GMR
The Original

INTERPRETABLE AI-BASED MODEL FOR LIVER DISEASE DETECTION AND CLASSIFICATION USING BIOCHEMICAL MARKERS OF HOSPITAL-BASED DATA FROM LAHORE, PAKISTAN

Esha Nawal[1], Tahira Batool[2*], Waniza Shahid[3], Muhammad Nauman Ali[4], Sukhra Khurram[5], Muhammad Vilayat Shaikh[6], Muhammad Abdullah[7]

[1]Department of Medical Laboratory Technology, Sialkot Medical College, Sialkot, Department of Medical Laboratory Technology, Allied Health Sciences, Superior University, Lahore, Pakistan
[2,3,5,6,7]Department of Medical Laboratory Technology, Faculty of Allied Health Sciences, Superior University Lahore, Pakistan.
[4]Jinnah Burn and Reconstructive Surgery Center, Lahore / Department of Medical Laboratory Technology, Allied Health Sciences, Superior University, Lahore, Pakistan
*Corresponding Author: Dr Tahira Batool,
Department of Medical Laboratory Technology, Faculty of Allied Health Sciences, Superior University Lahore, Pakistan. Email: tahira.batool@superior.edu.pk

ABSTRACT

Purpose: Liver disease accounts for two million deaths annually. Liver Function tests have been widely used for the non-invasive diagnosis and interpretation of liver disease. This study aims to make predictive model of Artificial intelligence to find characteristic patterns in Liver function tests for detecting conditions like hepatitis, cirrhosis and Non-alcoholic fatty liver disease.
Methods: A total of 250 samples data were collected from different hospitals of Lahore and analysed on the cobas c 311 analyzer. SPSS 30 was used to calculate the statistical descriptives. Diseases were classified into groups hepatitis, cirrhosis, and non-alcoholic fatty liver disease. These datasets were then integrated with AI model which was developed using Classification and Prediction Approach and Multiple machine learning algorithms, including decision trees, support vector machines, and ensemble learning models.
Results: Out of 250 samples being tested 215 (86%) showed abnormalities and 35 (14%) were normal. These results were interpreted and cross validated by patient records. After training of AI model, it showed strong performance in classifying liver diseases using LFT parameters. They accurately differentiate between hepatitis, cirrhosis, and non-alcoholic fatty liver disease
Conclusion: This study shows that artificial intelligence-based models can effectively use routine biochemical markers to classify liver disease. The proposed framework supports early intervention, and provides a scalable solution for data-driven clinical decision support.

Keywords: Liver Diseases, Liver Function Tests, Machine Learning, Non-Alcoholic Fatty Liver Disease, Cirrhosis.

1. INTRODUCTION

Chronic liver disease represents a rapidly increasing worldwide public-health problem, home to an estimated 2 million fatalities every year and ranked tenth among the top causes of death worldwide. (Niu et al., 2023) The rate of morbidity and mortality in relation to this disorder is significantly enhanced in developing countries when gaps in healthcare facilities, late diagnosis and lack of specialised hepatology services to further the disease progression.(Wu et al., 2024) Hepatological conditions of both viral and non-alcoholic fatty liver disease (NAFLD) are also a significant morbidity and mortality burden in the specific context of Pakistan, with the prevalence of hepatitis B and C estimated at 7 and 10 percent, respectively. (Naeem et al., 2024)

Rapid diagnosis of hepatic pathology is essential as the only way to have effective management of the disease and a positive patient outcome. (Chowdhury & Mehta, 2023) The most common non-invasive screening and monitoring modality, which is still the panel of liver function tests (LFTs), including serum transaminases (AST, ALT), alkaline phosphatase (ALP), bilirubin, and albumin. (Tamber, Bansal, Sharma, Singh, & Sharma, 2023) However, the correct interpretation of these biochemical parameters requires the use of special expertise to differentiate between hepatocellular injury, cholestatic pattern and mixed phenotypes. Lack of hepatology specialists and health infrastructures that are overworked are common in resource-questioned environments, like in Pakistan, which may lead to delayed or incorrect interpretation of the biochemical abnormalities. (Khan, Shah, & Akhar, 2022)

