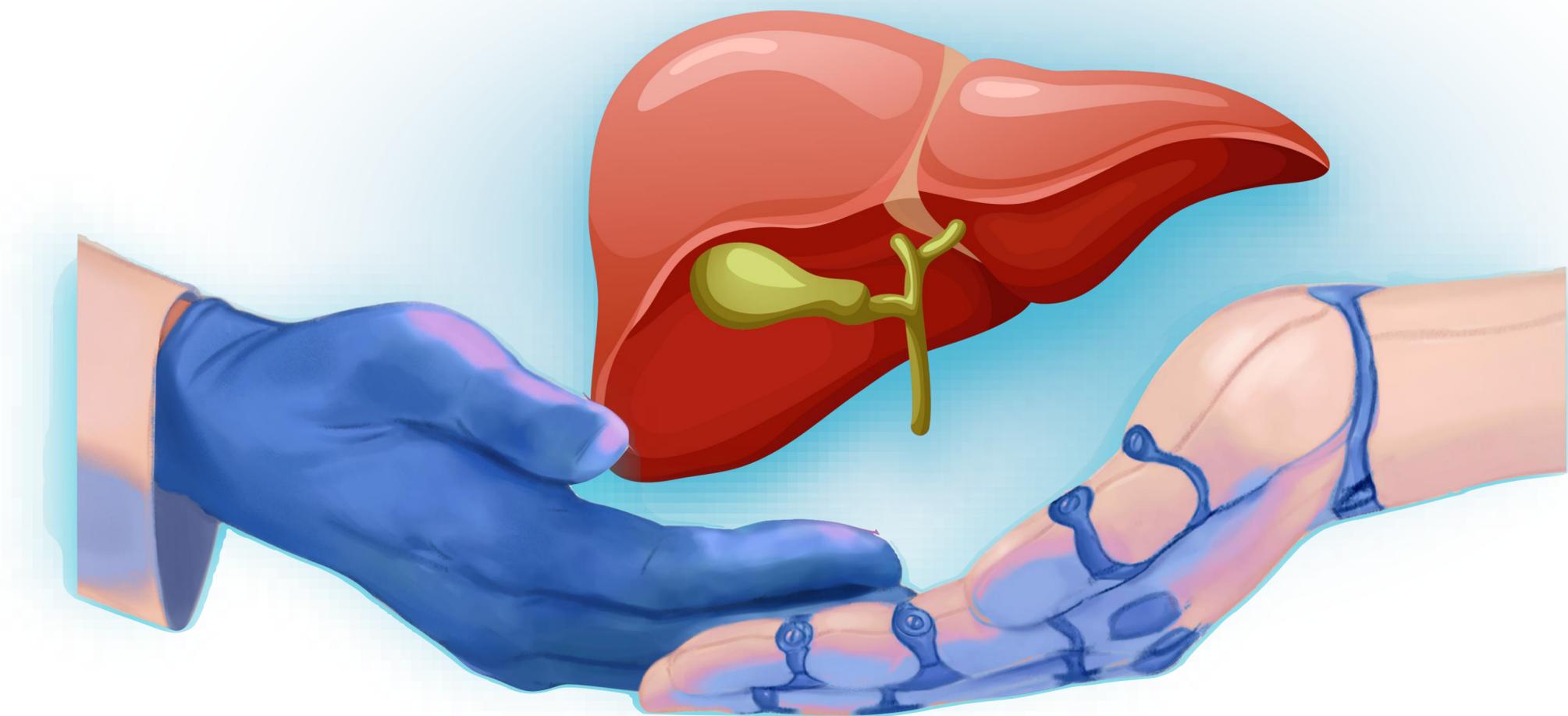


Use of AI in addressing unmet clinical needs for MASLD



Jana Lipkova

Block 1: Liver biopsy as “imperfect gold standard” in Clinical Trials

- ▶ Histopathology assessment is required for enrollment and evaluations of endpoints in clinical trials (CTs)

1) Inter- & intra-rater variability of biopsy assessment [1,2]:

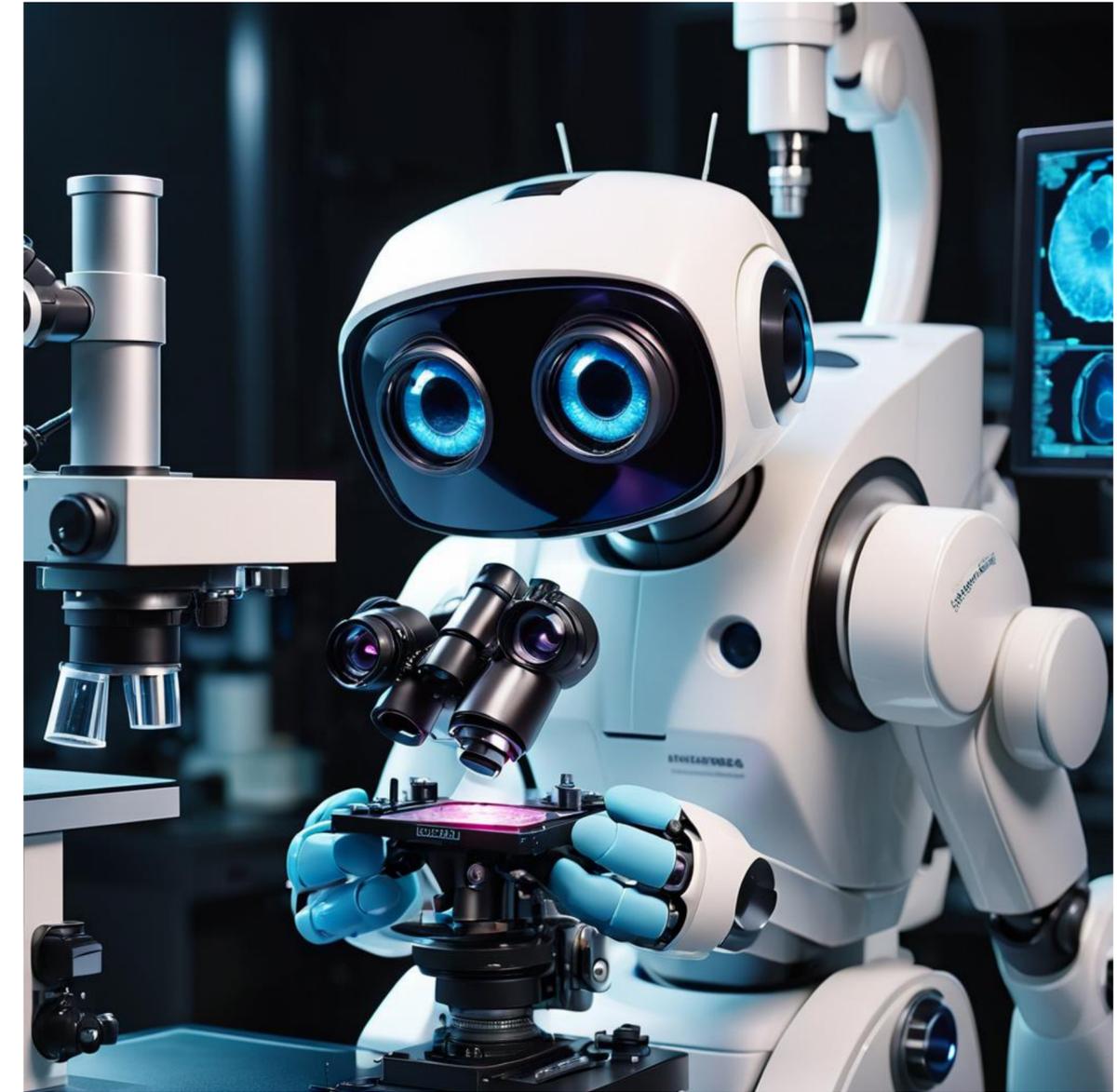
- **46.3%** patients in MASH clinical trials **did not meet enrollment criteria** upon re-evaluation by 2nd pathologist
- Variability in assessment of endpoints and inclusion criteria **reduce study power of CTs by 50%**

2) Ordinal grading system for continuous disease severity:

- Errors around the boundaries of the grades/stage
- Not sufficient to capture early changes in treatment response
- Difficult to quantify early changes and their extent

Unmet Needs:

- #1. Objective, automated & reproducible tool for biopsy assessment
- #2. Refined grading system to allow detection and quantification of early changes in treatment response



[1] Davison, B. A. et al. "Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials." *J. Hepatol.* (2020).

