Comparison of Artificial Intelligence and Eyeball Method in the Detection of Fatty Liver Disease

Evan J. Catron, MS1[1], Robert P. Passarelli, BS[1], Danielle Wilmes, MS3[1], Barry Wei, MS1[1], Thi M.U. Le, PhD[1], Ping Li, PhD[1], Wenjun Zhang, PhD[1], Jingmei Lin, MD, PhD[2], Marc L. Melcher, MD, PhD[3], Plamen V. Mihaylov, MD, PhD[1], Chandrashekhar A. Kubal, MD, PhD[1], Richard S. Mangus, MD, MS[1], Burcin Ekser, MD, PhD[1]

(1) Transplant Surgery, Department of Surgery, Indiana University School of Medicine; (2) Department of Pathology and Laboratory Medicine, Indiana University School of Medicine; (3) Transplant Surgery, Department of Surgery, Stanford University School of Medicine

Background: Quantification of liver fat content relies on visual microscopic inspection of liver biopsies by pathologists. Their percent macrosteatosis (%MaS) estimation is vital in determining donor liver transplantability; however, the eyeball method may vary between observers. Overestimations of %MaS can potentially lead to the discard of viable donor livers. We hypothesize that artificial intelligence (AI) could be helpful in providing a more objective and accurate measurement of %MaS.

Methods: Literature review identified HALO (image analysis) and U-Net (deep-learning) as high-accuracy AI programs capable of calculating %MaS in liver biopsies. We compared (i) an experienced pathologist's and (ii) a transplant surgeon's eyeball %MaS estimations from de-novo liver transplant (LT) biopsy samples taken 2h post-reperfusion to (iii) the HALO-calculated %MaS (Fig.1). 250 patients had undergone LT at Indiana University between 2020-2021, and 211 had sufficient data for inclusion. Each biopsy was digitized into 5 random non-overlapping tiles at 20x magnification (a total of 1,055 images). We used HALO software for analysis and set the minimum vacuole area to 10μm² to avoid the inclusion of microsteatosis. Microsteatosis was excluded by the pathologist and the surgeon by the eyeball method using the same 1,055 images. Each %MaS estimation was compared with early allograft dysfunction (EAD). EAD is defined by the presence of at least one of the following: INR >1.6 on postoperative day (POD) 7, total bilirubin >10mg/dL on POD7, or AST/ALT >2000IU/L within the first 7 days following LT.

Results: Of 211 LTs, 42 (19.9%) had EAD. The mean %MaS estimation of pathologist and transplant surgeon were 6.3% (SD: 11.9%) and 3.2% (SD: 6.4%), respectively. HALO yielded a significantly lower mean %MaS of 2.6% (SD: 2.6%) than the pathologist's eyeball method ($p<0.001$). The mean %MaS calculated by HALO was higher in EAD patients than in non-EAD ($p=0.032$), but this difference did not reach statistical significance in the pathologist's estimation ($p=0.069$).

Conclusions: Although mean %MaS measurements from all parties were mild (<10%), human eyeball estimations of %MaS were significantly higher than HALO's %MaS. The HALO-calculated %MaS differed significantly between the EAD and non-EAD LTs which might suggest a possible correlation between the AI's steatosis analysis and EAD outcomes. However, pathologic variables other than %MaS (necrosis or cholestasis) should be included in future analyses to determine whether %MaS is the dominant parameter predicting EAD. AI is a promising tool to quantify liver steatosis and will help pathologists and transplant surgeons predict liver transplant viability.

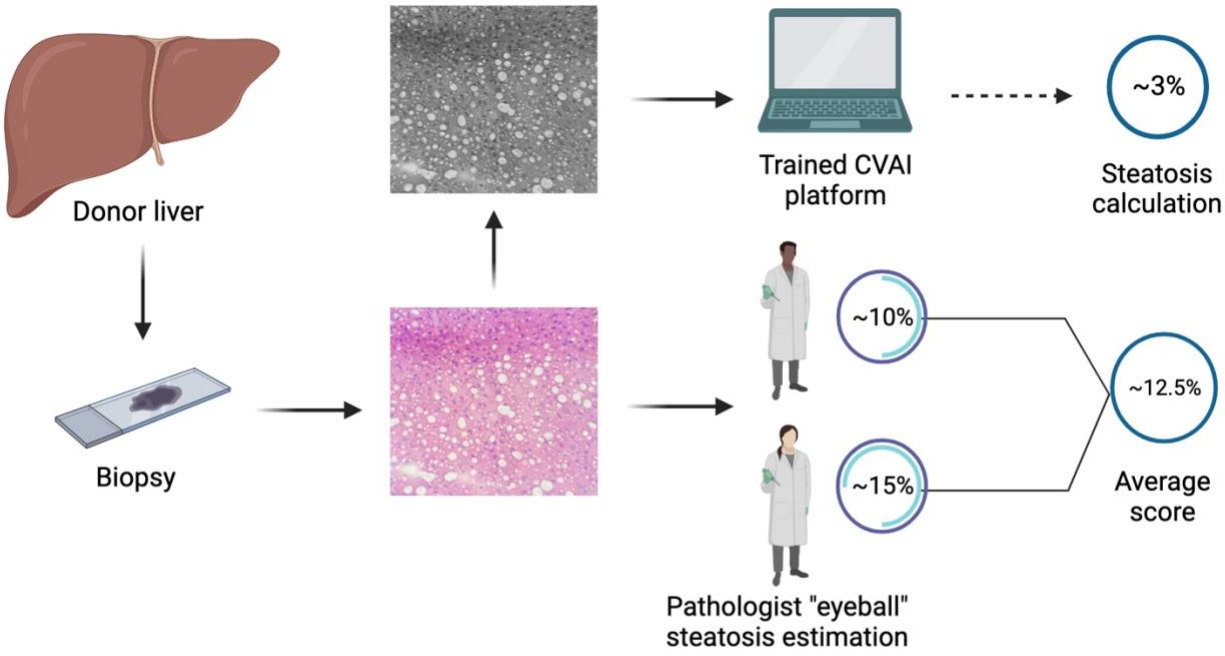

Fig. 1. Steps to compare AI %MaS calculation of biopsy to pathologist visual estimation. The same donor liver biopsies that were inspected by pathologists were digitized into high-resolution images and uploaded to AI platforms that calculate %MaS by fat vacuole identification. Most studies have found that %MaS estimated by pathologists is consistently higher and more variable than AI calculations.