Analysis of B-Scan Ultrasonography using Neural Networks to predict Risk of Fibrosis in Patients with Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD)

Josefine Stansch, Kien Vu Trung[1], Valentin Blank[2], Jakob Kather[3], Moritz Herzog[3], Thomas Berg[4], Johannes Wiegand[4], Thomas Karlas[1]

[1]Department of Medicine II, Division of Gastroenterology, Leipzig Medical Center, 04103 Leipzig
[2]Department of Medicine I, Interdisciplinary Ultrasound Department, Halle Medical Center, 06102 Halle
[3]Else Kroener Fresenius Center for Digital Health, Technical University Dresden, 01307 Dresden
[4]Department of Medicine II, Division of Hepatology, Leipzig Medical Center, 04103 Leipzig

BACKGROUND

Metabolic dysfunction-associated steatotic liver disease (MASLD) is globally relevant and expected to be the main cause of end-stage liver disease in the coming decades. Liver biopsy as the traditional reference standard for fibrosis risk stratification in patients with MASLD is not suitable for everyday clinical practice given the prevalence of the disease. Non-invasive screening methods such as liver elastography and serum-based risk indices are therefore preferred and are also explicitly recommended for risk stratification in the MASLD setting by current guidelines, although elastography devices are not accessible to many patients. Traditional ultrasound diagnostics of the liver is not considered to be sufficiently specific for fibrosis detection due to subtle morphological characteristics and examiner-dependent interpretation of the images. This is especially relevant when steatosis is present. Therefore, ultrasound ist currently not recommended for risk stratification, although the method is widely available.In this study, we evaluated whether an AI model can identify patients with MASLD at increased fibrosis risk when trained with B-scan liver images and liver elastography values as reference.

The aim of the study was to evaluate automated analysis of B-scan datasets with neural networks to classify patients with MASLD at increased risk of fibrosis, using liver elastography based risk categories as the reference standard.

METHODS

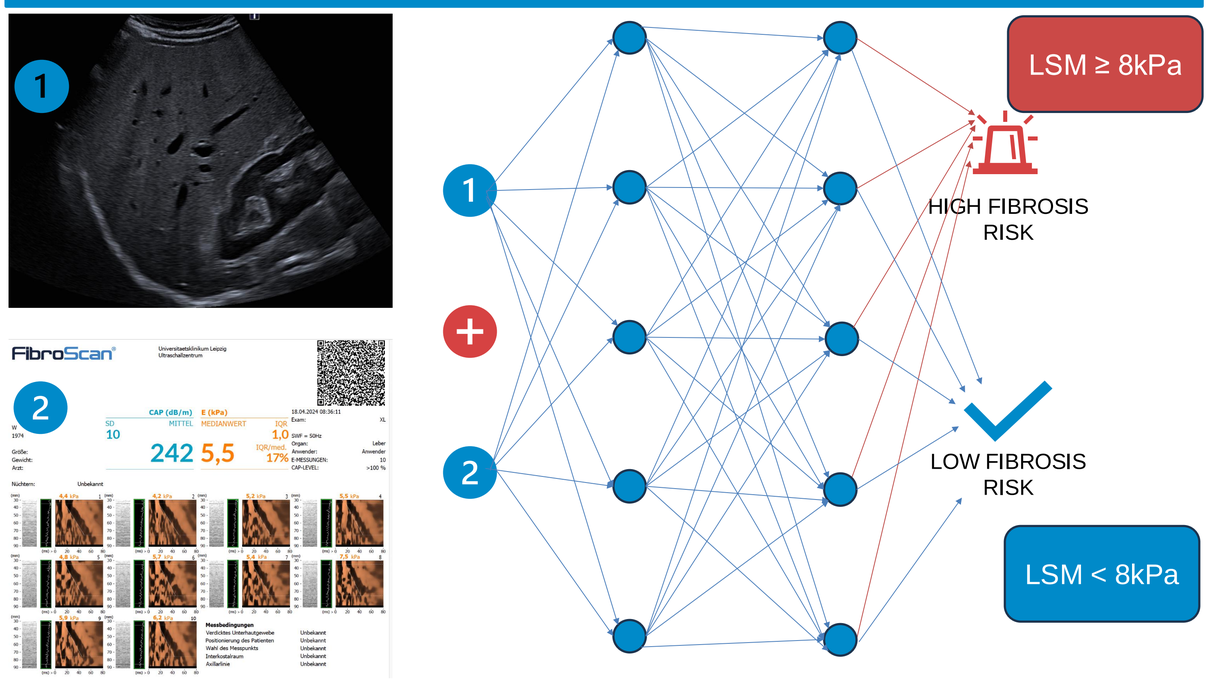

Figure 1- Simplified illustration of the applied deep learning algorithm

We systematically reviewed liver ultrasound images of well characterized study cohorts including 350 patients with MASLD (PMID: 33110165, 38702958, 34446354).
For this analysis we only included adult patients who met the current MASLD criteria, had undergone a valid Fibroscan® (liver elastography) examination and at least one high-quality ultrasound image of the hepatorenal region of the liver, using TOSHIBA and CANON Aplio ultrasound systems. If available, further images from different anatomic planes of the liver were also included.

