

# NASH IN FOCUS

Current Developments in the Management of Nonalcoholic Steatohepatitis

Section Editor: Stephen A. Harrison, MD

## Potential Use of Artificial Intelligence in the Histologic Assessment of Nonalcoholic Steatohepatitis



Quentin M. Anstee, MB BS, PhD, FRCP  
Deputy-Dean of Research & Innovation – Faculty of Medical Sciences  
Professor of Experimental Hepatology & Consultant Hepatologist  
Translational & Clinical Research Institute, Faculty of Medical Sciences  
Newcastle University  
Newcastle upon Tyne, United Kingdom

### **G&H** Where may there be a potential role for the use of artificial intelligence in the histologic assessment of nonalcoholic steatohepatitis?

**QA** When using histology to establish whether a biopsy shows nonalcoholic steatohepatitis (NASH) as opposed to a different liver disease (eg, autoimmune hepatitis), human assessment is key. However, that may not be the case when the diagnosis is established but there is a need to reproducibly quantify specific features within a biopsy. For example, one of the things regulators look for in clinical trials is NASH resolution, which is the transition from the inflammatory form of nonalcoholic fatty liver disease (NAFLD) back to simple steatosis, and so a disease state that is potentially less likely to progress. The single feature that is considered to be a marker of this is the ballooned hepatocyte; its absence or presence defines, to a certain extent, whether a patient has steatosis or NASH. However, it is extremely difficult to accurately quantify these features. That is where there is potential room for the use of artificial intelligence (AI), which could provide more reproducible histologic assessment that is less subject to intra- and interobserver variation.

### **G&H** What types of AI and machine learning techniques may aid in the evaluation of liver histology in NASH?

**QA** A number of different platforms are being proposed that can leverage AI and machine learning approaches to develop algorithms for feature recognition on histology.

At one end of the spectrum are techniques that essentially take images of the standard biopsies that a human pathologist would look at. These techniques then analyze and break down those images into individual features to produce either a continuous scale, for example, of the amount of scarring or fibrosis in the liver, or a categorical scale, similar to the 0-to-4 scale that a human pathologist would use but is based upon the algorithm's feature recognition. The other end of the spectrum consists of techniques that do not require stained biopsies and instead use, for example, second-harmonic generation and two-photon excitation (SHG/TPE) laser microscopy to deeply characterize features in the liver that may not be visible with standard light microscopy. There are increasing data to support the use of all these approaches, and indeed the algorithms are relatively similar in that they are all using AI to look for specific histologic details and trying to break them down. There is still much work to be done in this field, but it is rapidly evolving and these are very exciting technologies.

### **G&H** How well does AI assessment of liver histology correlate with that of expert pathologists?

**QA** It is important to first remember that we are discussing the specific situation of quantifying defined histologic features, not differentiating 2 liver diseases to establish a new diagnosis. In terms of making a diagnosis, human pathology remains essential. Nevertheless, data suggest that interobserver agreement between 2 pathologists can be very variable when it comes to quantifying specific

histologic features (ie, the amount of fat, inflammation, ballooning, and fibrosis in the liver). Even among the best histopathologists, some features such as inflammation and ballooning are very challenging and exhibit kappa values (a measure of interobserver agreement) that are often in the order of 0.5, not dissimilar to tossing a coin.

... AI assessment has the potential to produce a more consistent quantification of specific histologic features and disease severity.

Fortunately, for other features such as fibrosis, the kappa value can be much greater. When a new test like AI is being developed, the existing test (here, human pathology) is used as the reference standard to assess performance. In the current situation, that standard is something of a moving target and is itself just an approximation to the ground truth.

A number of studies have been published in this area. A recent study looked at interobserver variation in the identification of ballooning at the individual cell level. In that study, a trained AI approach appeared to fit within the range of accuracy of expert human pathologists. Thus, this approach was certainly comparable with a human pathologist. Because no test is perfect, and all are an approximation to the ground truth, one way to think of this is that AI approaches might be considered analogous to a pathologist who is consistently imperfect, as opposed to one who is inconsistently imperfect. Thus, AI assessment has the potential to produce a more consistent quantification of specific histologic features and disease severity.

**G&H** How reproducible and standardized has AI assessment of liver histology been shown to be?

**QA** AI assessment has theoretical advantages in terms of reproducibility and standardization, which are important. In the clinical trial space, AI's potential to robustly compare the trial-entry biopsy with the end-of-study biopsy and quantify differences is key.

