

ORIGINAL RESEARCH

Artificial intelligence-based multiclass classification of etiologies in patients with elevated liver enzymes: A prospective cohort study

 Yavuz Ozden,  Omer Yuzugulen,  Nuh Mehmet Buyukberber

Department of Gastroenterology, University of Health Sciences, Kayseri City Hospital, Kayseri, Türkiye

Highlights & Insights

- **Scientific Gap:** Persistent liver enzyme elevation has overlapping etiologies, making diagnostic evaluation sequential and test-intensive.
- **Key Finding:** An explainable multiclass AI model differentiated five major etiologies with an overall accuracy of 0.86 using routine clinical and laboratory data.
- **Clinical Impact:** The model may support structured triage and targeted confirmatory testing, but external validation is required before clinical use.

ABSTRACT

Background and Aim: Elevated liver enzymes are a common reason for hepatology referral, but etiological overlap complicates diagnosis. Most artificial intelligence applications address binary tasks, whereas routine practice requires multiclass discrimination. We developed and internally validated an explainable multiclass model to differentiate five major causes of persistent liver enzyme elevation.

Materials and Methods: This prospective proof-of-concept cohort included 280 consecutive adults with unexplained persistent aminotransferase elevation. Final diagnoses—metabolic dysfunction-associated steatotic liver disease, viral hepatitis, alcohol-related liver disease, drug-induced liver injury, or autoimmune hepatitis—were independently adjudicated by three hepatologists. Data were split into training (70%) and independent test (30%) sets, and four supervised algorithms were evaluated. To assess incorporation bias, a basic model using routine clinical and biochemical variables was compared with an extended model including serological markers. Interpretability was assessed using SHapley Additive exPlanations.

Results: eXtreme gradient boosting achieved the best test-set performance, with an overall accuracy of 0.86 (95% CI 0.78–0.93) and balanced accuracy of 0.87 (95% CI 0.82–0.92). Class-specific AUROC values ranged from 0.92 to 0.95. Serological markers improved discrimination for autoimmune hepatitis (AUROC increase, 0.15; $p < 0.03$) and viral hepatitis (increase, 0.11; $p < 0.02$). Interpretability analysis identified clinically plausible predictors. Misclassifications occurred in 12 cases (14.3%), mainly in overlapping drug-induced liver injury–autoimmune hepatitis presentations.

Conclusion: An explainable multiclass model differentiated five major etiologies of persistent liver enzyme elevation. Serological markers added value for autoimmune and viral etiologies. External validation is required before clinical implementation.

Keywords: Aminotransferases, artificial intelligence, liver diseases.

How to cite this article: Ozden Y, Yuzugulen O, Buyukberber NM. Artificial intelligence-based multiclass classification of etiologies in patients with elevated liver enzymes: A prospective cohort study. Hepatology Forum 2026; 7(3):249–255.

Received: February 16, 2026; **Revised:** February 25, 2026; **Accepted:** March 04, 2026; **Available Online:** July 02, 2026

Corresponding author: Yavuz Ozden; Department of Gastroenterology, University of Health Sciences, Kayseri City Hospital, Kayseri, Türkiye; **E-mail:** yavuzozden@gmail.com



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Introduction

Persistent elevation of liver enzymes is a frequent reason for hepatology referral and often presents a diagnostic dilemma rather than a final diagnosis. The main etiologies encountered in routine practice include metabolic dysfunction-associated steatotic liver disease (MASLD), viral hepatitis, alcohol-related liver disease (ALD), drug-induced liver injury (DILI), and autoimmune hepatitis (AIH). These conditions may present with overlapping clinical and biochemical features, particularly in patients with metabolic risk factors, borderline alcohol intake, recent drug exposure, or low-titer autoantibody positivity.^[1,2] Consequently, diagnostic evaluation often proceeds through sequential laboratory testing, imaging, and occasionally histopathology, with variable diagnostic yield.^[3–6]

Machine learning (ML) methods may support this process by modeling complex interactions among routine clinical and laboratory variables.^[4,7–10] However, most AI applications in hepatology have focused on binary classification or single-disease prediction within selected cohorts.^[6,11,12] This does not fully reflect routine hepatology practice, where several competing etiologies often need to be considered simultaneously in patients with abnormal liver biochemistry. Multiclass etiological discrimination is therefore a clinically relevant but still insufficiently studied application of artificial intelligence (AI).^[3,13]

