



# Interpretable Machine Learning Data Modeling for Liver Disease Risk Profiling: Insights from the Indian Liver Patient Dataset

Kamruzzaman Mithu <sup>1\*</sup>, Shahanara Begum <sup>2</sup>, Md Nesar Uddin <sup>1</sup>, Mohammad Nurul Huda <sup>1</sup>

## Abstract

Liver disease remains a significant global health concern, yet identifying subtle or latent risk factors from clinical data continues to be challenging. In this study, we sought to explore these less obvious patterns using the Indian Liver Patient Dataset (ILPD), a widely used benchmark dataset comprising 583 records, including both liver and non-liver cases. Rather than focusing solely on predictive accuracy, we aimed to balance performance with interpretability an aspect that is, perhaps, equally critical in clinical contexts. The dataset was carefully preprocessed, including label encoding of categorical variables and normalization of continuous features. Multiple supervised machine learning models were evaluated to determine the most suitable approach. Among them, Logistic Regression emerged as the most consistent performer, achieving a test accuracy of approximately 71%, while also providing probabilistic outputs conducive to clinical interpretation. To better understand the model's decision-making process, SHAP (SHapley Additive exPlanations) was employed for feature attribution. This analysis revealed that Total Proteins, Age, and Albumin were the most influential predictors of liver disease within the dataset. These findings align, to some extent, with established clinical indicators, lending credibility to the model's outputs. Overall, this study demonstrates that interpretable machine learning can offer meaningful insights into liver disease risk while

maintaining transparency. By translating predictions into individualized risk profiles, the approach supports more informed and human-centric healthcare decisions, aligning with emerging Industry 5.0 principles. **Keywords:** Liver disease prediction, Machine learning, SHAP interpretability, Risk factor analysis.

## Keywords

Liver disease prediction, Machine learning, SHAP interpretability, Risk factor analysis.

## 1. Introduction

Liver diseases continue to represent a substantial and persistent burden on global public health. Across both developed and developing regions, disorders affecting the liver contribute significantly to morbidity, long-term disability, and premature mortality. The liver itself is an extraordinary organ—indeed, the largest internal gland in the human body—and its physiological responsibilities are remarkably diverse. It regulates metabolic homeostasis, converts nutrients into usable biochemical forms, detoxifies harmful compounds, synthesizes essential proteins, and supports immune defense mechanisms against invading pathogens (Dritsas & Trigka, 2023). Because these processes are deeply integrated with almost every physiological system, even modest

**Significance |** This study bridges predictive modeling and clinical interpretability, enabling transparent, individualized liver disease risk assessment to support more informed, human-centered healthcare decisions.

### \*Correspondence.

Kamruzzaman Mithu, Department of Computer Science & Engineering United International University, Dhaka, Bangladesh. E-mail: kamruzzamansmithu@gmail.com, Editor Moushumi Afroza Mou, Ph.D., And accepted by the Editorial Board March 16, 2025 (received for review December 29, 2025)

## Author Affiliation:

<sup>1</sup> Department of Computer Science & Engineering United International University, Dhaka, Bangladesh.

<sup>2</sup> Department of Obstetrics and Gynaecology, Ibn Sina Medical College, Dhaka, Bangladesh.

**Please cite this article as:** Mithu, K., Begum, S., Uddin, M. N., Huda, M. N. (2026).

"Interpretable Machine Learning Data Modeling for Liver Disease Risk Profiling: Insights from the Indian Liver Patient Dataset", *Data Modeling*, 7(1), 1-12, 10695

© 2026 DATA MODELING, a publication of Eman Research, USA.  
Eman Research Publishing House, 42 Raritan Reach Rd, South Amboy, NJ 08879, USA  
This is an open access article under the CC BY-NC-ND license.  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)  
(<https://publishing.emanresearch.org>)

impairment of hepatic function can lead to complex systemic consequences. Viral infections, alcohol consumption, metabolic disorders, environmental toxins, and lifestyle-related risk factors all contribute to the progressive deterioration of liver health, often culminating in conditions such as cirrhosis, fatty liver disease, or hepatocellular carcinoma. Despite the medical importance of early detection, diagnosing liver disease at its initial stages remains difficult. In many cases, symptoms appear only after significant physiological damage has already occurred. Traditional diagnostic approaches—such as imaging techniques, liver biopsies, and biochemical assays—are undeniably valuable, yet they can be costly, time-consuming, or invasive. Moreover, these procedures often identify disease only after measurable structural damage has taken place. This challenge has prompted researchers to explore alternative strategies capable of detecting subtle patterns within clinical data that may indicate early disease risk. In this context, machine learning has gradually emerged as a promising analytical paradigm capable of transforming the landscape of medical diagnostics (Ding et al., 2022; Khan et al., 2022). Machine learning models possess the capacity to analyze complex, high-dimensional datasets and uncover relationships that may not be immediately visible through conventional statistical approaches. Clinical datasets often contain numerous biochemical indicators, demographic variables, and laboratory measurements whose interactions are nonlinear and multifaceted. Human interpretation alone may struggle to identify meaningful relationships among these variables. Machine learning algorithms, however, can evaluate large volumes of data simultaneously, learning from patterns embedded within them and generating predictive models capable of estimating disease probability with considerable accuracy. Over the past decade, such techniques have increasingly been applied to healthcare problems ranging from cancer detection to cardiovascular risk prediction, demonstrating encouraging results across multiple domains (Ding et al., 2022; Khan et al., 2022). Within the specific context of liver disease, several studies have explored the potential of machine learning algorithms for improving diagnostic performance and clinical decision-making. These approaches typically rely on patient clinical profiles—including laboratory markers such as bilirubin levels, liver enzyme concentrations, and demographic characteristics—to train predictive models capable of distinguishing between healthy individuals and those with

hepatic disorders. The capacity of machine learning systems to process these diverse inputs simultaneously makes them particularly well suited for this task. In fact, research has shown that certain machine learning algorithms can achieve impressive diagnostic performance when applied to liver disease prediction problems (Mostafa et al., 2021). These developments suggest that data-driven predictive frameworks may serve as valuable decision-support tools in clinical environments where early identification of risk factors is essential.