[2] Brunt, E. M. et al. "Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD". *J. Hepatol.* (2022)

AI-Solutions: Automated and Robust Biopsy Assessment

nature medicine



Article

<https://doi.org/10.1038/s41591-024-03172-7>

AI-based automation of enrollment criteria and endpoint assessment in clinical trials in liver diseases

Received: 2 May 2023

Accepted: 3 July 2024

Published online: 07 August 2024

 Check for updates

Janani S. Iyer, Dinkar Juyal, Quang Le, Zahil Shanis, ...,
Ilan Wapinski & Andrew Beck

Clinical trials in metabolic dysfunction-associated steatohepatitis (MASH, formerly known as nonalcoholic steatohepatitis) require histologic scoring for assessment of inclusion criteria and endpoints. However, variability in interpretation has impacted clinical trial outcomes. We developed an artificial intelligence-based measurement (AIM) tool for scoring MASH histology (AIM-MASH). AIM-MASH predictions for MASH Clinical Research

AIM-MASH: AI-based Measurement tool for scoring MASH histology

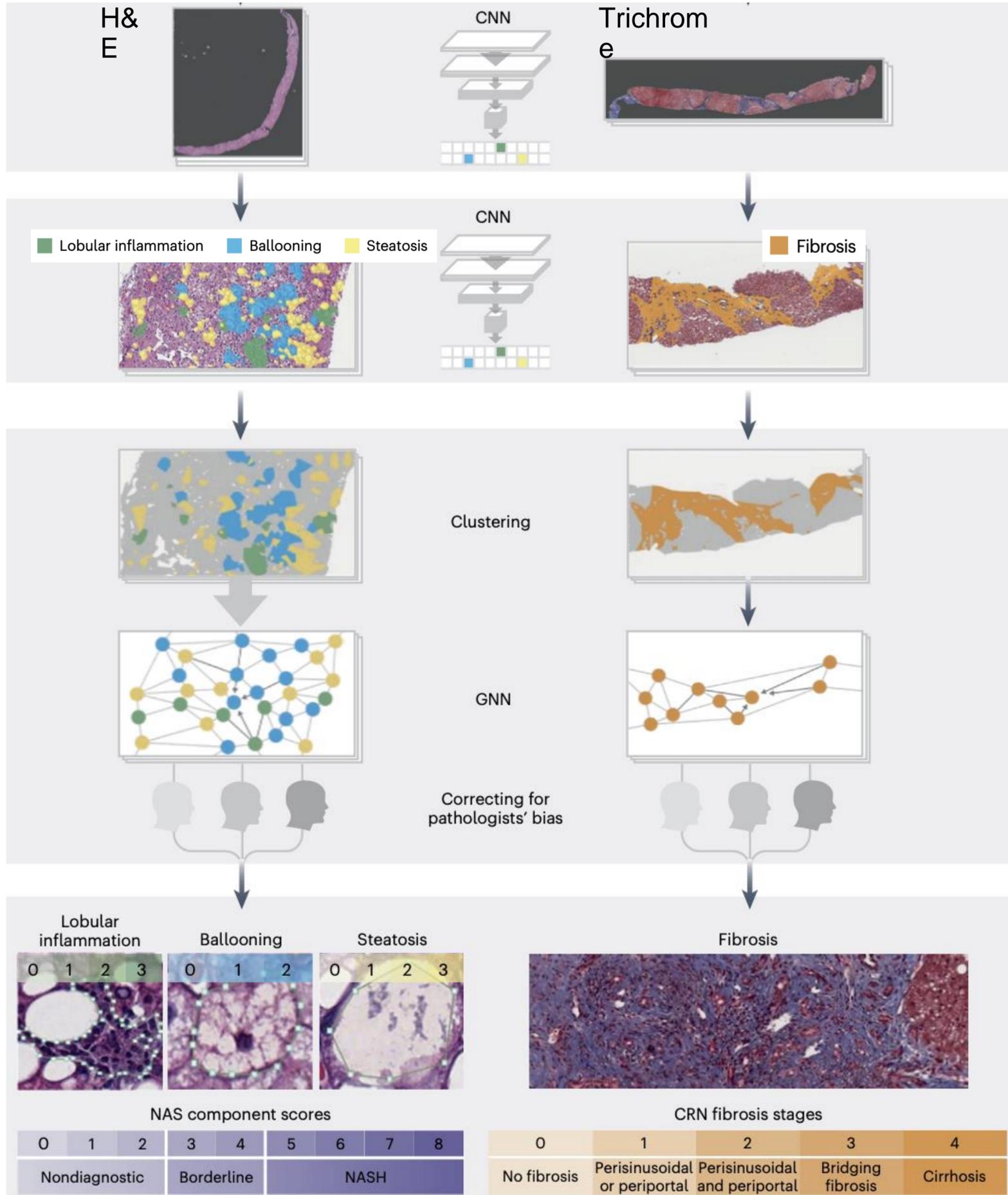
AI-based model for automated grading / staging of Steatosis, Lobular inflammation, hepatocellular Ballooning and Fibrosis from H&E and Masson Trichrome WSIs

Multistage AI Model

- a) Input: H&E and MT WSI
- b) Segment tissue regions with feature relevant for MASH scoring
- c) For each histology feature the model predicts a continuous scores reflecting the feature severity
- d) Continuous score is mapped to the corresponding ordinal MASH CRN grade/stage

Study Design

- ▶ Developed on **8,747 H&E** and **7,660 Masson trichrome WSIs**
- ▶ **6 MASH clinical trials** (6 phase 2b and phase 3)
- ▶ Annotations from **59 MASH expert pathologists**



AI-enables robust assessment of enrollment criteria, endpoints & treatment response