Before such models can be adopted clinically, several issues must be addressed, including class imbalance, small sample sizes in less frequent diagnoses, generalizability, and incorporation bias when disease-defining tests are used as predictors.^[14] Interpretability is also essential. SHapley Additive exPlanations (SHAP) can help determine whether model predictions are driven by clinically plausible variables rather than opaque statistical patterns.^[15,16]

We hypothesized that a multiclass ML model based on routinely available clinical and laboratory data could distinguish the major causes of persistent liver enzyme elevation. In this prospective proof-of-concept cohort study, we developed and internally validated an explainable AI-based model to differentiate MASLD, viral hepatitis, ALD, DILI, and AIH. To address incorporation bias, we compared a basic model excluding disease-defining serological markers with an extended model that included viral serologies, autoantibodies, and immunoglobulin G. We also examined SHAP-based explanations and misclassification patterns to assess clinical plausibility.

Materials and Methods

Study Design and Participants

We conducted a prospective, single-center observational cohort study at a tertiary academic medical center in Türkiye. Consecutive adults aged ≥ 18 years referred for newly detected, unexplained liver enzyme elevation were screened between October 15, 2025, and February 15, 2026. Persistent elevation was defined as alanine aminotransferase (ALT) and/or aspartate aminotransferase (AST) levels >1.5 times the upper limit of normal on two measurements at least 4 weeks apart.

The study was approved by the Kayseri City Hospital Institutional Ethics Committee (Approval No: 614; October 13, 2025) and conducted in accordance with the Declaration of Helsinki and International Council for Harmonisation Good Clinical Practice guidelines. All participants gave written informed consent. The study was reported in accordance with the TRIPOD statement, incorporating TRIPOD-AI recommendations where applicable.^[17] The completed checklist was submitted as Supplementary File 1.

Eligible patients had persistent ALT and/or AST elevation and completed a diagnostic evaluation allowing definitive etiological classification. Exclusion criteria were known chronic liver disease of established etiology, decompensated liver disease, active extrahepatic malignancy, pregnancy, incomplete diagnostic workup, or loss to follow-up before adjudication. Overall, 280 patients met the criteria and formed the final cohort.

Etiological Classification and Reference Standard

Each patient was assigned to one of five mutually exclusive categories using predefined criteria. MASLD ($n=100$, 35.7%) was defined as hepatic steatosis on imaging plus at least one cardiometabolic risk factor after exclusion of secondary causes.^[1] Viral hepatitis ($n=78$, 27.9%) was defined as chronic hepatitis B or C confirmed by positive hepatitis B surface antigen (HBsAg) with detectable HBV DNA or positive anti-hepatitis C virus antibody (anti-HCV) with detectable HCV RNA. DILI ($n=50$, 17.9%) was diagnosed using a Roussel Uclaf Causality Assessment Method (RUCAM) score ≥ 6 .^[2] ALD ($n=32$, 11.4%) was defined by alcohol intake exceeding established EASL thresholds without competing causes.^[18] AIH ($n=20$, 7.1%) was diagnosed using a simplified International Autoimmune Hepatitis score ≥ 7 or compatible histopathology.^[19] No patient met criteria for more than one category.

The primary outcome was the adjudicated final etiology. Three hepatologists independently reviewed all available clinical, laboratory, imaging, and histopathological data and were blinded to model outputs. Disagreements were resolved by consensus.

Data Collection and Preprocessing

Baseline data were collected prospectively using a structured electronic case report form. Candidate predictors were selected from previous literature and routine hepatology practice.^[1,3,5] Variables included age, sex, body mass index, alcohol intake, medication exposure, comorbidities, complete blood count, ALT, AST, alkaline phosphatase, γ -glutamyl transferase, total bilirubin, albumin, international normalized ratio, fasting glucose, lipid profile, HBsAg, anti-HCV, antinuclear antibody, anti-smooth muscle antibody, and immunoglobulin G. All patients underwent abdominal ultrasonography; transient elastography was recorded when available.

