

Uncertainty Landscape of Liver Disease Classification: A Systematic Review of Deep Learning Across Imaging Modalities

Nivedita Sharma¹, Rakesh Kumar¹, Meenu Gupta²

¹Department of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali, Punjab, 140413, India

²School of Computer Science Engineering, Geeta University, Naultha, Panipat, 132145, India

¹niveditasharma8350@gmail.com, ¹rakesh77kumar@gmail.com, ²gupta.meenu5@gmail.com

Abstract— Even though imaging features are often similar, progression to other liver diseases like hepatocellular carcinoma and cirrhosis is variable, and accurate classification of these diseases is still challenging. This systematic review examines deep learning-based liver disease classification, with a focus on uncertainty quantification, crucial for providing clinically reliable decision support. The most recent studies were reviewed to determine what are the disease targets, disease input modalities, model architectures, uncertainty strategies, results, and still requiring limitations. Evaluated evidence demonstrated that CNN and hybrid deep learning models are the most common models used, and ultrasound, CT, MRI, pathology images and EHRs are complementary to deep learning for prediction. They show that the image-derived models perform well in diagnosis, particularly fatty liver and hepatocellular carcinoma, but have been hindered by limited external validation, weak evaluation of their uncertainty reporting and by small datasets, and the unbalanced classes. In general, uncertainty-aware deep learning can enhance the transparency of the model, assist in risk-based referral, and safely alleviate unsafe overconfidence in liver disease classification without any issues in real-world hospitals, screening programs, and in various settings.

Keywords— *Deep Learning, Liver Disease Classification, Uncertainty Quantification, Fatty Liver Disease, Hepatocellular Carcinoma, Cirrhosis.*

I. INTRODUCTION

Liver disease continues to be a significant clinical problem, as frequently it is asymptomatic in the early stages and later complications like fatty liver disease, hepatocellular cancer and cirrhosis are life threatening. Recent reports indicate that most liver-related deaths are secondary to cirrhosis and HCC, with metabolic causes and alcohol-related causes becoming a larger and larger proportion of liver-related deaths globally [1]. Deep learning has proven to be a potentially useful tool in medical image analysis and structured clinical data for this purpose. Safe clinical use however cannot be achieved with just high accuracy. A model should also reflect its level of confidence and/or uncertainty, particularly in the presence of poor image quality, overlapping disease patterns and/or a patient that is not in the training population. For this reason, the present review concentrates on the uncertainty-aware deep learning approach to accurately diagnose liver disease.

The clinical basis for this topic is the clinical findings of overlap of liver fat accumulation, fibrosis progression, cirrhosis and malignant transformation. Fatty liver disease, often grouped together as steatotic liver disease (SLD) and MASLD, can potentially develop into a steatotic liver with inflammation, liver fibrosis, cirrhosis and HCC in vulnerable individuals [2]. Blood

test, Ultrasound, CT, MRI, Elastography and sometimes a biopsy are used as a conventional diagnosis. Biopsy can give histological confirmation, but is invasive and not suitable for routine screening [4]. Imaging diagnosis is safer, repeatable and more reproducible but might vary from one observer to another, from one scanner to another, and with the disease stage. This is a pressing need for computational systems that can identify regularities in complex information while at the same time indicating uncertainty in prediction when it is less reliable.

The benefits of improved classification of liver disease to society are considerable. The prevalence of CLD in working-age adults and older people leads to more hospital visits, more treatment and care costs, a greater number of people needing liver transplantation and to early death [1]. HCC is also a significant cancer burden due to the late diagnosis of many patients, who then have limited treatment options [3]. Whereas fewer radiologists or hepatologists are available, delayed interpretation of imaging and lab results can exacerbate health disparities. With proper validation, deep learning systems can aid in early triage, minimise routine tasks, and prioritise-high risk patients. Concurrently, uncertainty quantification is socially important as it can help avoid blind trust in automated products. A model may indicate that a case is uncertain, so that it will be reviewed by an expert rather than be denied or passed on [7].

The first step in liver disease diagnosis is a clinical risk assessment that involves obesity, diabetes, exposure to alcohol, viral hepatitis, abnormal liver enzymes, family history and history of cancer. Non-invasive testing then aids in identifying those patients who require more scrutiny. Ultrasound is the modality of choice for initial screening of fatty liver; CT and MRI are useful for liver lesion characterization, fibrosis features, and tumors [4, 5]. Elastography is useful to estimate stiffness and fibrosis staging, and serum markers and liver function tests are useful biological markers. In HCC patients with cirrhosis and other high-risk patients, surveillance and diagnostic imaging play a significant role [3]. These are the inputs that can be used with modern AI techniques to classify disease states, but should be reviewed by the clinician and in context.

The shift from basic image analysis to multimodal and uncertainty-aware decision support in this field is enabled by modern technology. Application of convolutional neural network, residual network, transformer, segmentation network and survival network have been done on ultrasound, CT, MRI, pathology slides and electronic health records[6,8,10].