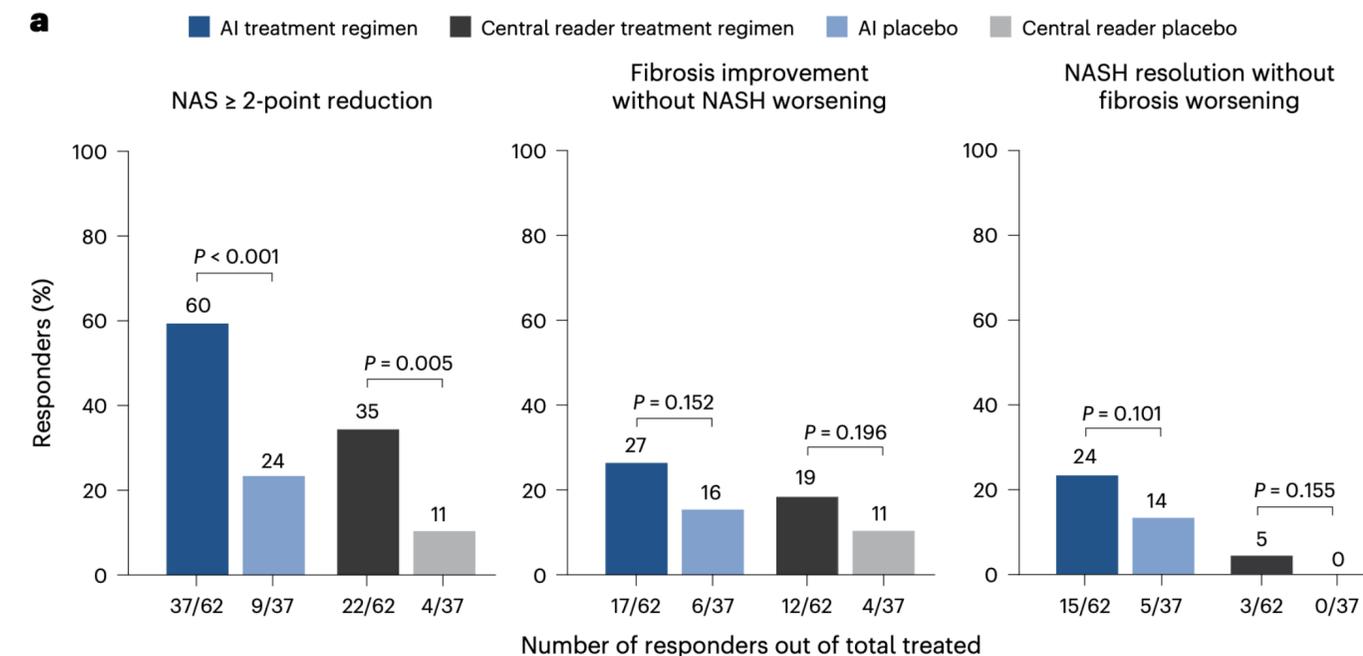
1) AI reached better agreement with pathologists' consensus than any individual pathologist:

- For grading and staging of all histology features
- For assessment trial enrollment criteria and endpoints
- 100% reproducibility (when tested on the same images)

Histologic feature	AIM-MASH versus consensus	Mean pathologist versus consensus
Lobular inflammation	0.67 (0.64–0.71)	0.64 (0.62–0.67)
Ballooning	0.70 (0.66–0.73)	0.66 (0.63–0.69)
Steatosis	0.74 (0.71–0.77)	0.69 (0.66–0.72)
Fibrosis	0.62 (0.58–0.65)	0.59 (0.57–0.62)

2) AI-based assessment of drug efficacy (ATLAS phase 2b)

- ▶ **Original trial (manual biopsy) no statically significant difference** in response rates between the treated vs placebo group
- ▶ **AI-based scoring:**
 - bigger difference between placebo vs treater group for all endpoints
 - **statistically significant difference for one endpoint**



→ By removing the inter-rater variability, AI can help improve reliability of CTs

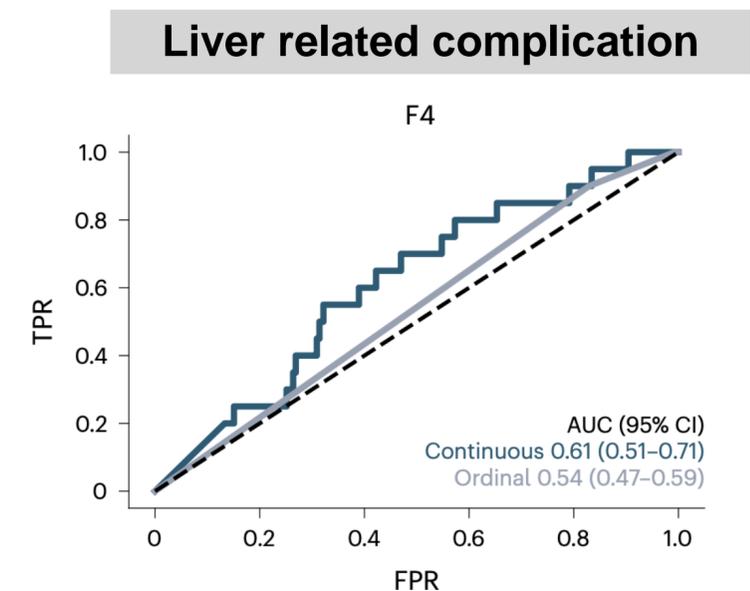
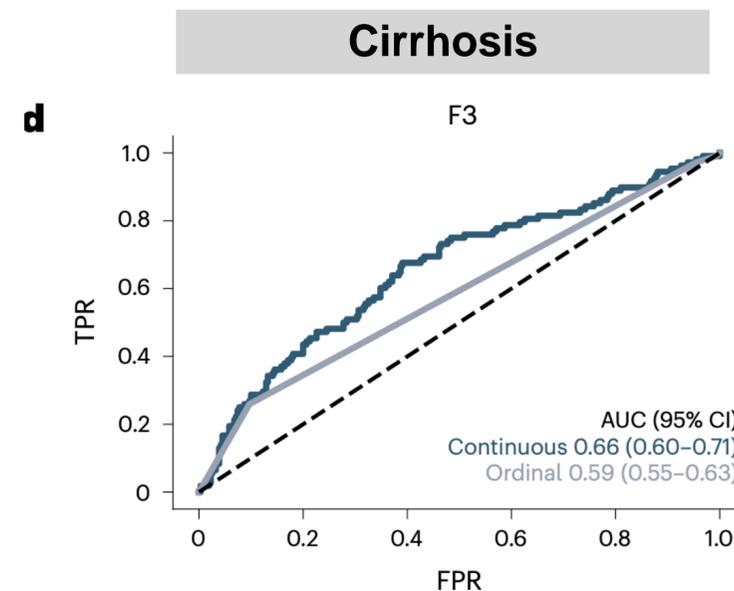
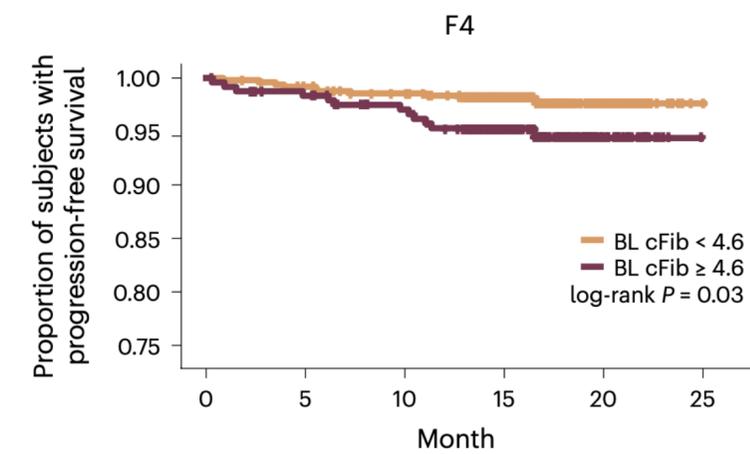
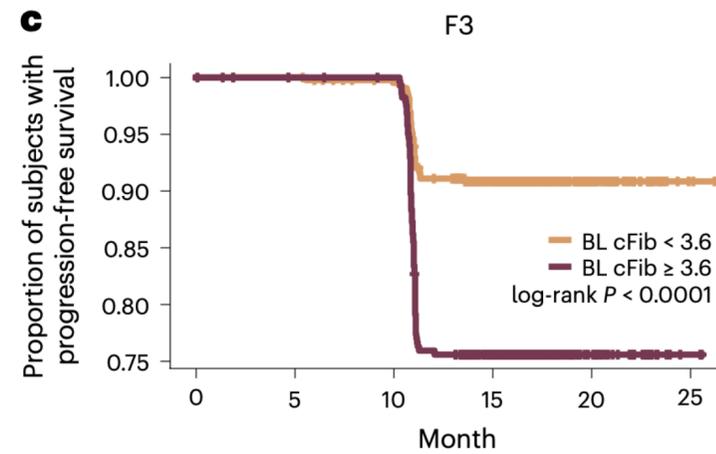
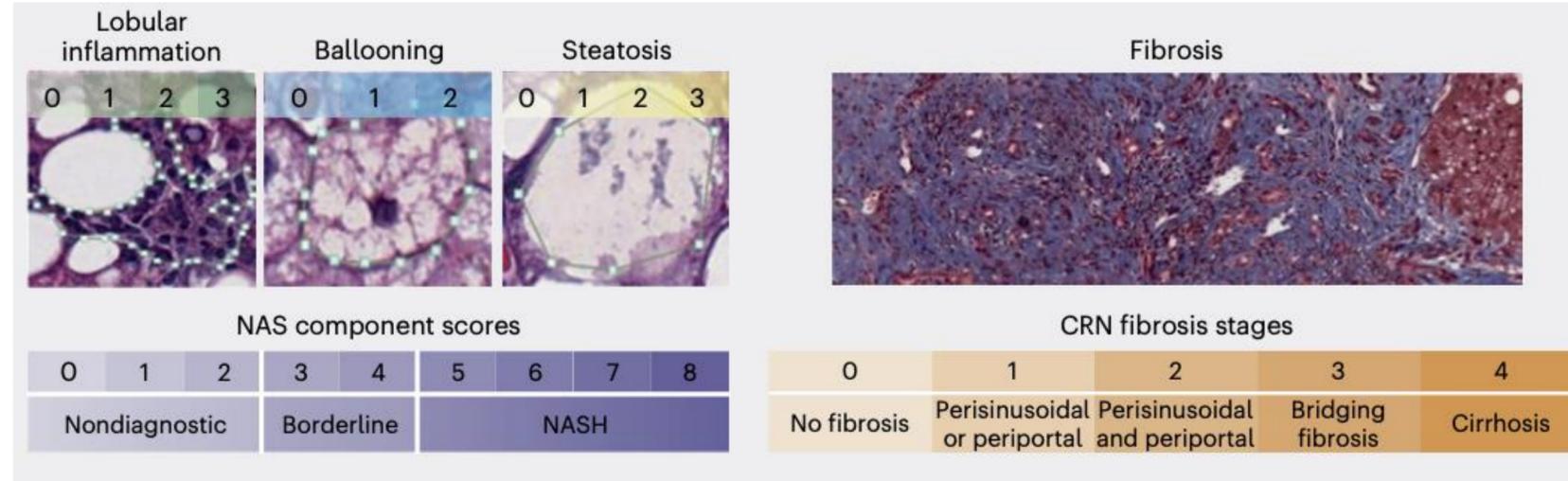
3) AI-enabled continuous MASH scoring