Missing data were $<10\%$ for all predictors and were imputed using multiple imputation by chained equations. Categorical variables were one-hot encoded. To avoid information leakage, synthetic minority over-sampling was applied only within cross-validation folds of the training set. Continuous variables were standardized using training-set parameters and applied unchanged to the test set.

Table 1. Baseline characteristics according to etiological category (n=280)

Characteristic	MASLD (n=100)	Viral hepatitis (n=78)	DILI (n=50)	ALD (n=32)	AIH (n=20)	p
Age (years)	55.8±12.1	49.6±14.8	52.4±13.9	48.3±15.5	41.9±16.8	<0.001
Male sex, n (%)	62 (62.0)	49 (62.8)	20 (40.0)	19 (59.4)	2 (10.0)	<0.001
BMI (kg/m²)	32.8±4.9	27.4±4.3	28.6±5.1	29.1±4.8	25.9±3.7	<0.001
ALT (U/L)*	82 (55–120)	115 (78–168)	148 (92–215)	96 (68–140)	130 (85–190)	<0.001
AST (U/L)*	60 (42–95)	78 (52–115)	100 (65–155)	112 (82–170)	95 (60–140)	<0.001
AST/ALT ratio*	0.72 (0.60–0.92)	0.75 (0.58–0.95)	0.68 (0.55–0.82)	1.50 (1.10–1.90)	0.73 (0.55–0.98)	<0.001
GGT (U/L)*	70 (40–125)	65 (38–110)	98 (55–210)	198 (140–380)	88 (50–160)	<0.001
ALP (U/L)*	92 (70–125)	105 (80–140)	110 (85–165)	125 (90–180)	118 (85–160)	0.012
Platelets (×10 ⁹ /L)	248±72	214±68	232±80	192±85	176±70	0.001

*Median (IQR), Continuous variables: ANOVA or Kruskal-Wallis; Categorical: χ^2 test; MASLD: Metabolic dysfunction-associated steatotic liver disease; DILI: Drug-induced liver injury; ALD: Alcohol-related liver disease; AIH: Autoimmune hepatitis; BMI: Body mass index; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: γ -glutamyl transferase; ALP: Alkaline phosphatase.

Model Development and Validation

Because this was a proof-of-concept study without established effect sizes, no formal a priori power calculation was performed. Enrollment was planned to include at least 20 patients in the smallest diagnostic class to support one-vs-rest AUROC estimation.

The dataset was split by stratified sampling into training (70%, n=196) and independent test (30%, n=84) sets. Four supervised algorithms were trained and compared: random forest, eXtreme gradient boosting, support vector machine with radial basis function kernel, and multilayer perceptron. Nested 5-fold cross-validation was performed within the training set, using macro-averaged F1-score as the primary optimization metric.

To address incorporation bias, we compared two modeling strategies. The basic model included demographic variables, clinical history, routine biochemistry, complete blood count, and metabolic parameters, but excluded viral serologies, autoantibodies, and immunoglobulin G. The extended model included all basic-model predictors plus HBsAg, anti-HCV, antinuclear antibody, anti-smooth muscle antibody, and immunoglobulin G. Class-specific AUROCs were compared using DeLong’s test.^[20]

Performance and Statistical Analysis

Final performance was assessed on the independent test set using overall accuracy, balanced accuracy, macro-precision, macro-recall, macro F1-score, and class-specific one-vs-rest AUROC. AUROC confidence intervals were calculated using the DeLong method,^[20] and other confidence intervals were estimated by bootstrapping with 1,000 iterations. SHapley Additive exPlanations (SHAP) were applied to the best-performing model to generate global and individual-level explanations.^[15,16]

Continuous variables are presented as mean±standard deviation or median with interquartile range, and categorical variables as counts and percentages. Between-group comparisons used analysis

of variance or Kruskal-Wallis tests for continuous variables and chi-square tests for categorical variables. Exact p values were reported where available. Analyses were performed in Python version 3.9.18. Two-sided p values <0.05 were considered statistically significant.