Let me give an example of the scale of variability we are dealing with. To help train a new AI algorithm, my colleagues and I recently conducted a study in which we asked 9 of the world's leading pathologists to look at 10 digital liver biopsy images and identify individual bal-

looned hepatocytes. Working independently, the pathologists circled every cell in the biopsies that they thought was ballooned. This was an enormous piece of work; the pathologists looked at more than 88,000 individual cells in total. More than 1100 cells were considered to be ballooned by at least one pathologist. However, when looking at the level of agreement on the individual cell level, there was only a single cell that all 9 pathologists considered to be ballooned. This highlights the challenge of judging treatment response in clinical trials; for any of those biopsies, one of the world's leading pathologists might have said that the patient had active NASH, but another leading pathologist might have said that there was no NASH and the disease had resolved. This level of variation means that we may be missing efficacious drugs and makes me think that better ways of assessing drug efficacy are needed in clinical trials. AI may help with this challenge.

AI assessment is also more scalable. There are a limited number of expert liver pathologists in the world. Considering that NAFLD affects 25% to 30% of the adult world population, the demand on human pathologists' time well outstrips the number of human pathologists available. Therefore, it is necessary to find ways of democratizing and standardizing the analysis of liver histology, which is where AI can potentially have a role.

**G&H** Has AI evaluation demonstrated prognostic value in NASH?

**QA** Good data have shown that human pathology assessment can predict disease outcomes. For example, a number of studies have demonstrated that fibrosis stage correlates well with long-term survival and the likelihood of liver-related events. Similarly, one study has demonstrated that long-term prognosis improves as fibrosis regresses, as measured with either histology or noninvasive biomarkers. At present, data are not available to show if AI techniques have similar prognostic value. Most work has focused on their diagnostic value and their utility of assessing fibrosis at a single point in time. Because human-measured fibrosis is known to correlate with long-term outcomes, and AI approaches correlate well with human pathologists, I think it is reasonable to assume that AI approaches could have prognostic value. Nevertheless, that is not proven, so more work is needed for definitive confirmation.

**G&H** How does histologic detail derived from AI compare with that from human pathologists?

**QA** The answer depends on which technique is used. As mentioned, several AI techniques use standard histologic stains. In such cases, computer-driven image analysis has the potential to derive more detail within optical

wavelengths. The other approach uses SHG/TPE microscopy, which uses lasers and so can conceivably capture additional details that may not be as easily detected with light microscopy (eg, how cords of fibrosis connect). It is quite possible that these techniques will provide greater insights into histologic features.

Beyond identifying features, these techniques may also look in greater detail at zonality. A computer-based algorithm can set a specific distance from a consistent feature, such as the central venule of each hepatic lobule, and that information can be used to tease out greater information about what is going on in different zones within different parts of the liver ultrastructure. This type of technique is still evolving, but I think it has a lot of potential. This is already being used to detect subtle features of

Only humans have the plasticity of thought to spot the unexpected and to accurately diagnose it.

fibrosis regression in clinical trial datasets and how features colocalize, as seen in, for example, the results from the TANDEM study. However, I suspect these techniques could provide even greater insights into pathophysiology if combined with novel scientific techniques such as single cell or spatial transcriptomics. In this example, we might begin to further define gene expression changes in different zones of the liver, not through AI pathology itself but by leveraging AI techniques to help support other forms of discovery science.

#### **G&H** What are the main limitations or drawbacks to using AI assessment of liver histology?

**QA** Currently, AI techniques are very good at counting. For example, once they are trained to detect a specific feature, they are very good at quantifying it. However, at present, AI techniques are unable to assimilate enough information to form an accurate diagnosis, for example, differentiating NASH from autoimmune hepatitis. Only humans have the plasticity of thought to spot the unexpected and to accurately diagnose it. It is important not to think about this as AI vs human pathology; it is about using the right technique and the right type of interpretation in the right situation. Human pathology is essential for diagnosing conditions using biopsies when

there is diagnostic uncertainty and for adapting when encountering unexpected findings. AI comes in when the diagnosis is established and there is a need for quantification. It can potentially assist human pathology. This is not about using a computer to do a person's job; it is more about not using a person to do a job that is better done by a computer (simple quantification). One of the biggest concerns about AI, not just in medicine but in any area, is that it will supplant human expertise. That still appears some way off in pathology.