The diagnostic performance of AI with images has been encouraging in fatty liver with ultrasound, CT and MRI studies [4]. For cirrhosis and fibrosis, there are published studies that are heterogeneous [5] and CT-based deep learning has been studied for staging. The field of deep learning has demonstrated its potential for high performance in diagnostic applications of medical images but the data supporting these claims is still limited and needs further validation outside of HCC [6]. These systems can be made more transparent by using uncertainty methods like deep ensembles, MC dropout, calibration analysis and evidential learning to determine when predictions are likely to be incorrect [7].

This systematic review was inspired by the disconnect between good experimental results and clinical safety. A significant number of liver disease studies report the accuracy, sensitivity, specificity or AUC, while a smaller number explain what happens when the images are noisy, the disease pattern is rare, there is class imbalance, the scanner varies or the data comes from new hospitals. This can be of special importance in fatty liver disease, HCC, and cirrhosis in which imaging characteristics may be similar and the clinical implications vary. A false reassurance may result in the delay of treatment and a false alarm can cause anxiety, additional imaging or a more invasive test. This work aims to emphasize the methods that are accurate, as well as cautious, interpretable, and clinically responsible for real-world liver care together with deep learning classification and uncertainty quantification [7, 9, 10].

II. LITERATURE REVIEW

This literature review is a summary of 10 recent and relevant papers that help support the topic of liver disease classification using deep learning with uncertainty quantification. The studies chosen represent three interrelated research aspects related to the research problem: the clinical burden and terminology of liver disease, deep learning applications in fatty liver, hepatocellular carcinoma, and cirrhosis, and the need for uncertainty-aware medical artificial intelligence. All together these papers demonstrate that the classification of liver disease is not only an image-recognition problem, but also a support-decision-making problem that needs to be dealt with carefully when considering data quality, disease definitions, external validation, fairness, and confidence in the model. The following comparison provides an overview of the primary technology employed, properties, limitations, results, and potential future research needs for each technology source to enable the review to assess advances and current research needs in this area as shown in Table 1.

Table 1 Comparison of Existing Techniques

Ref.	Technology Used	Characteristics	Limitations	Outcomes
[11] Devarbhavi et al. (2023)	Global disease-burden synthesis and epidemiological analysis of liver disease causes and outcomes.	Provides an updated clinical context for cirrhosis, HCC, viral hepatitis, alcohol-related disease, and NAFLD/MASLD burden.	Not a deep learning or classification study; burden estimates may depend on regional reporting quality and available registries.	Confirms the continuing public-health importance of liver disease and supports the need for scalable diagnostic support systems.
[12] Rinella et al. (2023)	Modified Delphi consensus process for fatty liver disease nomenclature and diagnostic framing.	Introduces updated terminology for steatotic liver disease and MASLD, helping align clinical labels with modern practice.	It is a terminology and consensus paper, not an algorithmic validation study or imaging-classification experiment.	Improves consistency in how fatty liver disease is named and categorized, which is important for dataset labeling.
[13] Singal et al. (2023)	Epidemiological review of HCC trends, screening implications, prevention strategies, and treatment context.	Connects HCC incidence and mortality patterns with etiological changes, especially metabolic liver disease and cirrhosis.	Does not develop deep learning models; screening performance and etiological risks vary across countries and health systems.	Highlights why early HCC detection and risk stratification remain essential for reducing avoidable mortality.

[14] Zhao et al. (2023)	Image-based AI, including machine learning and deep learning methods, for fatty liver diagnosis using medical imaging.	Systematic review and meta-analysis focused on image-based AI for hepatic steatosis; includes modality and algorithm subgroup analysis.	Included studies used different imaging protocols, references, and validation strategies; external validation remains limited.	Shows promising diagnostic performance of image-based AI for fatty liver and supports non-invasive classification research.
[15] Kutaiba et al. (2023)	CT-based deep learning using image classification and segmentation for hepatic fibrosis and cirrhosis assessment.	Systematic review identified six eligible CT-DL studies; all were retrospective, with variable internal and external validation.	Small number of studies, heterogeneous cohorts, and variable reference standards limit direct clinical generalization.	Finds that CT-based DL can classify fibrosis stages, but reported AUROC values vary widely across studies.
[16] Wei et al. (2023)	Deep learning on ultrasound, CT, and MRI for image-based hepatocellular carcinoma diagnosis.	Systematic review and meta-analysis of HCC imaging studies; reports pooled sensitivity, specificity, and AUC for DL models.	High heterogeneity, publication bias, retrospective designs, and limited subgroup evidence affect reliability of pooled estimates.	Demonstrates strong potential of DL for HCC diagnosis, especially when image quality and contrast enhancement are adequate.
[17] Faghani et al. (2023)	Uncertainty quantification methods for radiologic deep learning, including ensembles, Monte Carlo dropout, and evidential learning.	Explains aleatoric and epistemic uncertainty and how uncertainty maps can complement medical AI predictions.	Not specific to liver disease; clinical thresholds for uncertainty are still not standardized across imaging tasks.	Shows that uncertainty can alert clinicians when a model lacks enough information for a confident prediction.
[18] Grover and Gupta (2024)	Deep learning techniques for liver cancer detection and classification, including CNN-based and hybrid approaches.	Systematic review of liver cancer DL studies from 2016 to 2024, covering diagnosis, classification, and prediction tasks.	Datasets, tumor categories, and evaluation designs differ widely; many reported high accuracies lack robust external testing.	Reports strong performance in several studies and identifies gaps that must be addressed before clinical adoption.