Results

Cohort Characteristics and Etiological Distribution

The cohort included 280 patients. The mean age was 51.8±15.1 years, and 152 patients (54.3%) were male. MASLD was the most frequent etiology (n=100, 35.7%), followed by viral hepatitis (n=78, 27.9%), DILI (n=50, 17.9%), ALD (n=32, 11.4%), and AIH (n=20, 7.1%). Each patient was assigned to a single etiological category.

Baseline characteristics are summarized in Table 1. Age, sex distribution, BMI, liver biochemistry, and platelet count differed significantly across etiological groups. BMI was highest in MASLD, whereas the AST/ALT ratio and GGT levels were highest in ALD.

Model Performance

The dataset was split by stratified sampling into training (n=196, 70%) and independent test (n=84, 30%) sets. The test set included 30 MASLD, 23 viral hepatitis, 15 DILI, 10 ALD, and six AIH cases.

Among the four evaluated algorithms, XGBoost showed the most consistent performance and was selected as the final model. On the independent test set, XGBoost correctly classified 72 of 84 patients, corresponding to an overall accuracy of 0.86 (95% CI 0.78–0.93). Balanced accuracy was 0.87 (95% CI 0.82–0.92), macro precision 0.87 (95% CI 0.82–0.92), macro recall 0.89 (95% CI 0.84–0.94), and macro F1-score 0.87 (95% CI 0.83–0.93) (Table 2).

Sensitivity Analysis and Multiclass Discrimination

The extended model outperformed the basic model on the independent test set. Overall accuracy increased from 0.73 to 0.86,

Table 2. Comparative performance of machine learning models on the independent test set (n=84)

Model	Accuracy (95% CI)	Balanced accuracy (95% CI)	Macro precision (95% CI)	Macro recall (95% CI)	Macro F1-score (95% CI)
XGBoost	0.86 (0.78–0.93)	0.87 (0.82–0.92)	0.87 (0.82–0.92)	0.89 (0.84–0.94)	0.87 (0.83–0.93)
Random forest	0.83 (0.74–0.91)	0.84 (0.78–0.90)	0.83 (0.77–0.89)	0.85 (0.79–0.91)	0.84 (0.78–0.90)
SVM (RBF)	0.80 (0.71–0.88)	0.81 (0.74–0.87)	0.79 (0.72–0.85)	0.82 (0.76–0.88)	0.80 (0.74–0.86)
MLP	0.81 (0.72–0.89)	0.82 (0.75–0.88)	0.81 (0.74–0.87)	0.83 (0.77–0.89)	0.82 (0.76–0.88)

CIs obtained via bootstrapping (1,000 iterations). CI: Confidence interval; SVM: Support vector machine; RBF: Radial basis function; MLP: Multilayer perceptron.

Table 3. Class-specific AUROC comparison between basic and extended models

Etiology	Basic model AUROC (95% CI)	Extended model AUROC (95% CI)	ΔAUROC	p-value (DeLong)
MASLD	0.88 (0.81–0.94)	0.95 (0.91–0.98)	+0.07	0.07
Viral hepatitis	0.83 (0.74–0.91)	0.94 (0.89–0.97)	+0.11	<0.02
DILI	0.86 (0.78–0.93)	0.93 (0.87–0.96)	+0.07	0.06
ALD	0.84 (0.74–0.92)	0.92 (0.86–0.96)	+0.08	0.08
AIH	0.80 (0.66–0.92)	0.95 (0.90–0.98)	+0.15	<0.03

AUROC comparisons performed using DeLong test for correlated ROC curves. CI: Confidence interval; AUROC: Area under the receiver operating characteristic curve; MASLD: Metabolic dysfunction-associated steatotic liver disease; DILI: Drug-induced liver injury; ALD: Alcohol-related liver disease; AIH: Autoimmune hepatitis.