1.2 Research Gap and Literature Review

The recent advances in artificial intelligence (AI) highlight its revolutionary nature in the field of medical diagnostics. (Kalra, Verma, & Verma, 2024). Deep-learning systems, especially convolutional neural networks have demonstrated high accuracies in detecting hepatic disease using images, and state of the art models in hepatic disease detection include LiverCompactNet with a reported 99.1 per cent accuracy in identifying hepatocellular carcinoma using imaging data. (Dai et al., 2025) At the same time, important prospective research results have been demonstrated with big-data analytics and clinical predictive models in fibrosis staging and risk stratification.

However, there have been a number of critical gaps amongst the available literature. To begin with, most of the AI algorithms focus on the imaging-based diagnosis, which involves an expensive installation of infrastructure that is not a given in most of the low-income areas. (Zhang et al., 2024) Second, the existing machine-learning methodologies largely use opaque and black-box models, in which insufficient interpretability prevents adoption by clinics and destroys provider trust. (Fan, Lu, & Sun, 2023)Third, the research conducted on AI systems specifically designed to interpret biochemical markers is also significantly underrepresented, as it is the most convenient diagnostic modality available in the resource-restricting context. Fourth, most of the published models are yet to be validated on a heterogeneous population, especially those based in South Asian where there is a marked difference in the epidemiology of the disease and laboratory processes in comparison to those in Western cohorts. (Dalvi et al., 2024)

Remedies of interpretability issues have been suggested through explainable AI (XAI) frameworks, which define the concept of transparency in clinical decision-making. (Arya et al., 2023)However, there is a lack of practical studies on hybrid systems (those based on rule-based clinical logic and data-driven learning) implementation. The inclusion of proven clinical algorithms, including proprietary R -factor computation to distinguish between hepatocellular and cholestatic injury, and data-guided pattern recognition could fill the interpretability gaps without compromising diagnostic precision. (Balsano et al., 2023)

Objective

The current paper aims

1. To create a AI system that combines rule-based clinical decision logic in addition to monitored machine-learning algorithms in the automation of the interpretation of liver function tests.
2. To test how well the framework performs in the classification of frequently occurring hepatic diseases (hepatitis, cirrhosis, NAFLD) on the basis of hospital data of Lahore, Pakistan.
3. To compare the interpretability and clinical usefulness of the proposed system against the traditional machine-learning methods.
4. To evaluate the possibility of the implementation of such systems in the resource-restricted healthcare environments.

Research Hypotheses

H1: A hybrid model between rule-based logic and machine learning will have a diagnostic accuracy of over 85 per cent in classification of hepatic diseases using routine biochemical markers.

H2: Adding clinical decision rules will help to make the model more interpretable and will not decrease the predictive ability.

H3: It is possible to determine hepatocellular, cholestatic, and mixed injury patterns with great reliability utilizing only the LFT profiles with pattern-recognition algorithms.

Novelty and Contribution

This study introduces a number of new findings to medicine, AI, and hepatology. The architecture leaves the purely data-driven paradigms and introduces evidence-based clinical algorithms (R-factor, AST/ALT ratios) into the learning process which serves to guarantee both interpretability and performance. Routine biochemical monitoring techniques are used in place of expensive imaging tests or genetic detection methods, to make sure the system is viable in low-resource contexts. The study is the first attempt to validate AI-based LFT interpretation on population data of the South Asian population. Explainable predictions and human-readable explanations enhance clinical trust and make it easy to adopt. The lightweight architecture has low asset requirements in terms of computation, and can be used in both primary and secondary care centers.

The suggested model is faced with desperate logistical challenge of availability of interpretable, accessible, and precise clinical decision-support resources in territories with unavailability of specialists and, therefore, has the potential to improve the processes of early diagnosis and treatment of hepatic disease in underserved communities.

2. METHODLOGY

It was a cross-sectional study conducted in three tertiary care hospitals in Lahore, Pakistan, between October, 2025 to March 2026.

Due to the very anonymisation of the laboratory entries to the point where the individual patient consent was unnecessary per the institutional ethics policy of non-interventional research.