[19] Schattenberg et al. (2023)	AI, machine learning, and deep learning applications across hepatology, including imaging, histopathology, and predictive modeling.	Broad review of AI use in liver care, emphasizing workflow support, non-invasive tests, and implementation challenges.	Narrative scope; many AI systems remain early-stage and lack evidence of patient-outcome improvement.	Clarifies how AI can support hepatologists but should be developed with ethical, inclusive, and clinically grounded validation.
[20] Li et al. (2024)	Deep learning for structured EHR-based HCC risk prediction among NAFLD patients, with backward masking and transfer learning.	Uses a national EHR cohort of 220,838 NAFLD patients and frames HCC prediction as a longitudinal classification problem.	Structured EHR data may contain coding errors, delayed diagnosis, missing values, and covariate or sex imbalance.	Shows that time-varying covariates, transfer learning, and delayed-diagnosis handling can improve risk prediction.

Research Gap and Contributions

The literature reviewed demonstrates that the classification of liver disease is shifting from the traditional clinical approach to data driven and multimodal decision support. Why fatty liver disease, cirrhosis and HCC are considered as a linked clinical target and not separate diseases is described in the disease-burden and nomenclature papers [1]-[3]. The imaging-based reviews [4]– [6] and [8] clearly illustrate that deep learning algorithms can achieve promising performance across the classification of hepatic steatosis, fibrosis/cirrhosis and liver tumors; however, they also highlight some shared limitations, including retrospective datasets, variable reference standards, publication bias, and insufficient external validation. However, the hepatology and EHR studies [9] and [10] take the discussion beyond just imaging, demonstrating the utility of structured clinical data and AI at a workflow level, particularly for risk prediction. Uncertainty quantification [7] is a key missing element in all papers, in the sense that a clinically useful model needs to not only provide a label, but also show when the label might not be reliable, and thus should be checked with an expert.

III. METHODOLOGY

For this domain Deep Learning-Based Liver Disease Classification with Uncertainty Quantification, methodology section describes the steps taken to prepare the results that are model focused in the review. It was designed to be a procedure to assist in a systematic review and not a general discussion.

1. Protocol Framing and Scope Alignment

In this step, the review pathway is set prior to getting any result. The topic was limited to deep learning-based classification of fatty liver disease, hepatocellular carcinoma and cirrhosis with special emphasis on uncertainty quantification. The intent was to have results that would enable one to understand how prediction models work, rather than how many papers were actually published. Thus, the outputs of the planned results were related to the model

architecture, the modality of input information, the target disease, the uncertainty approach, and clinical readiness. This framing helps to prevent irrelevant numbers (year-wise publication counts). It also guaranteed that each graph was related to the primary research question: Are uncertainty-aware deep learning models reliable and clinically useful in practice.