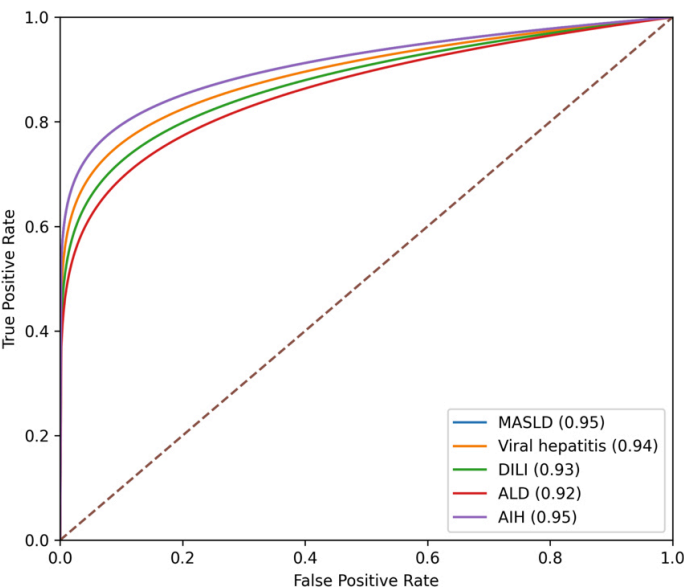


Figure 1. One-vs-rest receiver operating characteristic (ROC) curves for the extended XGBoost model. ROC curves demonstrating class-specific discrimination of the extended XGBoost model on the independent test set (n=84). Area under the ROC curve (AUROC) values (95% confidence intervals, DeLong method) were: (i) Metabolic dysfunction-associated steatotic liver disease: 0.95 (0.91–0.98). (ii) Viral hepatitis: 0.94 (0.89–0.97). (iii) Drug-induced liver injury: 0.93 (0.87–0.96). (iv) Alcohol-related liver disease: 0.92 (0.86–0.96). (v) Autoimmune hepatitis: 0.95 (0.90–0.98). The diagonal dashed line represents chance-level performance (AUROC=0.50). Class-specific ROC curves were generated using a one-vs-rest approach.

and macro F1-score improved from 0.72 to 0.87. The largest AUROC gains were observed for AIH (0.80 to 0.95; p=0.03) and viral hepatitis

(0.83 to 0.94; p=0.02). Improvements for MASLD, DILI, and ALD were modest and not statistically significant (Δ AUROC $\leq$ 0.08), including DILI (Δ AUROC=0.07; p=0.06) (Table 3). Because the independent test set included only six AIH cases, class-specific estimates for AIH should be interpreted cautiously.

One-vs-rest analysis showed strong class-specific discrimination for the extended XGBoost model (Fig. 1). AUROC values were 0.95 for MASLD, 0.94 for viral hepatitis, 0.93 for DILI, 0.92 for ALD, and 0.95 for AIH.

Explainability and Misclassification Analysis

SHAP analysis identified GGT, AST/ALT ratio, platelet count, age, and BMI as the most influential predictors (Fig. 2a). Higher GGT and a higher AST/ALT ratio contributed to ALD predictions, whereas higher BMI favored MASLD. Among serological markers, anti-HCV and ANA contributed mainly to viral hepatitis and AIH classification, respectively. A representative local SHAP plot for a DILI case is shown in Figure 2b.

Overall, 12 patients (14.3%) were misclassified by the extended model. The most frequent bidirectional error involved DILI and AIH (n=6), including four DILI cases classified as AIH and two AIH cases classified as DILI. Four misclassifications occurred between ALD and MASLD, and two occurred between DILI and viral hepatitis. These errors largely reflected overlapping clinical and biochemical features rather than systematic overprediction of a single category (Table 4).

Discussion

In this prospective proof-of-concept cohort study, an explainable multiclass machine learning model differentiated five major etiologies of persistent liver enzyme elevation with good performance on an

Table 4. Systematic characterization of misclassified cases (n=12)

True diagnosis	Predicted diagnosis	Key features	Contributing factors
DILI	AIH	ANA 1:80, recent antibiotic use	Borderline RUCAM (5–6)
DILI	AIH	Mixed hepatocellular-cholestatic injury	Low-titer autoantibodies
DILI	AIH	ALT 180 U/L	ANA positive
DILI	AIH	Drug exposure	Overlapping serology
AIH	DILI	Recent antibiotic use	Incomplete antibody profile
AIH	DILI	No biopsy	Borderline IgG elevation
ALD	MASLD	BMI 34 kg/m ²	Alcohol near threshold
ALD	MASLD	Steatosis on USG	Metabolic syndrome
ALD	MASLD	GGT elevated	Overlapping metabolic features
DILI	Viral hepatitis	Mild anti-HCV positivity	Drug exposure
Viral hepatitis	DILI	ALT 160 U/L	Recent medication
MASLD	ALD	AST/ALT ratio borderline	Alcohol intake near diagnostic threshold