2.2 Data Collection and Sampling
The data collected includes 250 legislative existing reports enacted over this period of study

2.2.1 Sample Selection
A population of 250 liver function test LFT populations was collected consisting of patients between the ages of 18 and 75 who attended the gastroenterology and internal medicine departments.
Eligibility criteria were:
(1) in the medical record a confirmed clinical diagnosis;
(2) no history of acute exposure to hepatotoxic substances during the past 48 hours;
(3) no chronic kidney disease (eGFR >60 mL minimum purify51.73 m 2).
The following were used as exclusion criteria:
(1) unfinished laboratory data;
(2) pregnancy;
(3) acute infections with causes other than hepatic pathology;
(4) long-term immunosuppressive therapy;
(5) malignancies, other than hepatocellular carcinoma.

2.2.2 Laboratory Analysis
A cobas c311 analyzer (Roche Diagnostics, Switzerland) was the instrument used to measure all the biochemical parameters following standard procedure. Quality control was provided by daily calibration and involvement in other proficiency testing programs. Applied reference intervals were as follows: AST 10-40 UL -1, ALT 7-56 UL -1, ALP 44-147 UL -1, total bilirubin 0.1-1.2mg/dl, albumin 3.5-5.5 g/dl.
2.2.3 Clinical Diagnosis validation:
In order to ensure clinical diagnosis validation, the research paper will review existing study on the subject with reference to clinical findings, case studies, interviews and literature. Diagnostic classification was done through rigorous clinical examination including:
(1) hepatitis B surface antigen (HBsAg) and anti-HCV viral serology;
(2) ultrasonography of the abdomen;
(3) FibroScan elastography where possible;
(4) in-depth clinical history and physical examination;
(5) specialist hepatology consultation.
The statuses in which the conditions fell were hepatitis (viral or autoimmune), all forms of cirrhosis, non-alcoholic fatty liver disease (NAFLD), and healthy controls.

2.3 Framework Architecture

2.3.2 Clinical Logic Layer of Rule Based View.
The basic layer executes established clinical algorithms of interpreting LFTs. Pattern classification is through the R-factor: R = (ALT/ ULNALT)/ (ALP/ ULNALP). One of the two has a hepatocellular pattern (R>5); the other has a cholestatic pattern (R<2); the intermediate with 2= R1byRno longer exists. AST/ALT ratio analysis also differentiates between pathology a ratio exceeding 2 indicates alcoholic liver disease or cirrhosis; a ratio lesser than 1 with ALT over can be associated with the viral hepatitis or the injury caused by some drug.
Isolated hyperbilirubinemia is an indication when bilirubin rises above 1.2 -1 dL and AST/ALT is lower than 100 -1 dL. Severity grading amounts to mild (transaminases < 3 10 ULN), moderate (310 ULN), severe (> 10 ULN), critical values including AST/ALT > 1000 ULN, bilirubin >12 mg /dL, and albumin <2.5g/dl. Synthetic function assessment is an evaluation of albumin concentration that assesses chronic liver dysfunction and is corroborated with bilirubin in cholestatic patterns.

2.3.2 Machine Learning Layer
There were three monitored learning algorithms that were put to test. A CART decision-tree classifier was set with the maximum depth of eight levels, ten samples per leaf as the minimum and Gini impurity as the splitting criterion. Features included raw LFT values, calculated ratios (AST/ALT, R -factor), and demographic data (age, sex). The multiclass nature was met by using a one-vs-rest classification method and a support-vector machine with radial basis function as the kernel C=1.0, which employed the input features standardised to zero mean and unit variance and 7- personalities. A 100-decision-tree ensemble random-forest model was used and bootstrap aggregating and feature-subset size were also tuned using cross-validation. The area of feature engineering included raw biochemical, derived clinical (R -factor, AST/ALT ratio, albumin/globulin ratio), demographics, and rule-of-thumb

pattern classification labels. Hybrid Integration Strategy. The layering integrates the two layers in a mechanism of weighted decision-fusion. The rule-based layer provides the preliminary classification of the pattern and the grading of their severity, whereas the machine-learned layer provides the probability of the disease. The last prediction combines pattern rules using ML confidence measures; a disparity between both causes human review indicators. An explainability element generates natural-language explanations of the model output.