#### **G&H** Do you foresee the use of AI becoming widespread in histologic assessment in the future?

**QA** I think there is definitely a future for AI in the clinical trial space to support human pathology and assist in the robust identification and quantification of features. In clinical practice, I suspect that AI will have some benefit, but there is a question about whether biopsy will still be the standard for assessing patients with NAFLD in 5 years. A strong move toward using noninvasive tests is already being seen. Biopsy will still have a place, but it will be used more for assisting with diagnosis when there is uncertainty and, at least until noninvasive biomarkers are qualified by regulators, in drug development.

#### **G&H** What other applications may AI and machine learning have in NASH?

**QA** There is a lot of interest and excitement about the potential use of AI in many different fields. It remains to be seen whether reality lives up to these expectations. That said, correctly trained and applied, AI or machine learning-based approaches offer great potential to support medical practice in a range of areas beyond image analysis. NAFLD is largely asymptomatic, and the symptoms that patients do have tend to be very nonspecific. The ability to flag at-risk individuals using an algorithmic approach based upon routinely available features in medical records would be very powerful. A great example of this is the recent publication in *Hepatology* from the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) consortium. In that study, we leveraged machine learning approaches against routinely available blood tests and other clinical data to develop a computerized algorithm that can help identify individuals with advanced liver fibrosis or at-risk NASH that was not necessarily clinically apparent. This also speaks to the potential value of being able to run an algorithm through electronic patient records to identify individuals who may not have been picked up by standard techniques. Electronic patient records are used in the hospital setting all the time.

Algorithms need to be built to help flag individuals at risk of liver disease or other conditions to make sure that the clinician seeing these patients does not focus too much on one single organ system or disease area, using technology to support a more holistic approach to patient care. This will be an area of active research for many working in the AI and machine learning space going forward.

### Disclosures

Professor Anstee has received grants or contracts from AstraZeneca, Boehringer Ingelheim, and Intercept. He has received royalties or licenses from Elsevier Ltd. He has received consulting fees on behalf of Newcastle University from Alimientiv, Akero, AstraZeneca, Axcella, 89bio, Boehringer Ingelheim, Bristol Myers Squibb, Galmed, Genfit, Genentech, Gilead, GlaxoSmithKline, Hanmi, HistoIndex, Intercept, Inventiva, Ionis, IQVIA, Janssen, Madrigal, Medpace, Merck, NGM Bio, Novartis, Novo Nordisk, PathAI, Pfizer, PharmaNest, ProSciento, Poxel, Resolution Therapeutics, Roche, Ridgeline Therapeutics, RTI, Shionogi, and Terns. He has received payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing, or educational events from Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal, Medscape, and Springer Healthcare. He has participated on a Data and Safety Monitoring Board on behalf of Newcastle University for Medpace (NorthSea Therapeutics DSMB). He has received support from the LITMUS consortium, which is funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under Grant Agreement 777377.

### Suggested Reading

- Anstee QM, Lucas KJ, Francque S, et al. Tropifexor plus cenicriviroc combination versus monotherapy in non-alcoholic steatohepatitis: results from the phase 2b TANDEM study [published online May 11, 2023]. *Hepatology*. doi:10.1097/HEP.0000000000000439.
- Brunt EM, Clouston AD, Goodman Z, et al. Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. *J Hepatol*. 2022;76(5):1030-1041.
- Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol*. 2020;73(6):1322-1332.
- Dinani AM, Kowdley KV, Noureddin M. Application of artificial intelligence for diagnosis and risk stratification in NAFLD and NASH: the state of the art. *Hepatology*. 2021;74(4):2233-2240.
- Kleiner DE, Brunt EM, Van Natta M, et al; Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a histological scoring system for non-alcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313-1321.
- Lee J, Westphal M, Vali Y, et al; LITMUS investigators. Machine learning algorithm improves detection of NASH (NAS-based) and at-risk NASH, a development and validation study [published online March 31, 2023]. *Hepatology*. doi:10.1097/HEP.0000000000000364.
- Liu F, Goh GB, Tiniakos D, et al. qFIBS: an automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis. *Hepatology*. 2020;71(6):1953-1966.
- Naoumov NV, Brees D, Loeffler J, et al. Digital pathology with artificial intelligence analyses provides greater insights into treatment-induced fibrosis regression in NASH. *J Hepatol*. 2022;77(5):1399-1409.
- Taylor-Weiner A, Pokkalla H, Han L, et al. A machine learning approach enables quantitative measurement of liver histology and disease monitoring in NASH. *Hepatology*. 2021;74(1):133-147.
- Wang S, Li K, Pickholz E, et al. An autocrine signaling circuit in hepatic stellate cells underlies advanced fibrosis in nonalcoholic steatohepatitis. *Sci Transl Med*. 2023;15(677):eadd3949.