Misclassifications reflected overlapping clinical and biochemical patterns rather than systematic model bias. MASLD: Metabolic dysfunction-associated steatotic liver disease; DILI: Drug-induced liver injury; ALD: Alcohol-related liver disease; AIH: Autoimmune hepatitis; BMI: Body mass index; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: γ-glutamyl transferase; ANA: Antinuclear antibody; USG: Ultrasound; RUCAM: Roussel uclaf causality assessment method; IgG: Immunoglobulin G.

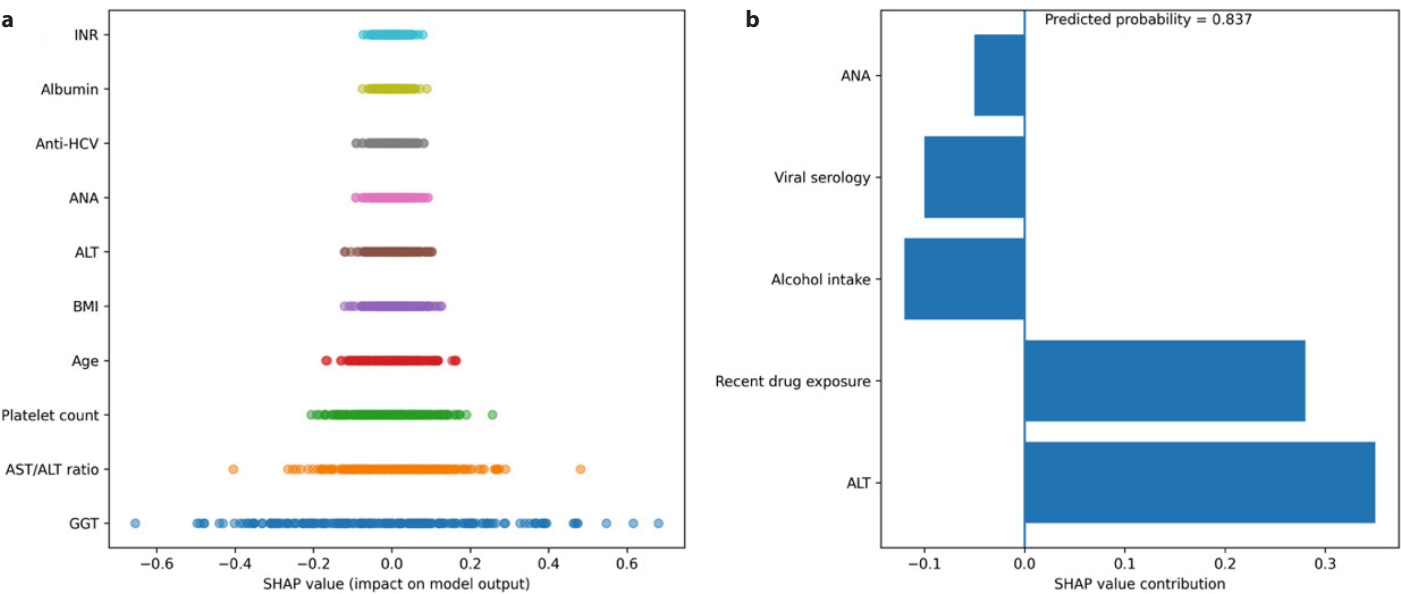


Figure 2. SHAP-based interpretability analysis of the extended model **(a)** Global feature importance (SHAP summary plot). Features are ranked according to mean absolute SHAP values, reflecting their overall contribution to model predictions across the independent test set. Each point represents an individual patient. Horizontal dispersion reflects the magnitude and direction of feature impact on model output. **(b)** Representative local SHAP contribution plot for a DILI case. Horizontal bar plot illustrating feature-level contributions to the predicted probability of drug-induced liver injury. Positive values increase predicted probability, whereas negative values decrease it. In this example, elevated ALT and recent drug exposure contributed positively, while low alcohol intake and negative viral serology reduced competing etiologies. The final predicted probability for DILI was 0.837.