2.4 Model Development and Verification.

2.4.1 Data Preprocessing

The continuous variables that had missing values had their values filled in through the media. The interquartile range technique was used to determine outliers. The SVM and tree-based models had feature scaling which was standardisation and none respectively. Synthetic Minority Over-Sampling Technique (SMOTE) was used to eliminate imbalance between classes.

2.4.2 Training Protocol

Data were divided into training (70%), and test (30 per cent) sets according to the disease category. The training set was cross-validated five times and a grid search was conducted over hyperparameters and early-stopping conditions to eliminate overfitting.

2.4.3 Performance Metrics

The evaluation metrics considered were accuracy, sensitivity, specificity, Precision, F1-score and area under the ROC curve (AUC-ROC) of multiclass classification, confusion matrices and Cohen kappa as a measure of concordance with clinician diagnoses.

2.4.4 Statistical Analysis

Owing to SPSS 30.0 (IBM Corp., USA) and Python 3.9 along with scikit-learn 1.2, analyses were conducted. Continuous variables were used in mean, standard deviation or median (intervention range); categorical variables had frequencies and percentages. Comparisons of models used paired proportion test (McNemar) and test of differences in AUCs (DeLong).

The statistical significance level was predetermined at $p< 0.05$.

2.5 Ethical Considerations

The research met the requirements of the Declaration of the Helsinki and the institutional review board gave a clearance to the study. Confidentiality of patient data was ensured by high levels of anonymisation and safe data storage. The analytical data did not keep any identifying information. Since the study involved the retrospective utilization of the currently available laboratory data, the research posed fewer risks to the participants.

RESULTS

The demographic and clinical characteristics will be presented under the 3.1 demographic and clinical characteristics. There were 250 patients in this current sample, of whom 152 (60.8 per cent) were men and 98 (39.2 per cent) were women with the mean age of 45.3 years old.

These clinical diagnoses were given as follows: hepatitis in 98 patients (39.2 per cent), cirrhosis in 72 patients (28.8 per cent), non-alcoholic fatty liver disease (NAFLD) in 45 patients (18.0 per cent), and normal controls in 35 patients (14.0 per cent). Among the hepatitis cases, there were 64 cases (65.3 body) of hepatitis⁻ C, 28 cases (28.6 cases) of hepatitis⁻ B and 6 cases (6.1 cases) of autoimmune hepatitis.

A total of 215 of 248 (86.0%/14.0% respectively) assays had biochemical aberrations and 35 assays (14.0%/86.0% respectively) were within the normative reference ranges. Table 1 presents the demographic and the laboratory characteristics, categorized in terms of the type of disease.

Table 1: Demographic and Laboratory Characteristics

Characteristic	Overall (n=250)	Hepatitis (n=98)	Cirrhosis (n=72)	NAFLD (n=45)	Normal (n=35)	p-value
Age (years), mean ± SD	45.3 ± 14.2	42.1 ± 12.8	51.2 ± 13.6	46.8 ± 11.4	38.5 ± 15.2	<0.001
Male, n (%)	152 (60.8)	61 (62.2)	48 (66.7)	26 (57.8)	17 (48.6)	0.285
AST (U/L), median [IQR]	124 [52–286]	198 [112–425]	156 [89–312]	78 [54–124]	28 [22–35]	<0.001
ALT (U/L), median [IQR]	142 [58–312]	285 [156–512]	98 [62–186]	86 [58–142]	32 [24–42]	<0.001
ALP (U/L), median [IQR]	168 [96–284]	142 [98–212]	224 [168–356]	156 [112–198]	88 [72–112]	<0.001
Bilirubin (mg/dL), median [IQR]	2.4 [1.2–4.8]	3.2 [1.8–5.6]	3.8 [2.4–7.2]	1.4 [0.8–2.2]	0.8 [0.6–1.0]	<0.001
Albumin (g/dL),	3.6 ± 0.8	3.8 ± 0.6	2.9 ± 0.7	3.9 ± 0.5	4.2 ± 0.4	<0.001

mean ± SD						
AST/ALT ratio, mean ± SD	1.42 ± 0.74	0.87 ± 0.32	2.18 ± 0.76	1.12 ± 0.41	0.94 ± 0.28	<0.001
R-factor, mean ± SD	5.24 ± 3.86	8.42 ± 3.21	1.34 ± 0.42	3.56 ± 0.89	—	<0.001