- ▶ AI provides continuous scores that can **detect mild histologic changes happening within the range of single ordinal grade**
- ▶ **Continuous fibrosis score:**
 - Enable patient risk stratification within F3 & F4 stage
 - Enable prediction of progression to cirrhosis & liver related complication

➔ Refined **AI-based continuous grading** might be able to **detect the treatment response earlier** than current histology grading

Next steps:

- ▶ AIM-MASH is being **evaluated by FDA & EMA as a drug discovery tool** for the use in clinical trials



Block 2: Non-Invasive Tests (NITs) for MASLD screening & monitoring

Liver biopsy limitations:

- Not suitable for general population screening & sequential monitoring
- Might discourage patients from participating in CTs

Existing NITs did not yet replace biopsy

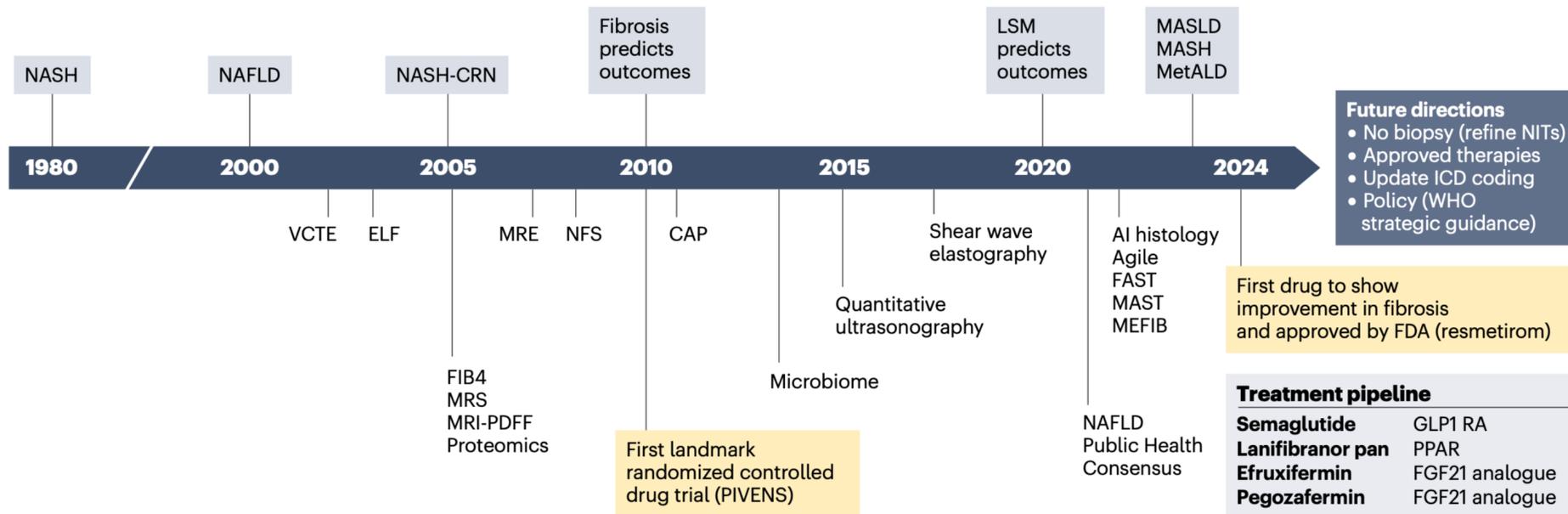
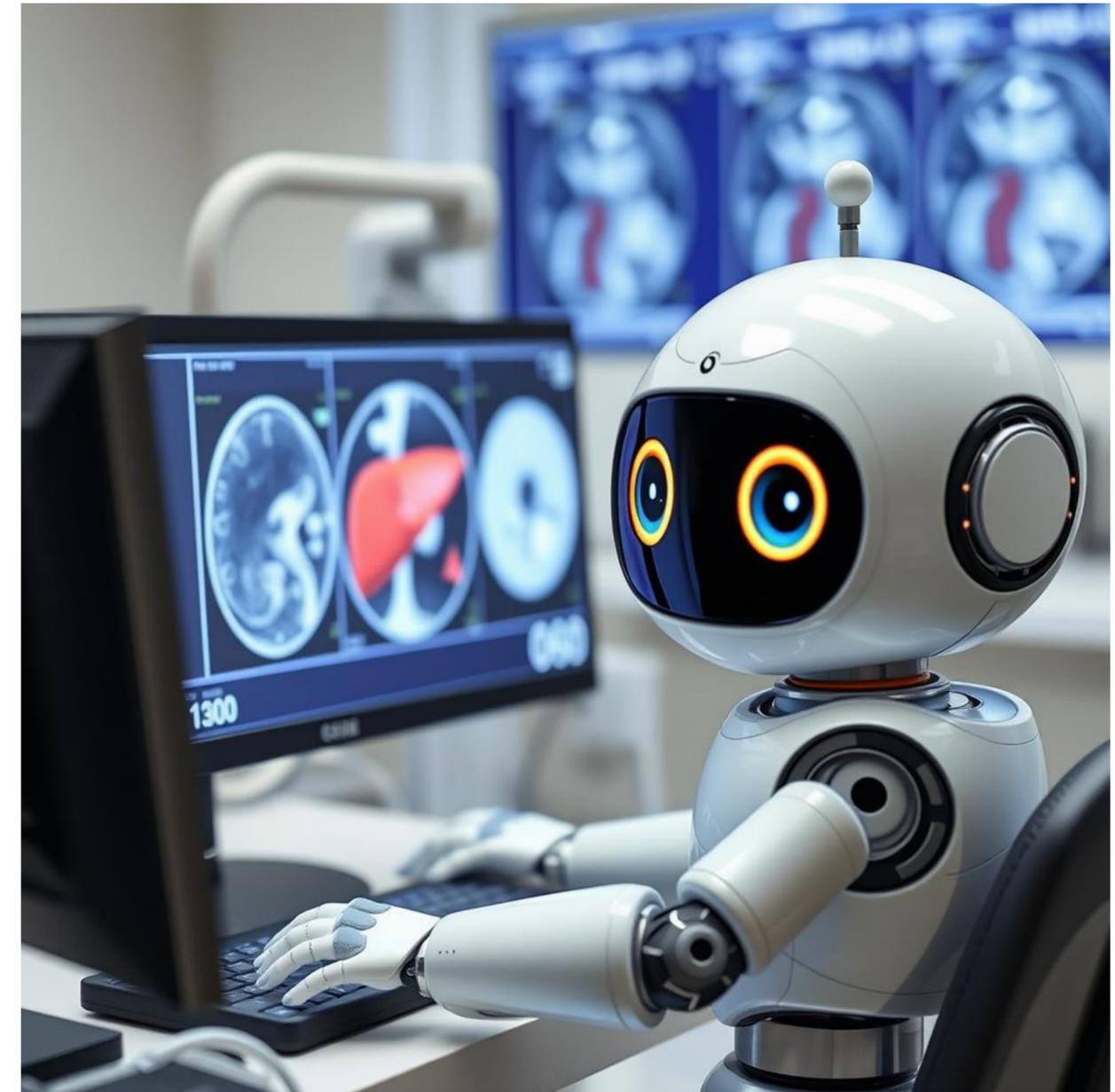