SHAP: SHapley Additive exPlanations; DILI: Drug-induced liver injury; ALT: Alanine aminotransferase.

independent test set. The final XGBoost model achieved an overall accuracy of 0.86, and class-specific discrimination was consistently high across MASLD, viral hepatitis, DILI, ALD, and AIH. SHAP-based analysis showed clinically plausible predictor patterns, supporting the interpretability of the model outputs.^[15,16] These findings suggest that multiclass AI may support the structured evaluation of patients with unexplained liver enzyme abnormalities, particularly when several etiologies are clinically plausible.

A key concern in this setting is incorporation bias, because some predictors are also used in diagnostic adjudication. To address this, we compared a basic model excluding disease-defining serological markers with an extended model including viral serologies, autoantibodies, and immunoglobulin G. The extended model improved macro F1-score from 0.72 to 0.87. The largest AUROC gains were seen for AIH (0.80 to 0.95; $p=0.03$) and viral hepatitis (0.83 to 0.94; $p=0.02$), whereas improvements for MASLD, DILI, and ALD were modest and not statistically significant ($\Delta\text{AUROC} \leq 0.08$). The AIH estimates should be interpreted cautiously because only six AIH cases were included in the independent test set.

Most previous AI studies in hepatology have focused on binary or single-disease tasks, such as steatotic liver disease detection, fibrosis assessment, hepatocellular carcinoma risk stratification, or outcome prediction.^[5–7,11,12] In contrast, the present study addressed a multiclass diagnostic problem that more closely reflects routine hepatology practice. Patients with persistent liver enzyme elevation often have overlapping metabolic, alcohol-related, drug-related, viral, or autoimmune features.^[1,2] A model that estimates several etiologies simultaneously may therefore be more clinically useful than a single-disease classifier in this setting.

The SHAP analysis supported the clinical plausibility of the model. GGT and the AST/ALT ratio contributed to ALD prediction,^[18] whereas BMI favored MASLD classification.^[1] Anti-HCV and ANA contributed mainly to viral hepatitis and AIH classification, respectively. Thus, the model appeared to integrate clinical, biochemical, and serological information rather than relying on a single diagnostic marker. The misclassification analysis also supported real-world diagnostic overlap. The most frequent bidirectional error involved DILI and AIH ($n=6$), a recognized diagnostic gray zone.^[2,19] Additional errors occurred between ALD and MASLD, and between DILI and viral hepatitis. These patterns suggest clinical overlap rather than systematic model bias.

The main clinical value of this approach lies in structured triage rather than autonomous diagnosis. If externally validated, such a model could help clinicians prioritize confirmatory testing, medication review, alcohol-use assessment, or autoimmune evaluation. However, this study did not assess effects on clinician behavior, diagnostic testing, time to diagnosis, or patient outcomes.^[21,22] Strengths include prospective consecutive enrollment, blinded adjudication by three hepatologists, separation of training and independent test sets, and nested cross-validation.^[17] Limitations include the single-center design, limited AIH sample size, possible residual incorporation bias, limited assessment of isolated cholestatic presentations, and non-uniform use of liver biopsy. External validation in larger multicenter cohorts, calibration analysis, and decision-curve analysis are needed before clinical implementation.^[14,23]

Conclusion

An explainable multiclass machine learning model differentiated five major etiologies of persistent liver enzyme elevation using routinely available clinical and laboratory data. Serological markers added value for AIH and viral hepatitis, while routine clinical and biochemical variables retained triage value for MASLD, DILI, and ALD. External validation is required before clinical implementation.

Online Appendix Link: <https://file.papilino.com/hepatologyforum/files/1785308991-appendix-en.pdf>

Ethics Committee Approval: The study was approved by the Kayseri City Hospital Institutional Ethics Committee (Approval No: 614; October 13, 2025).

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Use of AI for Writing Assistance: The authors declare that no artificial intelligence was used in this study.

Author Contributions: Concept: YO; Design: YO; Supervision: YO, NMB; Materials: YO; Data Collection and/or Processing: YO, OY; Analysis and/or Interpretation: YO, OY, NMB; Literature Search: YO, OY; Writing: YO; Critical Reviews: YO, OY, NMB.