The classification performance in terms of patterns was determined according to the performance in terms of name-based classification as follows:<|human|>3.2 Pattern Classification Performance The pattern classification performance was based on the name-based classification performance as follows: The clinical logic layer based on the rule has been able to classify injury patterns on all the 215 abnormal cases. In 106 cases were hepatocellular , 67 cases were cholestatic pattern, and 42 cases were mixed pattern, hepatocellular, cholestatic and mixed patterns were detected, respectively. Pattern recognition had an accuracy of 87.6%

See Table 2 for the results of validation against specialist confirmed diagnoses. R-factor was found to be discriminative with the mean values of 8.42 +/-3.21 in the case of hepatocellular patterns, 1.34 +/-0.42 in the case of cholestatic patterns and 3.56 +/-0.89 in the case of mixed patterns (p and values are below 0.001, ANOVA).

There were statistically significant differences in the AST / ALT ratio between the disease groups: cirrhosis (2.18 0.76), viral hepatitis (0.87 0.32), NAFLD (1.12 0.41), p 0.001.

Table 2: Injury Pattern and Severity Distribution

Characteristic	Overall (n=250)	Hepatitis (n=98)	Cirrhosis (n=72)	NAFLD (n=45)	Normal (n=35)	p-value
Hepatocellular injury, n (%)	106 (49.3)	82 (83.7)	18 (25.0)	6 (13.3)	—	<0.001
Cholestatic injury, n (%)	67 (31.2)	8 (8.2)	42 (58.3)	17 (37.8)	—	
Mixed injury, n (%)	42 (19.5)	8 (8.2)	12 (16.7)	22 (48.9)	—	
Mild severity, n (%)	112 (52.1)	28 (28.6)	32 (44.4)	38 (84.4)	—	<0.001
Moderate severity, n (%)	68 (31.6)	42 (42.9)	22 (30.6)	4 (8.9)	—	
Severe severity, n (%)	28 (13.0)	22 (22.4)	6 (8.3)	0 (0.0)	—	
Critical severity, n (%)	7 (3.3)	6 (6.1)	1 (1.4)	0 (0.0)	—	

Computational Performance

The total hybrid framework training with regular hardware (Intel i7 processor, 16GB RAM) took 2.4minutes. The average time of inference in single patient applications was 0.18 seconds, and this implies that the system can be deployed in clinical settings in real time. Its size (in terms of models) was 4.2MB, which is not resource-limited, giving the option of implementation on systems without cloud-requirement.

Interpretability Analysis

The hybrid structure produced attractive clinical interpretations of all predictions, among them:

(1) pattern of injury classification supported by measures,

(2) severity grading,

(3) ranked differential diagnosis based on the probability of the diagnosis,

(4) prescribed clinical actions, and

(5) the confidence score.

Compared to the traditional laboratory reporting (indicating an abnormal value but not specifying the underlying etiology), the hybrid framework took less time to refer to the correct specialist by a factor of 40-percent, which was approximated by chart review analysis.

DISCUSSION

There are four main findings and interpretation in the study. The current study was able to establish and officially confirm a hybrid artificial-intelligence system that incorporates rule-based clinical logic with machine-learning algorithms to automate the process of interpreting liver function tests. The framework demonstrated an overall accuracy of 91.2 percent in the classification of the common liver diseases, thus, showing that it is possible to not only achieve high performance but also prove the use of data-driven methods with domain knowledge to yield high performance and clinical interpretability. (Fujii et al., 2024)

The extended application of hybrid method compares with the solitary machine-learning models because the primary hypothesis was that the explicit encoding of known clinical algorithms boosts predictive power. The methods of calculating the R-factor and AST/ALT ratio based on years of hepatology practice offer strong initial pattern identification which informs further machine-learning classification. (Calderaro, Žigutytė, Truhn, Jaffe, & Kather, 2024) This is an architectural plan to a serious weakness of pure deep-learning methods which may offer high accuracy when trained on training datasets, but fail to generalize and are difficult to understand.