Figure: Timeline for the MASLD field. (Source Allen et al.: (2024))

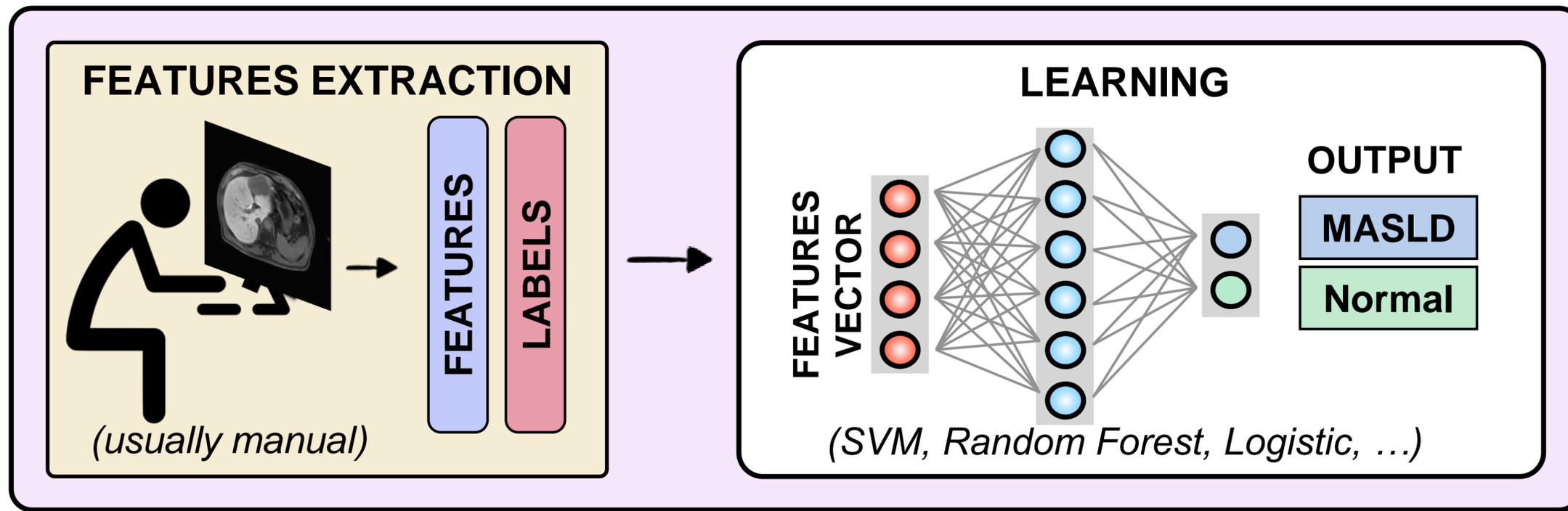
Unmet Needs:

- #1. Enhance accuracy of existing NITs to eliminate need for biopsy
- #2. Identify new non-invasive biomarkers