Acknowledgments: The authors are grateful to the clinical staff for their support in patient evaluation and systematic data collection.

Peer-review: Externally peer-reviewed.

References

1. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, Kleiner DE, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology* 2023;77(5):1797-1835. [CrossRef]
2. Chalasani NP, Maddur H, Russo MW, Wong RJ, Reddy KR. ACG Clinical Guideline: Diagnosis and Management of Idiosyncratic Drug-Induced Liver Injury. *Am J Gastroenterol* 2021;116(5):878-898. [CrossRef]
3. Atabaki-Pasdar N, Ohlsson M, Whiting S, Frau F, Pomares-Millan H, Haid M, et al. Predicting and elucidating the etiology of fatty liver disease: a machine learning modeling and validation study in the IMI DIRECT cohorts. *PLoS Med* 2020;17(6):e1003149. [CrossRef]
4. Ahn JC, Connell A, Simonetto DA, Hughes C, Shah VH. Application of artificial intelligence for the diagnosis and treatment of liver diseases. *Hepatology* 2021;73(6):2546-2563. [CrossRef]
5. Ghosh S, Zhao X, Alim M, Brudno M, Bhat M. Artificial intelligence applied to 'omics data in liver disease: towards a personalised approach for diagnosis, prognosis and treatment. *Gut* 2025;74(2):295-311. [CrossRef]
6. Huang G, Jin Q, Mao Y. Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: prospective cohort study. *J Med Internet Res* 2023;25:e46891. [CrossRef]
7. Calderaro J, Seraphin TP, Luedde T, Simon TG. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J Hepatol* 2022;76(6):1348-1361. [CrossRef]
8. Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, Huang K. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun* 2021;12(1):3445. [CrossRef]
9. Clift AK, Coupland CAC, Hippisley-Cox J. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. *BMJ* 2023;380:e073800. [CrossRef]

10. Ye J, Zhuang X, Li X, Gong X, Sun Y, Wang W, et al. Novel metabolic classification for extrahepatic complication of metabolic associated fatty liver disease: a data-driven cluster analysis with international validation. *Metabolism* 2022;136:155294. [\[CrossRef\]](#)
11. Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan WK, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol Hepatol* 2020;5(4):362-373. [\[CrossRef\]](#)
12. Fujii H, Iwaki M, Hayashi H, Toyoda H, Oeda S, Hyogo H, et al. Clinical outcomes in biopsy-proven nonalcoholic fatty liver disease patients: a multicenter registry-based cohort study. *Clin Gastroenterol Hepatol* 2023;21(2):370-379. [\[CrossRef\]](#)
13. Choi BK, Choi YJ, Sung MD, Ha WS, Chu MK, Kim WJ, et al. Development and validation of an artificial intelligence model for the early classification of the etiology of meningitis and encephalitis: a retrospective observational study. *EClinicalMedicine* 2023;61:102051. [\[CrossRef\]](#)
14. Mohsen F, Al-Absi HRA, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digit Med* 2023;6(1):197. [\[CrossRef\]](#)
15. Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: a systematic literature review. *Comput Biol Med* 2023;166:107555. [\[CrossRef\]](#)
16. Nazir S, Ali S, Khan A, Iqbal J, Daudpota SM, Akram MU. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. *Comput Biol Med* 2023;153:106668. [\[CrossRef\]](#)
17. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162(1):55-63. [\[CrossRef\]](#)
18. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of alcohol-related liver disease. *J Hepatol* 2018;69(1):154-181. [\[CrossRef\]](#)
19. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Autoimmune hepatitis. *J Hepatol* 2015;63(4):971-1004. Erratum in: *J Hepatol* 2015;63(6):1543-4. [\[CrossRef\]](#)
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837-845. [\[CrossRef\]](#)
21. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195. [\[CrossRef\]](#)
22. Allam A, Feuerriegel S, Rebhan M, Krauthammer M. Analyzing patient trajectories with artificial intelligence. *J Med Internet Res* 2021;23(12):e29812. [\[CrossRef\]](#)
23. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:m1328. [\[CrossRef\]](#)