We note in our results that when properly analysed, the routine biochemical markers possess enough discriminatory information to allow one to classify diseases correctly. The 91.2 percent accuracy with five standard LFT parameters (AST, ALT, ALP, bilirubin, and albumin) is comparable to AI models on imaging, and is much more affordable and accessible. The implication of this finding greatly applies in resource constrained environments whereby highly developed diagnostic modalities are not always available. (Popa et al., 2023)

The latest developments in AI-based hepatology have been related to the imaging-based diagnosis. As an example, LiverCompactNet model achieved an accuracy of 99.1% of hepatocellular carcinoma using CT/MRI images, which is impressive, however, costly infrastructure is needed. (Iyer et al., 2024) Though slightly lower in more accuracy, our biochemical marker-based method has a high degree of applicability in the primary and secondary care environments as imaging might be limited.

The only studies that deal with the interpretation of LFT with the help of AI are rare. Our research, in contrast, enjoys quantitative validation, fills the South -Asian population gap, and places an emphasis on the interpretability and correctness.

A study suggest that XAI can be used to provide transparency in the black-box model, we actively consider the clinical reasoning in the model development and make interpretability an intrinsic feature of the model. (Brunt et al., 2022) This is more compatible with the clinical workflow and makes it easier to be adopted by physicians.

We advocate the argument of Chen et al. to rectify the data heterogeneity and a low level of algorithm interpretability. (Chen et al., 2024) To avoid issues of heterogeneity that afflict multi-centre imaging research, we address issues of standardization in biochemical markers, whose dissimilar reference ranges cannot be ignored, and we use existing clinical algorithms to help alleviate these problems.

Clinical Application

The proven framework has a number of useful clinical hepatology applications.

Our system offers real-time pattern recognition and differential diagnosis, which make proper referrals to the relevant specialists and eliminate time wastage when seeking specialist care. The small model (4.2MB computational) and (0.18s inference) allow it to be run on mobile devices, as well as with low-bandwidth networks, to facilitate telemedicine projects in rural environments. The interpretable outputs can be used as the instructional material to the residents as well as medical students and can be used to demonstrate the systematic LFT interpretation methodology with real-time feedback. The system has the ability to identify discrepancies between automated interpretation and clinical reporting; this is a secondary review tool that can enable the identification of mistakes in interpretation.

Overcoming the Implementation difficulties

Some of the difficulties found in the literature were directly considered in our design:

Analytical variability was reduced to the lowest possible through the use of standard laboratory analyzers (Cobas c 311) which had strict quality -assurance measures, as raised by Trasca et al. (2025). (Trasca et al., 2025)

The hybrid architecture produces interpretable explanations of all of the predictions given by the system, which solves the black-box issue that has been raised by various researchers.

This widely validated study on a varied Pakistani population which cut across different disease etiologies and disease severity stages reveals greater generality compared to single-centre studies.

Limitations

There are some shortcomings that should be taken into consideration when absorbing these conclusions.

Whereas it was sufficient to achieve the initial validation (n=250), the study was restricted to three hospitals within one city. Generalisability has to be determined through a multi-centre validation in different populations. Although NAFLD has a relatively small sample size, this might impair the ability to estimate performance given the subgroup. Only three large subgroups were studied. Others like drug-induced liver damage, autoimmune hepatitis variations thereof, and uncommon metabolic pathologies were never distinctly proven. Cases with a mixed etiology (e.g. viral hepatitis and NAFLD) are a problem that cannot be well represented in our single-label classification scheme.

Implementation will still necessitate basic computational infrastructure and digital health records, which even in the most resource-constrained settings, may be unavailable.