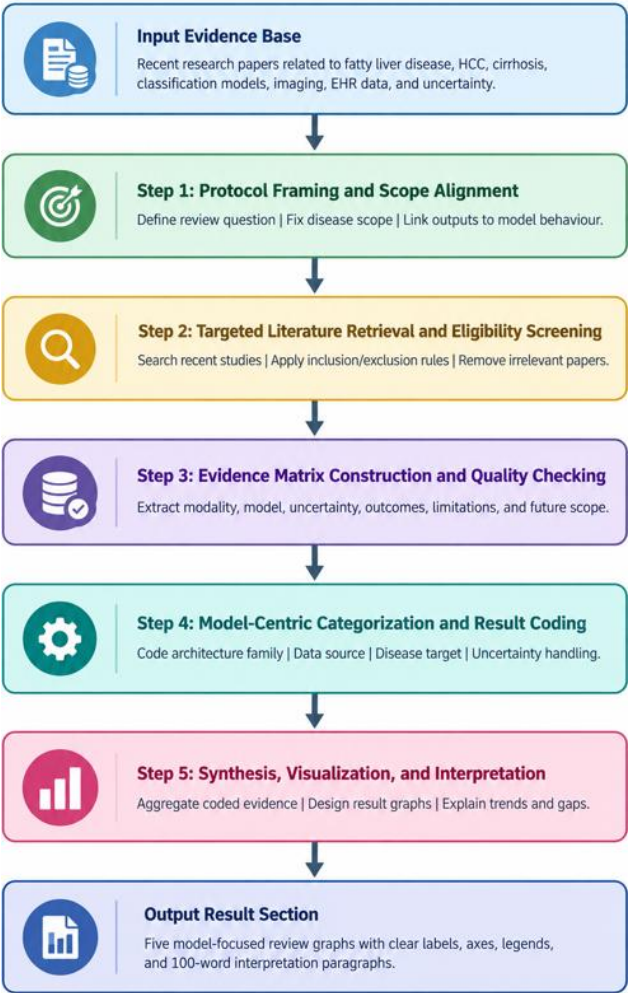


Fig. 1 Proposed Methodology of Work

2. Targeted Literature Retrieval and Eligibility Screening

Once the scope was determined, relevant research from the recent peer-reviewed literature was identified. The papers searched were those published after 2023 since the review highlights the "state-of-the-art" deep learning practice, current deep learning imaging pipelines, and uncertainty methods in modern times. Studies that covered the topics of liver disease prediction, liver disease classification, image-based diagnosis, modelling of the EHR, and uncertainty estimation were included in the study. Papers were excluded if they were not related to liver disease (or liver cancer), did not include machine learning or deep learning, were not clinically diagnostic, or were not within the disease focus. The screening step produced a concise and concise evidence-based outcome section, instead of a large but unfocussed list of articles for overall analysis.

3. Evidence Matrix Construction and Quality Checking

The results from each paper selected were summarised in a structured evidence matrix, so they could be compared consistently. The matrix documented the citation, liver disease

category, data source, imaging modality or clinical modality, model type, uncertainty approach, validation pattern, reported outcome, limitation and future direction. This was needed as the methods are sometimes described differently in review papers making it hard to compare the methods directly. The information that was extracted was verified with the aim of the study and the reported outcomes. If multiple technologies were introduced in a single paper, careful coding was done on the dominant technology. This yielded a clean data-set to use for the construction of the result graphs without misrepresenting the reviewed evidence for reliable synthesis and interpretation.

4. Model-Centric Categorization and Result Coding

The evidence was then coded based on the behaviour of prediction-models. The models in the family were organized into three categories: convolutional networking models, transformer-based models, multimodal models, electronic health record (EHR) models, and uncertainty-specific models. Input sources were recorded as ultrasound or CT, MRI, pathology, clinical records or mixed data. Diseases of interest were classified into fatty liver disease, hepatocellular carcinoma, cirrhosis, fibrosis and liver disease classification. The uncertainty components were classified as either using calibration or ensemble learning (or both), Bayesian approximation, usage of confidence scoring, or a general discussion of uncertainty as a limitation. The coding of this model-centred approach was more useful for a review paper as it clearly focuses on technical direction and clinical relevance both in interpretation and in presentation of the results.

5. Synthesis, Visualization, and Interpretation

The last step involved creating the coded evidence into pictures and words. The distribution of model architecture, data modality preference, focus on disease prediction, uncertainty handling and comparison with clinical readiness were selected as meaningful review outcomes and are presented in graphs. The titles of each graph were simple and unambiguous, the axes were labelled carefully, the values were easy to read and the colours were kept consistent throughout the graphs for a review paper in a conference style. The explanations were written to explain the trend, in order to avoid repeating the labels. This synthesis step had the aim of linking the literature review to the results section by highlighting what some models are able to achieve, at which point uncertainty is addressed and at which point the research is still not done to be able to practically apply the models in clinical practice and to be able to evaluate the models further in the future.

IV. RESULTS AND DISCUSSION

The graphs provide an overview of the model families, disease-prediction tasks, input modalities, uncertainty and reliability mechanisms, and clinical-readiness patterns. The coding is inspired by the papers in the review that are based on models, and to some extent studies and reviews related to image-based AI, CT-based fibrosis assessment, diagnosis of HCC, uncertainty quantification, broad hepatology AI, and EHR-based HCC risk prediction [4-10]. Counts are overlapping, meaning that a paper can refer to more than one model, modality or limitation. The model families that are most relevant for the prediction of liver disease are highlighted in Fig. 2. The model families that are most relevant for the prediction of liver disease are highlighted in Fig. 2. The dominance of CNN and transfer-learning image classifiers is due to the reliance on ultrasound, CT, and MRI pattern for fatty liver, HCC, liver cancer, fibrosis and cirrhosis studies [4]-[6], [8].