Patients (n=205) were classified in two groups according to the liver Fibroscan® result (<8kPa indicating low risk and ≥8kPa indicating increased risk) and to the current guideline recommendations. The ultrasound images and the Fibroscan®-based risk classification were used to train a deep learning network, based on the pre-trained DenseNet121 (Figure 1).

Data augmentation and class weighting techniques were employed to address imbalances of the data. A range of model combinations were subjected to training to ascertain the most effective combination of pretrained network, activation function and optimizer. The stratified k-fold method was implemented to train a model that is as generalizable as possible with minimal overfitting.

The Receiver Operating Characteristic (ROC) curve and the Youden index characterized the diagnostic properties of the models (Figure 2). ROCs were compared using the DeLong method.

In addition, we compared the results of the AI-based classificiation with the judgment of three human expert ultrasound examiners. The McNemar test was used to statistically analyze the diagnostic properties, both for the individual examiner and in a pooled variant for the three examiners in relation to the AI.

Thomas.Karlas@medizin.uni-leipzig.de Josefine.Stansch@medizin.uni-leipzig.de

RESULTS

Variables	Total cohort		Low Fibrosis Risk		High Fibrosis Risk		
		n= 205		n= 153		n= 52	p-value
Number of images in total	636		478		158		
Number of females	104	(50.7%)	75	(49.0%)	29	(55.8%)	p=1.000
Age (years)	60.0	[51.0; 66.0]	60.0	[49.0; 66.0]	61	[52.0; 68.0]	p=0.378
BMI (kg/m2)	30.2	[26.6; 34.9]	28.90	[26.3; 34.1]	32.6	[29.7; 36.6]	p=0.002
Diabetes mellitus type 2	94	(45.9%)	73	(47.7%)	21	(40.4%)	p=1.000
Number of images	3.0	[2.0; 4.0]	3.0	[2.0; 4.0]	3.0	[2.0; 4.0]	p=0.598