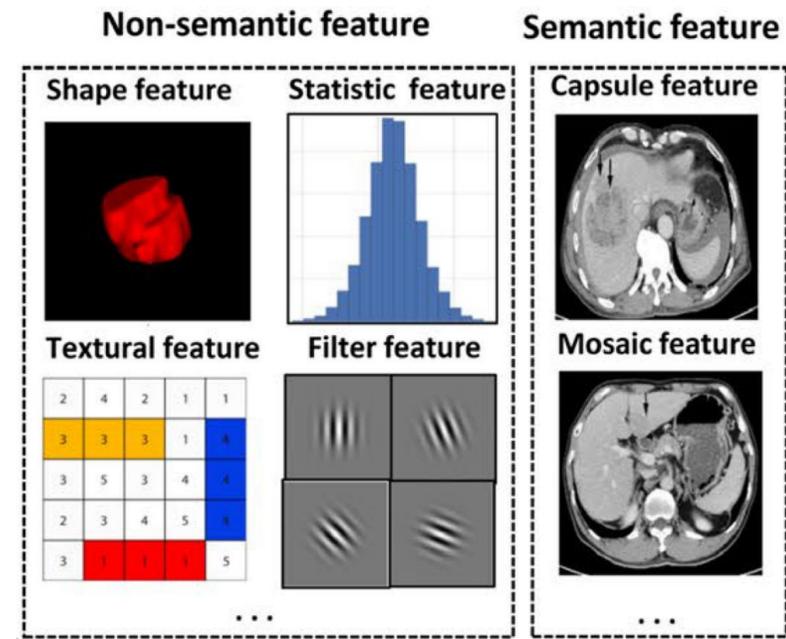


Radiomics in Liver Diseases

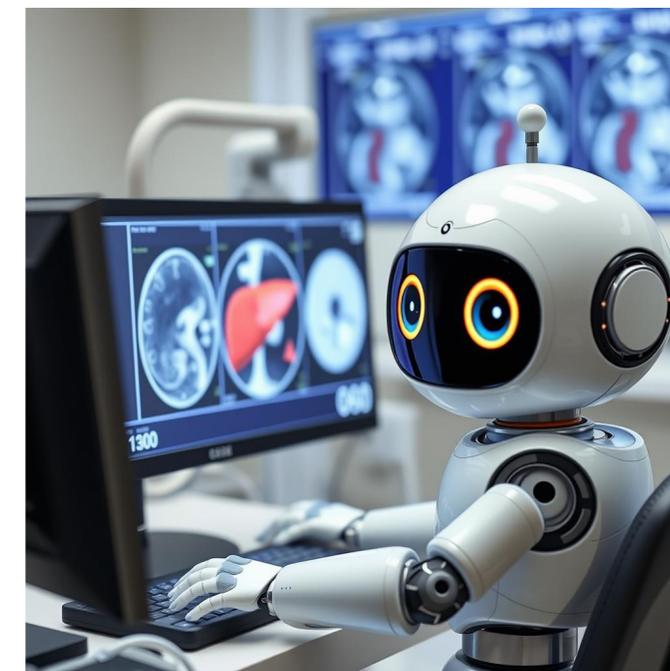
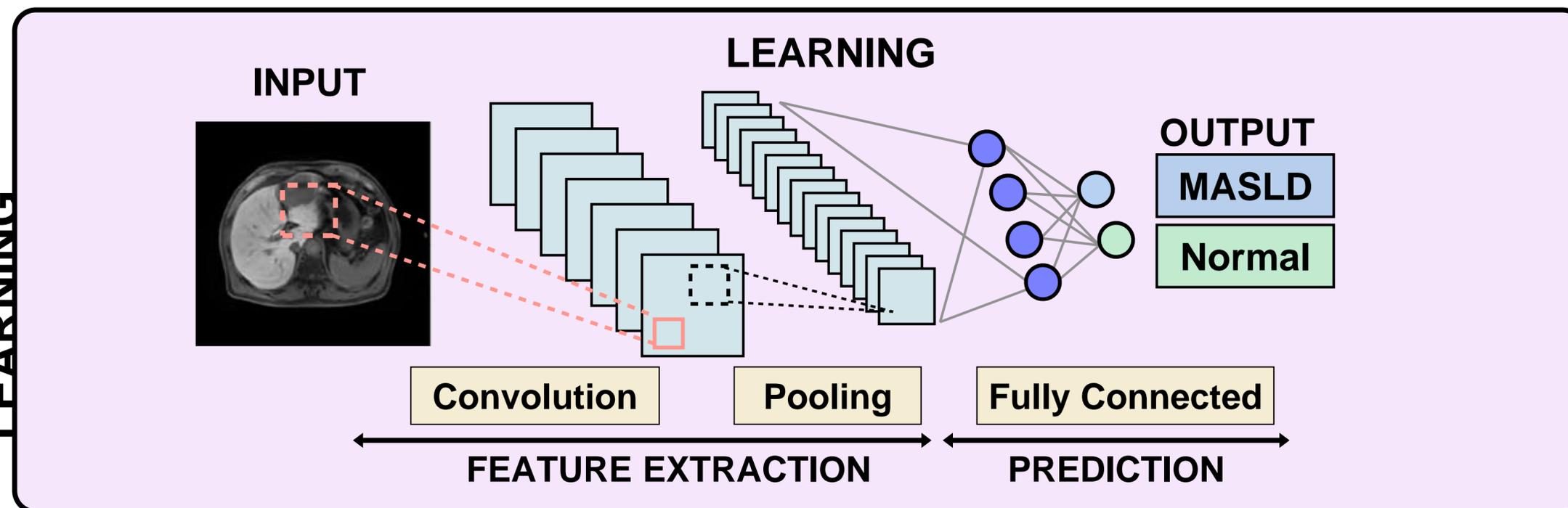
HAND-CRAFTED



Feature extraction



DEEP LEARNING



Application of Radiomics in MASLD

Systematic Review (Zamanian et al.: 2024)

- ▶ Demonstrate feasibility of AI-based image diagnosis
 - MASLD diagnosis: AUC=0.98
 - MASH diagnosis AUC = 0.80
- ▶ Mostly retrospective studies, smaller cohorts

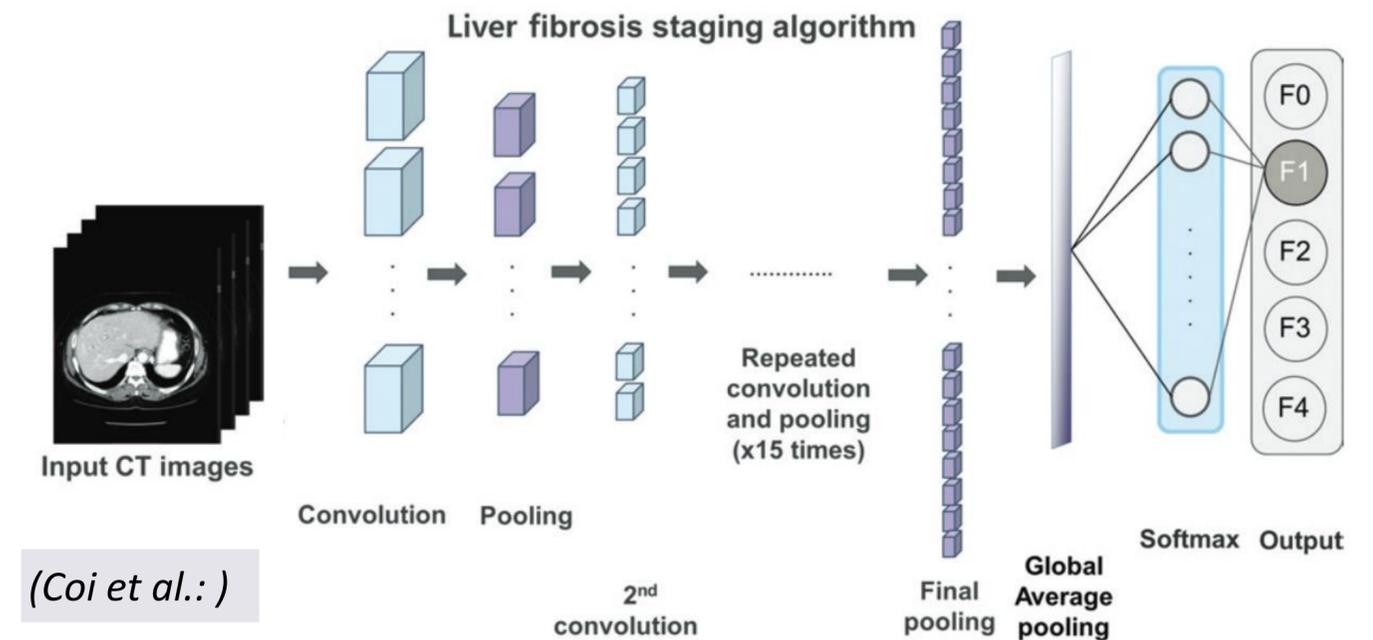
Radiomics for Fibrosis Staging:

Shear wave elastography (SWE) (Wang et al.: 2019)

- ▶ Prospective multi-center study
 - Cirrhosis (F4): AUC = 0.97
 - Advanced fibrosis ($\geq F3$): AUC=0.98
 - Significant fibrosis ($\geq F2$): AUC = 0.85
- ▶ DL model outperforms standard liver stiffness measurement

Next steps for AI-based image diagnosis:

- ▶ Benchmark for objective comparison of AI methods
- ▶ Prospective clinical trial to assess their clinical potential



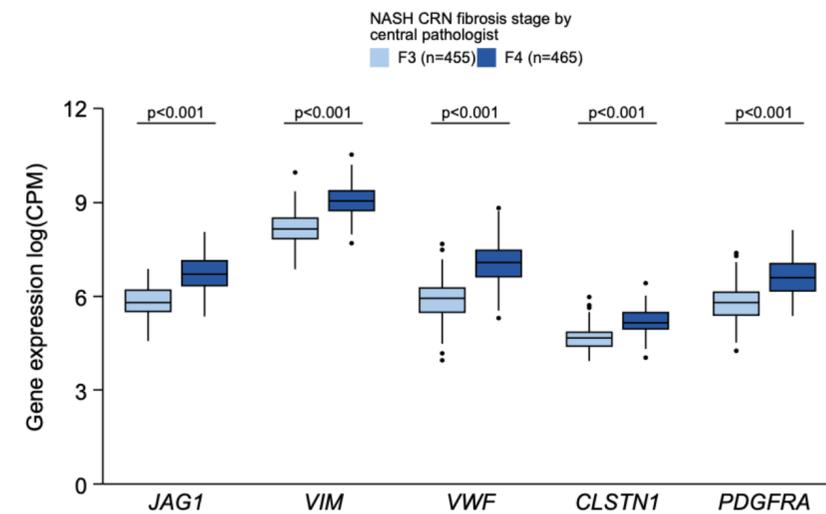
Computer Tomography Images (Choi et al 2018)