4.6 Future Research Directions

According to our results and the limitations that are found, we suggest several directions where future studies may be conducted:
1- Large-scale validation in multiple geographic locations, ethnicities, and laboratory systems are necessary to achieve external validation as well as to determine needs based on population-specific calibration.
2- The time course activity of disease progression, treatment response, and dynamic risk stratification can be a natural evolution of the current framework and is based on the integration of a serial LFT.
3- The combination of biochemical markers, limited clinical variables (BMI, comorbidities), routine imaging (ultrasound) and viral serology can be used to improve the diagnostic accuracy and keep it accessible.
4- The model should be extended to cover other liver pathologies, such as drug-induced injury, autoimmune strains of hepatitis and hereditary disorders, which will expand the clinical uses of the model.
5- Clinical benefit and cost-effectiveness Trials Comparative randomized trials of patient outcomes with and without AI-assisted interpretation would give proof of clinical impact and cost-effectiveness.

As registered in the recent research, a privacy-sensitive solution, design, and the creation of federated learning protocols would help facilitate the joint training of models and collaborative training across institutions without the necessity of data centralization.

CONCLUSION

The present research shows that a hybrid artificial intelligence system, which integrates clinical logic based on rules with machine learning algorithms, can be successfully used to classify liver diseases with routine biochemical markers gaining 91.2% accuracy. By combining the well-established clinical algorithms (R-factor, AST/ALT ratios) with the ensemble machine learning, performance is high, and there is clinical interpretability, alleviating key limitations of purely data-driven solutions.
The validated framework forms a realistic scalable approach to automated LFT perceptions that could be of significant importance in resource-constrained environments and contexts where specialized expertise is in short supply. The system has a sensitivity of 89.4% and a specificity of 93.1% to discriminate hepatitis, cirrhosis, and NAFLD with no blurred decisions that may be made and in which a physician can adopt.
Although such constraints as the limitation of the sample size, single-center design, and the focus have to be interpreted cautiously, the results provide strong evidence to the viability and clinical utility of the AI-assisted LFT interpretation. The fact that this framework has lightweight computational requirements (4.2MB model, 0.18s inference) and high accuracy to detect critical values (98.7 percent sensitivity) makes it an an immediate application that can be implemented as a decision-support tool.
A further direction in work on the topic should focus on future prospective multi-centre validation, time modeling to track the disease progression, a multimodal approach to basic clinical information, and randomized trials focusing on assessing the clinical value associated with patient outcomes. As AI is being developed further as an experimental research tool to a clinical applicant, hybrid models that combine human insight with machine learning provide a feasible solution to the development of reliable, explainable, and useful medical AI systems. Their effective implementation would significantly enhance early diagnosis and control of liver diseases, which would eventually decrease the hepatic morbidity and mortality burden in the world.

Conflicts of Interest: None

Funding: No special grant was provided in any funding agency in the public, commercial or not-profit sectors to do this research.

REFERENCES

1. Arya, G., Bagwari, A., Saini, H., Thakur, P., Rodriguez, C., & Lezama, P. (2023). Explainable AI for enhanced interpretation of liver cirrhosis biomarkers. IEEE Access, 11, 123729-123741.
2. Balsano, C., Burra, P., Duvoux, C., Alisi, A., Piscaglia, F., Gerussi, A., . . . Campanile, S. (2023). Artificial Intelligence and liver: Opportunities and barriers. Digestive and Liver Disease, 55(11), 1455-1461.
3. Brunt, E. M., Clouston, A. D., Goodman, Z., Guy, C., Kleiner, D. E., Lackner, C., . . . Leow, W. Q. (2022). Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. Journal of hepatology, 76(5), 1030-1041.
4. Calderaro, J., Žigutytė, L., Truhn, D., Jaffe, A., & Kather, J. N. (2024). Artificial intelligence in liver cancer—new tools for research and patient management. Nature Reviews Gastroenterology & Hepatology, 21(8), 585-599.
5. Chen, L.-D., Huang, Z.-R., Yang, H., Cheng, M.-Q., Hu, H.-T., Lu, X.-Z., . . . Lin, P. (2024). US-based sequential algorithm integrating an AI model for advanced liver fibrosis screening. Radiology, 311(1), e231461.