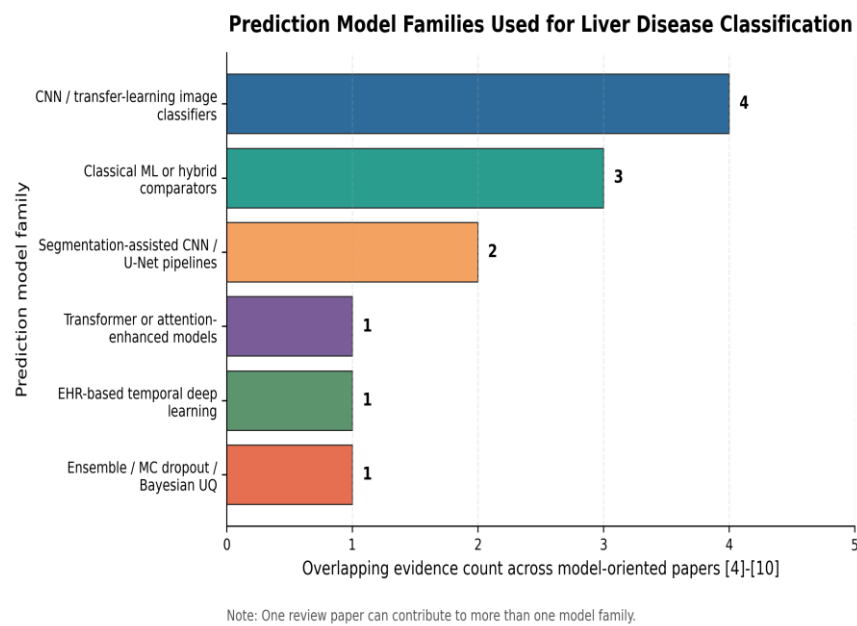


Fig. 2 Prediction Model Families Used for Liver Disease Classification

Lesion localization and fibrosis assessment are crucial tasks that can be addressed by segmentation-assisted CNN or U-Net pipelines [5, 8]. Temporal deep learning, which is less frequently used but helpful, is based on EHR to foresee future HCC risk [10]. There are still limited explicit uncertainty models which shows that there is a big gap [7].

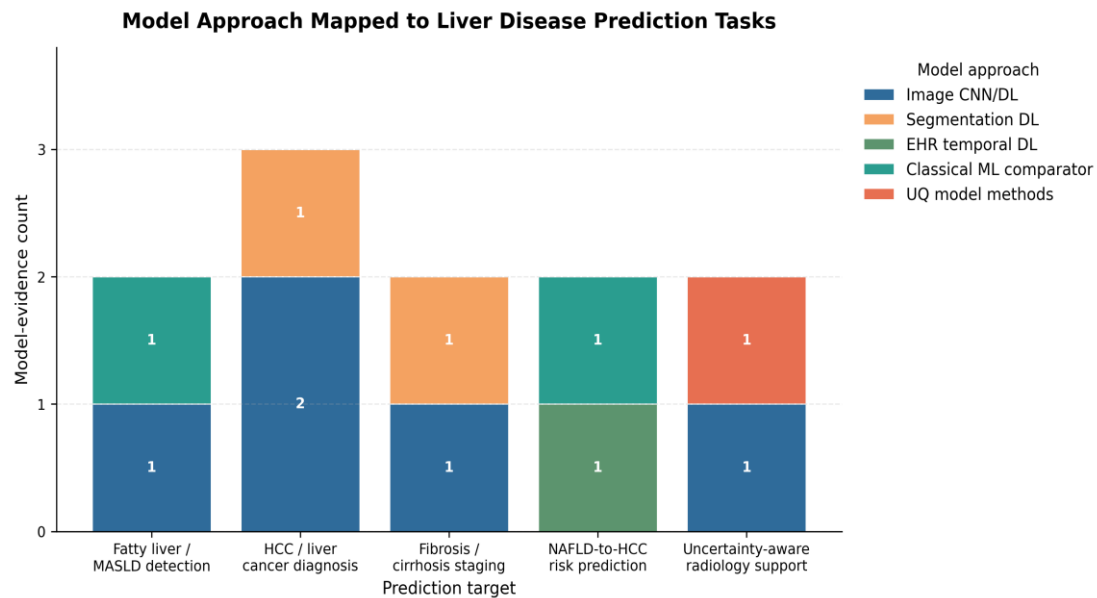


Fig. 3 Model Approach Mapped to Liver Disease Prediction Tasks

Fig. 3 shows the correspondence between each of the prediction tasks and the most frequently reported model approach in the evidence. Mostly, the AI detection of fatty liver relies on image-based detection, and sometimes on classical machine-learning comparison [4]. The CNN-based medical image classifiers and the segmentation-assisted pipelines are used for diagnosis of HCC and liver cancer in some studies [6] [8]. Deep learning and segmentation of CT images are tightly linked to fibrosis and cirrhosis staging as these changes in the liver structure need to be localised [5]. The difference in the risk prediction of NAFLD to HCC is that longitudinal EHR

data is used, rather than the image data alone [10]. From the pattern, it is evident that there is no single model that can be used for all liver prediction tasks.