Table 1- Baseline characteristics of the study cohort

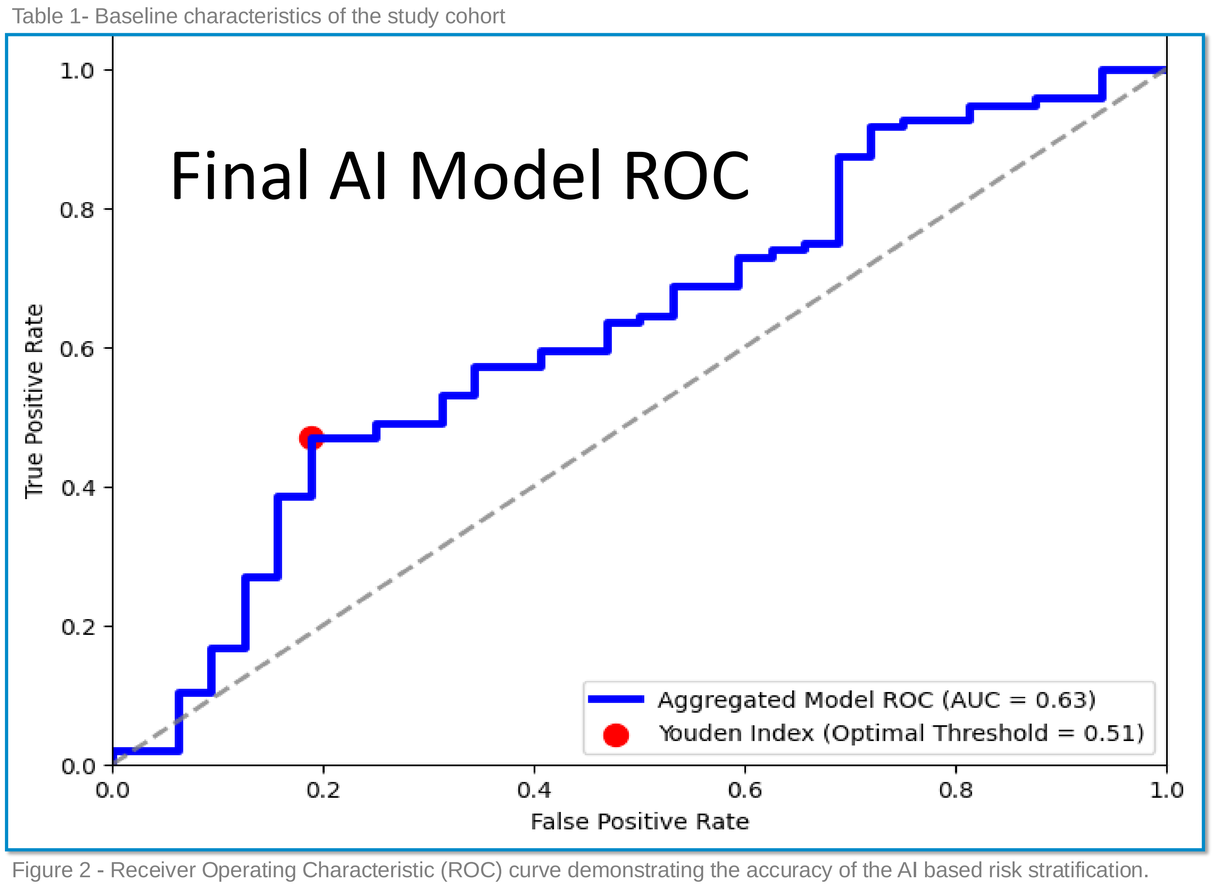

Figure 2 - Receiver Operating Characteristic (ROC) curve demonstrating the accuracy of the AI based risk stratification.

The final AI model revealed an accuracy of 0,75 and an ROC of 0,63 (Figure 2).
The comparison of the pooled assessment of the human experts and two different AI-based approaches revealed significant difference in the performances of the risk classification compared to human experts. The approach with highest overall accuracy performed significantly better than the experts' judgement ($p < 0,001$), whereas only little differences were seen when a Youden-Index-optimized approach was used.

	ACCURACY	PRECISION	SENSITIVITY	SPECIFICITY
Examiner 1	0,6 [0.51, 0.68]	0,22 [0.15, 0.29]	0,26 [0.18, 0.33]	0,71 [0.63, 0.79]
Examiner 2	0,63 [0.55, 0.71]	0,32 [0.24, 0.4]	0,45 [0.37, 0.54]	0,69 [0.61, 0.77]
Examiner 3	0,7 [0.62, 0.78]	0,27 [0.19, 0.34]	0,13 [0.07, 0.19]	0,89 [0.83, 0.94]
Examiners combined	0.64 [0.59, 0.69]	0.27 [0.23, 0.32]	0.28 [0.23, 0.32]	0.76 [0.72, 0.80]
AI version 1 [best approach with high specificity]	0.76 [0.69, 0.84]	0.57 [0.49, 0.66]	0.13 [0.07, 0.19]	0.97 [0.94, 1.00]
AI version 2 [optimized for Youden Index]	0.61 [0.53, 0.70]	0.33 [0.25, 0.42]	0.58 [0.49, 0.67]	0.62 [0.54, 0.71]

Table 2 – Statistical analysis comparing human evaluation and AI performance

CONCLUSION

- **AI-based liver ultrasound shows promise for liver fibrosis risk stratification.**
- **The approach demands validation and refinement in prospective cohorts.**
- **Perspectively, AI-based point-of-care ultrasound may contribute to optimized care of patients at risk for metabolic liver disease.**

REFERENCES

Kalapala, R., Rughwani, H., & Reddy, D. N. (2023). Artificial Intelligence in Hepatology- Ready for the Primetime. *Journal of clinical and experimental Hepatology*, *13*(1), 149–161.
Cao, L. L., Peng, M., Xie, X., Chen, G. Q., Huang, S. Y., Wang, J. Y., Jiang, F., Cui, X. W., & Dietrich, C. F. (2022). Artificial intelligence in liver ultrasound. *World Journal of Gastroenterology*, *28*(27), 3398–3409.
Boursier, J., Hagström, H., Ekstedt, M., Moreau, C., Bonacci, M., Cure, S., Ampuero, J., Nasr, P., Tallab, L., Canivet, C. M., Kechagias, S., Sánchez, Y., Dincuff, E., Lucena, A., Roux, M., Riou, J., Trylesinski, A., & Romero-Gomez, M. (2022). Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events. *Journal of Hepatology*, *76*(5), 1013–1020