- ▶ Retrospective, single center
 - Cirrhosis (F4): AUC = 0.95
 - Advanced fibrosis ($\geq F3$): AUC=0.97
 - Significant fibrosis ($\geq F2$): AUC = 0.96

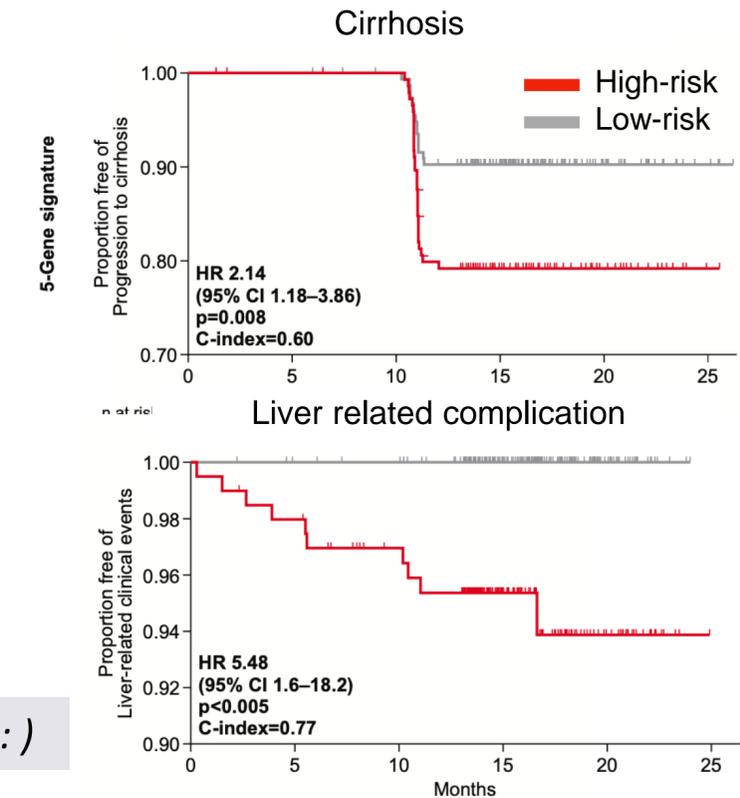
AI-guided Discovery of Non-invasive Biomarkers

RNA-seq Biomarkers (Conway et al., 2023)

- AI-model: correlates MASH histology features & RNA-seq data to **identify 5-gene signatures** associated with severe fibrosis:
 - differentiate between **F3 and F4** patients at the **gene-level**
 - predict progression to cirrhosis (in F3) and clinical events (in F4)



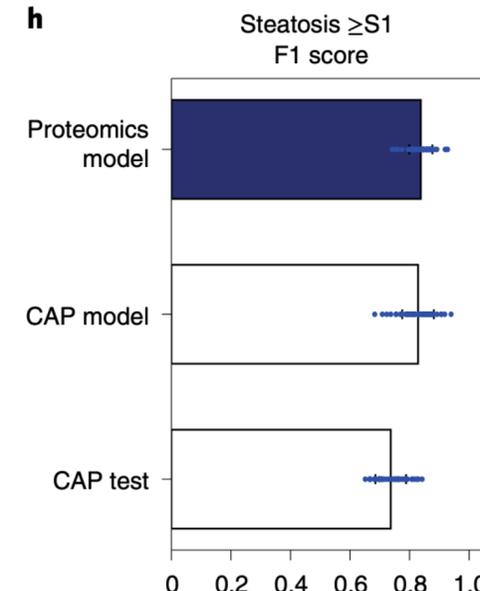
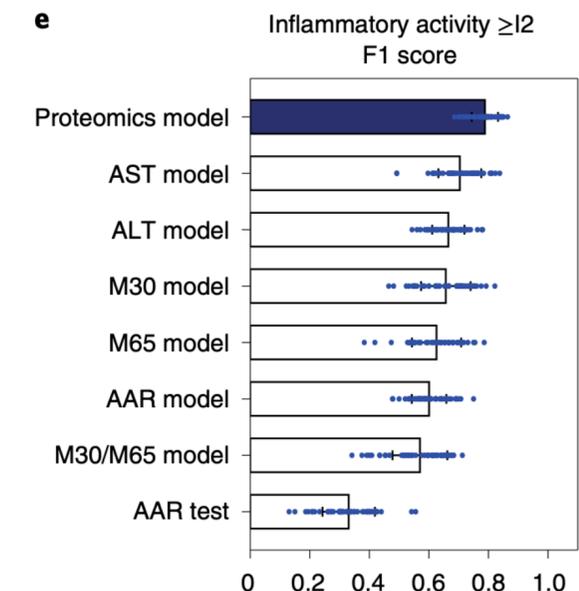
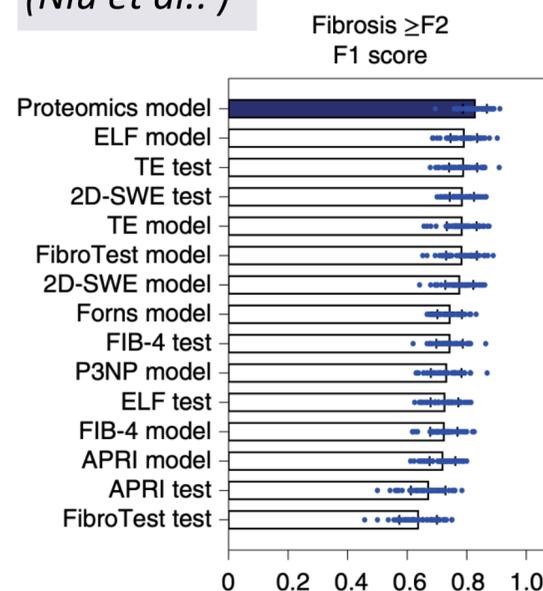
(Conway et al.)



Blood plasma biomarkers: (Niu et al., 2022)

- AI model: correlate histology and mass spectrometry proteomics to identify proteomics biomarker panels
- Detect fibrosis ($\geq F2$) and mild inflammation ($\geq I2$) more accurately than existing clinical assays (eg. TE or ELF)
- Predict of liver relevant events (c=0.90) and mortality (c=0.79)

(Niu et al.)



Future MASLD Applications:

- Explore other modalities: proteomics, lipidomics, metabolomics, transcriptomics, genomics, epigenetic, pharmacokinetics

Block 3: Personalized Medicine

1) Personalized, Risk-based Screening Schedule

Efficient screening schedule:

- Balance between **early detection vs over-screening**
- **Risk factors vary across patients**
- Risk fluctuates **over time even** for the same patient

