantification of hepatic–renal echogenicity difference in hepatic steatosis diagnosis. *Dig Dis Sci* 58:2993–3000
  - 37. Ballestri S, Romagnoli D, Nascimbeni F, Francica G, Lonardo A (2015) Role of ultrasound in the diagnosis and treatment of nonalcoholic fatty liver disease and its complications. *Expert Rev Gastroenterol Hepatol* 9:603–627
  - 38. Chauhan A, Sultan LR, Furth EE, Jones LP, Khungar V, Sehgal CM (2016) Diagnostic accuracy of hepatorenal index in the detection and grading of hepatic steatosis. *J Clin Ultrasound* 44:580–586
  - 39. Mancini M, Prinster A, Annuzzi G, Liuzzi R, Giacco R, Medagli C, Cremoni M, Clemente G, Maurea S, Riccardi G, Rivellese AA, Salvatore M (2009) Sonographic hepatic–renal ratio as indicator of hepatic steatosis: comparison with 1 h magnetic resonance spectroscopy. *Metab Clin Exp* 58:1724–1730
  - 40. Nguyen A, Yosinski J, Clune J (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 427–436
  - 41. Oelze ML, Mamou J (2016) Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. *IEEE Trans Ultrason Ferroelectr Freq Control* 63:336–351
  - 42. Sadeghi-Naini A, Suraweera H, Tran WT, Hadizad F, Bruni G, Rastegar RF, Curpen B, Czarnota GJ (2017) Breast-lesion characterization using textural features of quantitative ultrasound parametric maps. *Sci Rep* 7:13638
  - 43. Byra M, Nowicki A, Wroblewska-Piotrkowska H, Dobruch-Sobczak K (2016) Classification of breast lesions using segmented quantitative ultrasound maps of homodyned K distribution parameters. *Med Phys* 43:5561–5569
  - 44. Lin SC, Heba E, Wolfson T, Ang B, Gamst A, Han A, Erdman JW, O'Brien WD, Andre MP, Sirlin CB, Loomba R (2015) Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat using a new quantitative ultrasound technique. *Clin Gastroenterol Hepatol* 13:1337–1345
  - 45. Paige JS, Bernstein GS, Heba E, Costa EAC, Fereirra M, Wolfson T, Gamst AC, Valasek MA, Lin GY, Han A, Erdman JW, O'Brien WD, Andre MP, Loomba R, Sirlin CB (2017) A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology determined steatosis grade in adult nonalcoholic fatty liver disease. *Am J Roentgenol* 208:168–177