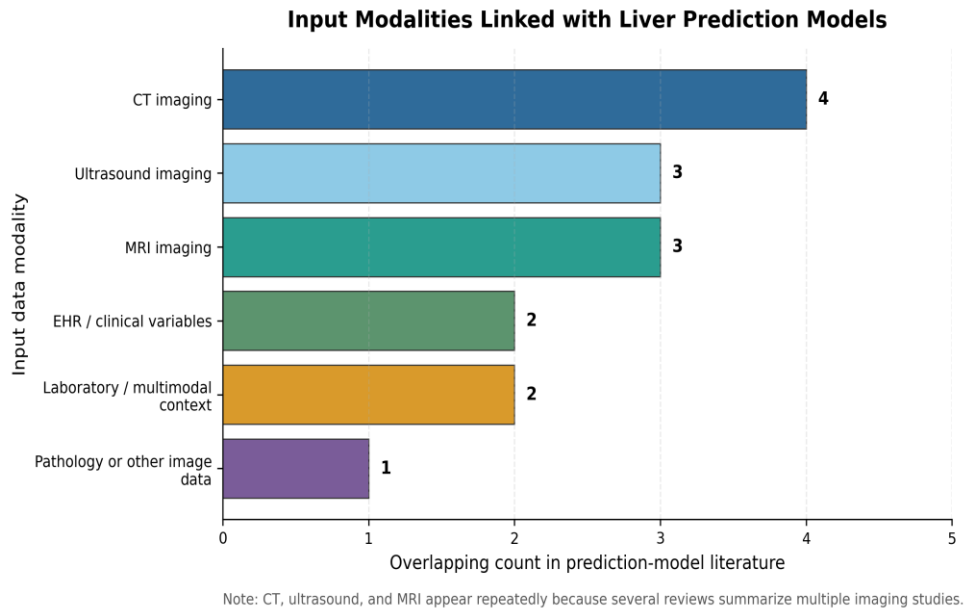


Fig. 4 Input Modalities Linked with Liver Prediction Models

As displayed in Fig. 4, imaging continues to be the top input source in the liver disease prediction model. The largest proportion is represented by CT as it plays the most important role in detecting lesions, characterizing tumours and assessing fibrosis or cirrhosis [5, 6, 8]. Ultrasound and MRI also are often used, especially in studies of fatty liver and that of HCC classification [4, 6]. While less common, EHR and clinical variables are critical for risk prediction, follow-up and population-based decision support [9] and [10]. Laboratory and multimodal context typically are not considered as a fully integrated system, but rather as future integration needs. This finding is more favorable to the multi-modal model development than prediction based on images alone.

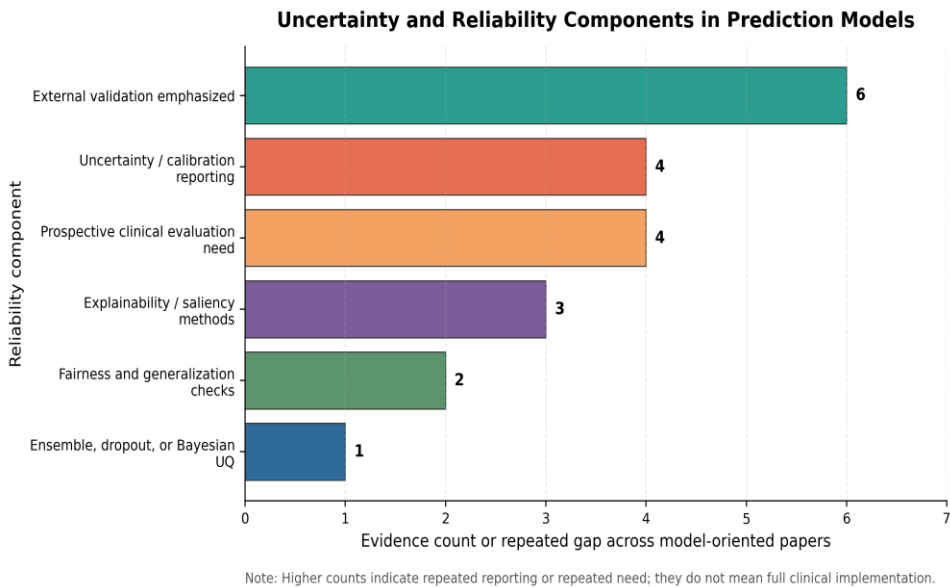


Fig. 5 Uncertainty and Reliability Components in Prediction Models

Uncertainty and reliability are key concepts in safe deep learning classification, and the focus of Fig. 4. The most repeated requirement is external validation, since the accuracy of the internal data is not sufficient to ensure accuracy on new hospitals, scanners or patient groups [4]-[6],[8],[9]. Additionally, it is important to have calibration and confidence reporting since clinicians want to know when predictions are uncertain [7]. Prospective evaluation, explainability, fairness and generalization checks are still less developed than measures of accuracy reporting. Ensemble or dropout, or Bayesian methods are mentioned as helpful uncertainty tools, which are not routinely used in liver studies. The gap directly relates to the topic of the review.