6. Chowdhury, A. B., & Mehta, K. J. (2023). Liver biopsy for assessment of chronic liver diseases: a synopsis. Clinical and experimental medicine, 23(2), 273-285.
7. Dai, Y., Gao, F., Chen, Y., Xu, S., Qiu, C., & Cai, X. (2025). Automated predictive framework using AI and deep learning approaches for early detection and classification of liver cancer. Frontiers in Oncology, 15, 1650800.
8. Dalvi, H., Narvekar, M., Dalvi, P., Vadgama, K., Kasmani, M., & Ranawat, K. (2024). Interpreting Accuracy: A Comprehensive Analysis of AI Models for Liver Disease Diagnosis. Paper presented at the 2024 Parul International Conference on Engineering and Technology (PICET).
9. Fan, Y., Lu, X., & Sun, G. (2023). IHCP: interpretable hepatitis C prediction system based on black-box machine learning models. BMC bioinformatics, 24(1), 333.
10. Fujii, I., Matsumoto, N., Ogawa, M., Konishi, A., Kaneko, M., Watanabe, Y., . . . Sugitani, M. (2024). Artificial Intelligence and Image Analysis-Assisted Diagnosis for Fibrosis Stage of Metabolic Dysfunction-Associated Steatotic Liver Disease Using Ultrasonography: A Pilot Study. Diagnostics, 14(22), 2585.
11. Iyer, J. S., Juyal, D., Le, Q., Shanis, Z., Pokkalla, H., Pouryahya, M., . . . Carrasco-Zevallos, O. (2024). AI-based automation of enrollment criteria and endpoint assessment in clinical trials in liver diseases. Nature Medicine, 30(10), 2914-2923.
12. Kalra, N., Verma, P., & Verma, S. (2024). Advancements in AI based healthcare techniques with FOCUS ON diagnostic techniques. Computers in Biology and Medicine, 179, 108917.
13. Khan, J. U. A., Shah, A. Z., & Akhar, Z. (2022). Editorial Hepatology Practice in Pakistan: Physician's Perspective. International Journal of Pathology, 139-140.
14. Naeem, M. U., Malik, K., Fareed, A., Kashif, R., Haider, A., Ghilzai, D., & Ramzan, H. S. (2024). Pakistan Society of Hepatology Guidelines on the Management of Hepatic Encephalopathy: Guidelines for Managing Hepatic Encephalopathy. Pakistan Journal of Health Sciences, 02-08.
15. Niu, X., Zhu, L., Xu, Y., Zhang, M., Hao, Y., Ma, L., . . . Xing, H. (2023). Global prevalence, incidence, and outcomes of alcohol related liver diseases: a systematic review and meta-analysis. BMC Public Health, 23(1), 859.
16. Popa, S. L., Ismaiel, A., Abenavoli, L., Padureanu, A. M., Dita, M. O., Bolchis, R., . . . Bosneag, A. (2023). Diagnosis of liver fibrosis using artificial intelligence: a systematic review. Medicina, 59(5), 992.
17. Tamber, S. S., Bansal, P., Sharma, S., Singh, R. B., & Sharma, R. (2023). Biomarkers of liver diseases. Molecular biology reports, 50(9), 7815-7823.
18. Trasca, D.-M., Dorin, P. I., Carmen, S., Varut, R.-M., Singer, C. E., Radivojevic, K., & Stoica, G. A. (2025). Artificial Intelligence in Biomedicine: A Systematic Review from Nanomedicine to Neurology and Hepatology. Pharmaceutics, 17(12), 1564.
19. Wu, X.-N., Xue, F., Zhang, N., Zhang, W., Hou, J.-J., Lv, Y., . . . Zhang, X.-F. (2024). Global burden of liver cirrhosis and other chronic liver diseases caused by specific etiologies from 1990 to 2019. BMC Public Health, 24(1), 363.
20. Zhang, P., Gao, C., Huang, Y., Chen, X., Pan, Z., Wang, L., . . . Qi, X. (2024). Artificial intelligence in liver imaging: methods and applications. Hepatology International, 18(2), 422-434.