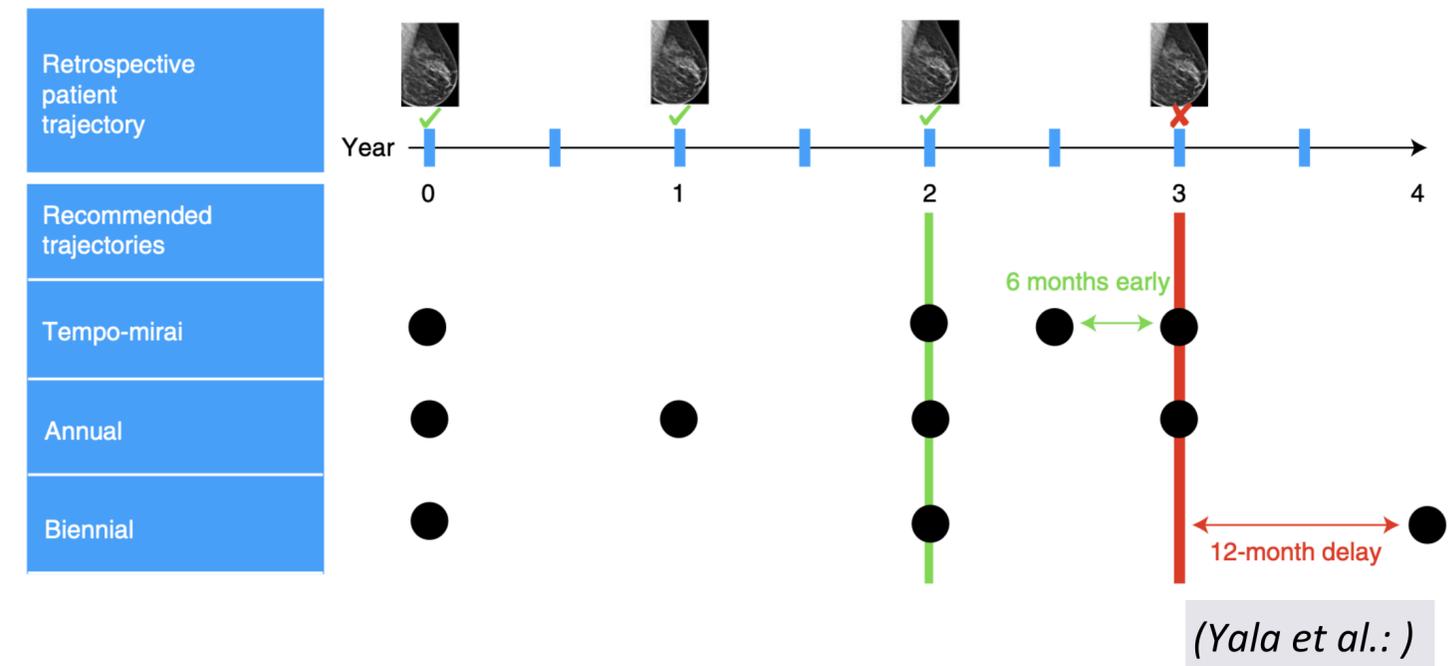


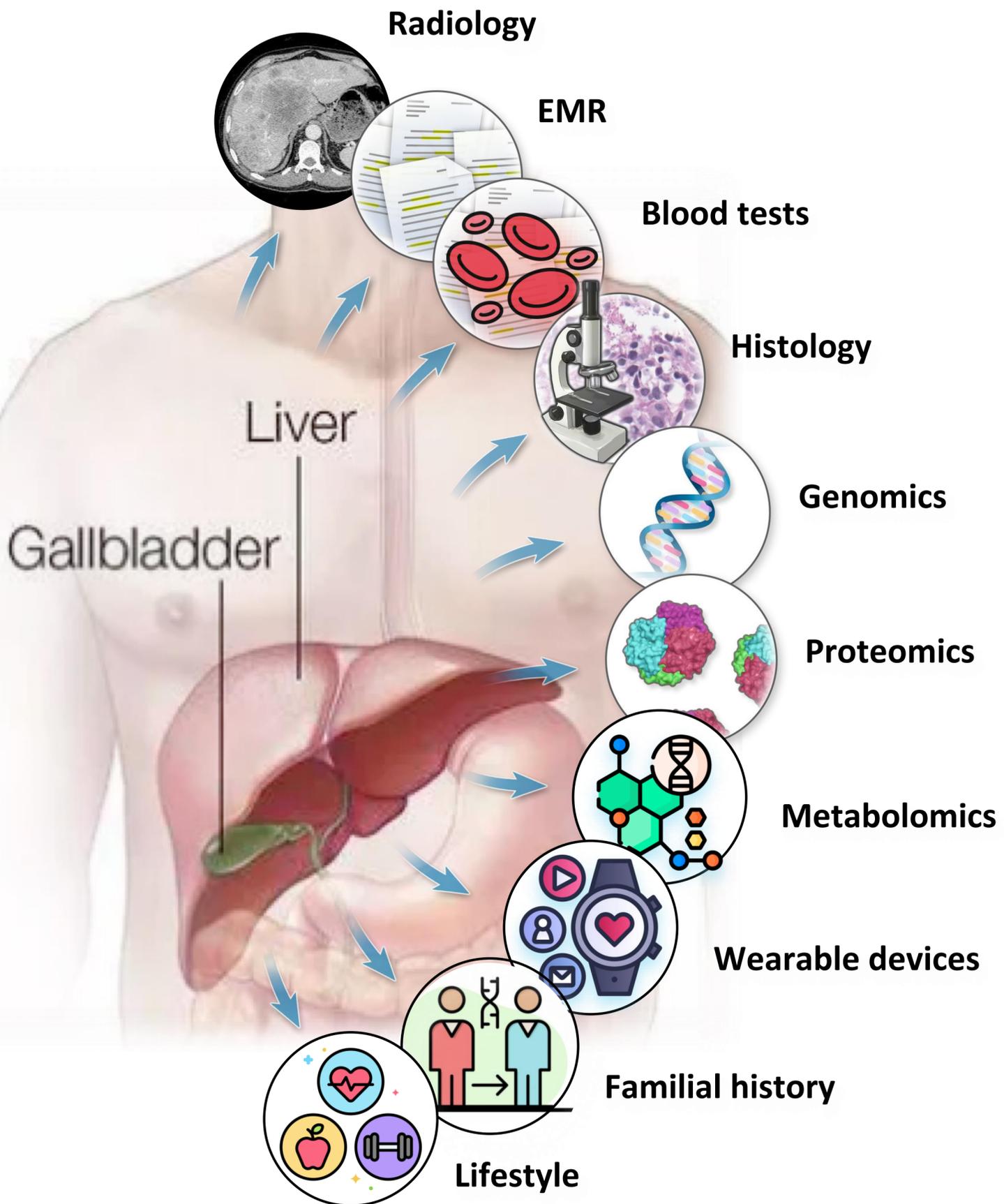
AI-optimized personalized screening schedule (Yala et al. 2022)

- AI: predict **risk of future events** from available past patient data
- Recommend **personalized follow-up schedule** based on the predicted patient's risks
- AI-schedule was significantly more efficient than annual screening

Future MASLD Applications:

- Optimize screening schedule w.r.t specific NIT
- Determine optimal combination or sequence of NIT
- Include other risk factors (e.g. type 2 diabetes)





AI-based Multimodal Data Fusion

Clinical Context Matter

- ▶ Disparities in patient outcomes even with similar diagnosis
 - e.g baseline comorbidities such as type 2 diabetes
- ▶ Patient unique state described by range of modalities
- ▶ Volume and complexity of data: challenging to manually assess patient state under larger clinical context

Potential of AI-based data fusion

- ▶ Identify relevant patterns in complex medical data and associations across modalities
- ▶ AI can leverage complementary and shared information in diverse modalities to provide more accurate and robust predictions

Implication for MASLD

- ▶ Multimodal data for improved patient risk stratification
- ▶ Integrate multiple NITs to improve **screening** accuracy
- ▶ **Novel** multimodal biomarkers for diagnosis & prognosis
- ▶ Personalized treatment plans

Conclusion

- ▶ AI give us "powers" to aid medical interventions at speed, scale & cost that was impossible before
- ▶ Lot of challenges, lot of potential to serve patients & clinicians

Path Forward:

Data sharing accelerates innovation:

- The better and larger datasets → the better AI models
- E.g. PathAI used data from 6 CTs

Active participation of all stakeholders:

- Bring together computer scientist and medical professionals
- Identify medical needs and meaningful AI solutions
- Paris MASH → fostering the interdisciplinary collaboration



→ Thanks to organizers and all of you