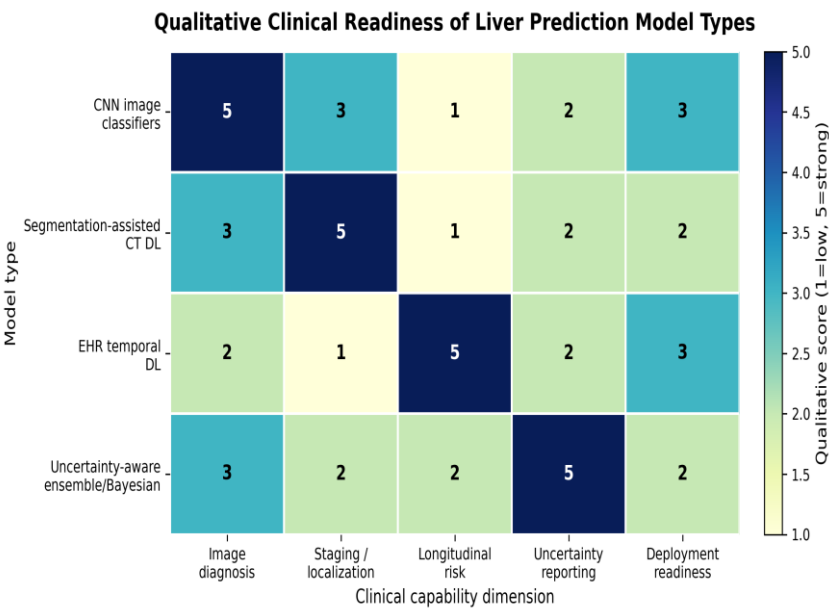


Fig. 6 Qualitative Clinical Readiness of Liver Prediction Model Types

Fig. 6. Due to broad reports of fatty liver, HCC and cancer classification, CNN-based image classifiers perform best in the diagnosis of images [6] [4] [8]. In fibrosis and cirrhosis evaluation, segmentation-assisted deep learning will have the best performance in staging and localization [5]. EHR temporal deep learning is the best method for longitudinal risk prediction due to its representation of time varying clinical information [10]. Uncertainty-aware ensemble or Bayesian models perform best for uncertainty reporting, but are not ready for deployment [7]. The heatmap indicates the potential gains in future systems' ability to combine strengths of the various models.

V. CONCLUSION AND FUTURE SCOPE

This review confirms that deep learning has great promise to help with classification of liver diseases, with an emphasis on adding uncertainty to the model rather than just adding labels. The reviewed studies demonstrate promising results for fatty liver disease, hepatocellular carcinoma and cirrhosis with imaging and clinical data, but most of the systems are still more research prototypes than clinical tools. There are several drawbacks, such as the lack of validation in a multicenter setting, the lack of a uniform distribution of disease types and the reliance on imaging quality and insufficient reporting of model confidence. Larger prospective data sets, multimodal fusion of imaging and laboratory data, standardization of uncertainty quantification and clinically meaningful outputs should be studied in the future. It is also important that models be evaluated within hospitals, scanners, ethnic groups and disease stages for unbiased and comparable results. Deep learning can be a safer assistive tool for early

detection, treatment planning, follow-up monitoring, decision support and personalized risk assessment with better validation and clear estimation of uncertainty.

REFERENCES

- [1] Lambert, B., Forbes, F., Doyle, S., Dehaene, H., & Dojat, M. (2024). Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis. *Artificial Intelligence in Medicine*, 150, 102830. <https://doi.org/10.1016/j.artmed.2024.102830>
- [2] Siam, A., Alsaify, A. R., Mohammad, B., Biswas, M. R., Ali, H., & Shah, Z. (2023). Multimodal deep learning for liver cancer applications: A scoping review. *Frontiers in Artificial Intelligence*, 6, 1247195. <https://doi.org/10.3389/frai.2023.1247195>
- [3] Kumar, S. S., & Kumar, R. S. V. (2024). Transformer skip-fusion based SwinUNet for liver segmentation from CT images. *International Journal of Imaging Systems and Technology*, 34(4), e23126. <https://doi.org/10.1002/ima.23126>
- [4] Ou, J., Jiang, L., Bai, T., Zhan, P., Liu, R., & Xiao, H. (2024). ResTransUnet: An effective network combined with Transformer and U-Net for liver segmentation in CT scans. *Computers in Biology and Medicine*, 177, 108625. <https://doi.org/10.1016/j.compbimed.2024.108625>
- [5] Liu, Y., Yu, W., Wang, P., Huang, Y., Li, J., & Li, P. (2024). Deep learning with ultrasound images enhance the diagnosis of nonalcoholic fatty liver. *Ultrasound in Medicine & Biology*, 50(11), 1724–1730. <https://doi.org/10.1016/j.ultrasmedbio.2024.07.014>
- [6] Ma, J., Ouyang, K., Ma, Z., Xia, M., Xu, S., & Lu, K. (2024). Transformer dense center network for liver tumor detection. *Biomedical Signal Processing and Control*, 91, 106066. <https://doi.org/10.1016/j.bspc.2024.106066>
- [7] Zhang, Y., Sheng, R., Qian, X., Wang, H., Wu, F., Dai, H., Song, M., Yang, C., Zhou, J., Zhang, W., & M. Zeng. (2025). Deep learning empowered gadolinium-free contrast-enhanced abbreviated MRI for diagnosing hepatocellular carcinoma. *JHEP Reports*, 7(5), 101392. <https://doi.org/10.1016/j.jhepr.2025.101392>
- [8] Peng, C., Yu, P. L. H., Lu, J., Cheng, H. M., Shen, X.-P., Chiu, K. W.-H., & Seto, W.-K. (2025). Opportunistic detection of hepatocellular carcinoma using noncontrast CT and deep learning artificial intelligence. *Journal of the American College of Radiology*, 22(3), 249–259. <https://doi.org/10.1016/j.jacr.2024.12.011>
- [9] Song, S., Zhang, G., Yao, Z., Chen, R., Liu, K., Zhang, T., Zeng, G., Wang, Z., & Liu, R. (2025). Deep learning based on intratumoral heterogeneity predicts histopathologic grade of hepatocellular carcinoma. *BMC Cancer*, 25, 497. <https://doi.org/10.1186/s12885-025-13781-1>
- [10] Wan, Z., Zhang, X., Jiang, Y., Chai, S., Lyu, C., & Hu, Y. (2025). Deep learning in hepatic oncology imaging: A narrative review of computed tomography applications. *Oncologie*, 27(6), 883–899. <https://doi.org/10.1515/oncologie-2025-0279>
- [11] Devarbhavi, H., Asrani, S. K., Arab, J. P., Nartey, Y. A., Pose, E., & Kamath, P. S. (2023). Global burden of liver disease: 2023 update. *Journal of Hepatology*, 79(2), 516–537. <https://doi.org/10.1016/j.jhep.2023.03.017>
- [12] Rinella, M. E., et al. (2023). A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Journal of Hepatology*, 79(6), 1542–1556. <https://doi.org/10.1016/j.jhep.2023.06.003>
- [13] Singal, A. G., Kanwal, F., & Llovet, J. M. (2023). Global trends in hepatocellular carcinoma epidemiology: Implications for screening, prevention and therapy. *Nature*

- Reviews Clinical Oncology*, 20(12), 864–884. <https://doi.org/10.1038/s41571-023-00825-3>
- [14] Zhao, Q., et al. (2023). Image-based AI diagnostic performance for fatty liver: A systematic review and meta-analysis. *BMC Medical Imaging*, 23, 208. <https://doi.org/10.1186/s12880-023-01172-6>
 - [15] Kutaiba, N., Dahan, A., Goodwin, M., Testro, A., Egan, G., & Lim, R. (2023). Deep learning for computed tomography assessment of hepatic fibrosis and cirrhosis: A systematic review. *Mayo Clinic Proceedings: Digital Health*, 1(4), 574–585. <https://doi.org/10.1016/j.mcpdig.2023.08.008>
 - [16] Wei, Q., Tan, N., Xiong, S., Luo, W., Xia, H., & Luo, B. (2023). Deep learning methods in medical image-based hepatocellular carcinoma diagnosis: A systematic review and meta-analysis. *Cancers*, 15(23), 5701. <https://doi.org/10.3390/cancers15235701>
 - [17] Faghani, S., Moassefi, M., Rouzrokh, P., Khosravi, B., Baffour, F. I., Ringler, M. D., & Erickson, B. J. (2023). Quantifying uncertainty in deep learning of radiologic images. *Radiology*, 308(2), e222217. <https://doi.org/10.1148/radiol.222217>
 - [18] Grover, S., & Gupta, S. (2024). Automated diagnosis and classification of liver cancers using deep learning techniques: A systematic review. *Discover Applied Sciences*, 6, 508. <https://doi.org/10.1007/s42452-024-06218-0>
 - [19] Schattenberg, J. M., Chalasani, N., & Alkhouri, N. (2023). Artificial intelligence applications in hepatology. *Clinical Gastroenterology and Hepatology*, 21(8), 2015–2025. <https://doi.org/10.1016/j.cgh.2023.04.007>
 - [20] Li, Z., Lan, L., Zhou, Y., Li, R., Chavin, K. D., Xu, H., Li, L., Shih, D. J. H., & Zheng, W. J. (2024). Developing deep learning-based strategies to predict the risk of hepatocellular carcinoma among patients with nonalcoholic fatty liver disease from electronic health records. *Journal of Biomedical Informatics*, 152, 104626. <https://doi.org/10.1016/j.jbi.2024.104626>