Nevertheless, as machine learning models become increasingly sophisticated, a new challenge has emerged—interpretability. Many advanced predictive models, particularly those based on ensemble methods or deep learning architectures, are often described as “black boxes.” While such models may achieve high predictive accuracy, their internal decision-making processes can be difficult to interpret. In medical contexts, this lack of transparency can raise serious concerns. Healthcare professionals must understand not only what a model predicts but also why it reaches a particular conclusion. Without such understanding, clinicians may hesitate to rely on algorithmic recommendations when making critical diagnostic or therapeutic decisions. This growing concern has led to a parallel research movement focused on explainable artificial intelligence (XAI). Explainability aims to provide insights into how machine learning models generate predictions, thereby improving transparency, trust, and accountability in automated decision systems. Among the various techniques developed to address this issue, SHAP (SHapley Additive exPlanations) has gained particular prominence. Derived from cooperative game theory, SHAP values quantify the contribution of each feature to an individual prediction, allowing researchers and clinicians to interpret model behavior in a consistent and mathematically grounded manner (Stenwig et al., 2022). By decomposing predictions into feature contributions, SHAP makes it possible to identify which variables exert the strongest influence on model outcomes. The importance of explainability in healthcare extends beyond technical transparency. Interpretable models can reveal clinically meaningful insights that might otherwise remain hidden within complex datasets. For example, identifying which biochemical markers contribute most strongly to liver disease prediction can enhance clinical understanding of disease progression and support more targeted diagnostic strategies. Explainable machine learning can therefore serve not only as a predictive tool

but also as a mechanism for generating new medical knowledge. From ethical and regulatory perspectives, explainability is equally critical. Healthcare systems must ensure that algorithmic decisions are fair, understandable, and aligned with established clinical principles (Amann et al., 2020). In this study, the identification of hidden risk factors associated with liver disease is investigated using the Indian Liver Patient Dataset (ILPD), a widely used clinical dataset available through the UCI Machine Learning Repository (Ramana & Venkateswarlu, 2012). The ILPD dataset contains multiple biochemical indicators and demographic variables associated with liver health, making it a suitable benchmark for evaluating predictive models. Yet, as with many medical datasets, the relationships between these variables and disease outcomes are complex. Some predictors may exert strong direct effects, whereas others may interact in subtle ways that are not immediately obvious through traditional analytical approaches. To address this challenge, the present research integrates machine learning modeling with explainable artificial intelligence techniques. The dataset undergoes rigorous preprocessing and exploratory analysis before being used to train predictive models capable of distinguishing between liver disease patients and healthy individuals. Subsequently, SHAP-based interpretability methods are applied to analyze model behavior and identify the most influential features contributing to prediction outcomes. Through this combined approach, the study seeks not only to achieve accurate classification performance but also to illuminate the underlying risk factors that may contribute to liver disease development. Understanding these determinants holds important implications for clinical practice. If certain biochemical indicators consistently emerge as dominant predictive features, they may serve as valuable diagnostic markers in routine medical screening. Furthermore, insights derived from interpretable machine learning models can support clinicians in making evidence-based decisions, improving both diagnostic accuracy and patient management strategies. The remainder of this paper is organized as follows. Section II reviews previous research related to machine learning applications in liver disease diagnosis and prediction. Section III presents the dataset description, preprocessing procedures, exploratory data analysis, model development, evaluation methods, and interpretability framework used in this study. Section IV discusses the experimental results and their implications. Finally, Section V concludes the paper and outlines

potential directions for future research in explainable machine learning for medical diagnostics.

## 2. Methodology

### 2.1 Study Design and Analytical Framework

This study was conceived as a structured, reproducible machine learning–based investigation aimed at identifying clinically meaningful risk factors associated with liver disease. Although computational approaches have become increasingly common in hepatology research, methodological transparency—and, perhaps more importantly, reproducibility—still varies across studies (Khan et al., 2022; Singh et al., 2022). With this in mind, we adopted a clearly delineated analytical pipeline to minimize ambiguity and ensure that each step could be replicated independently.

The workflow consisted of sequential stages: data preprocessing, exploratory data analysis (EDA), model development, performance evaluation, risk quantification, and interpretability assessment. Rather than committing to a single predictive model at the outset, multiple supervised learning algorithms were evaluated comparatively. This approach aligns with recommendations in clinical machine learning literature, where model benchmarking is often necessary to balance predictive performance with interpretability (Dritsas & Trigka, 2023; Mostafa et al., 2021).

A schematic overview of the complete analytical pipeline is provided in Figure 1.

### 2.2 Data Source and Dataset Characteristics

The dataset utilized in this study was the Indian Liver Patient Dataset (ILPD), obtained from the UCI Machine Learning Repository (Ramana & Venkateswarlu, 2012). This dataset has been widely employed as a benchmark in liver disease prediction studies, enabling comparability across machine learning approaches (Jin et al., 2014; Wu et al., 2019).

The ILPD dataset contains a total of 583 patient records, including 416 individuals diagnosed with liver disease and 167 non-liver (control) cases. The demographic distribution is somewhat imbalanced, comprising 441 male and 142 female participants.

Ten clinical attributes were included in the analysis: age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase (ALT/SGPT),

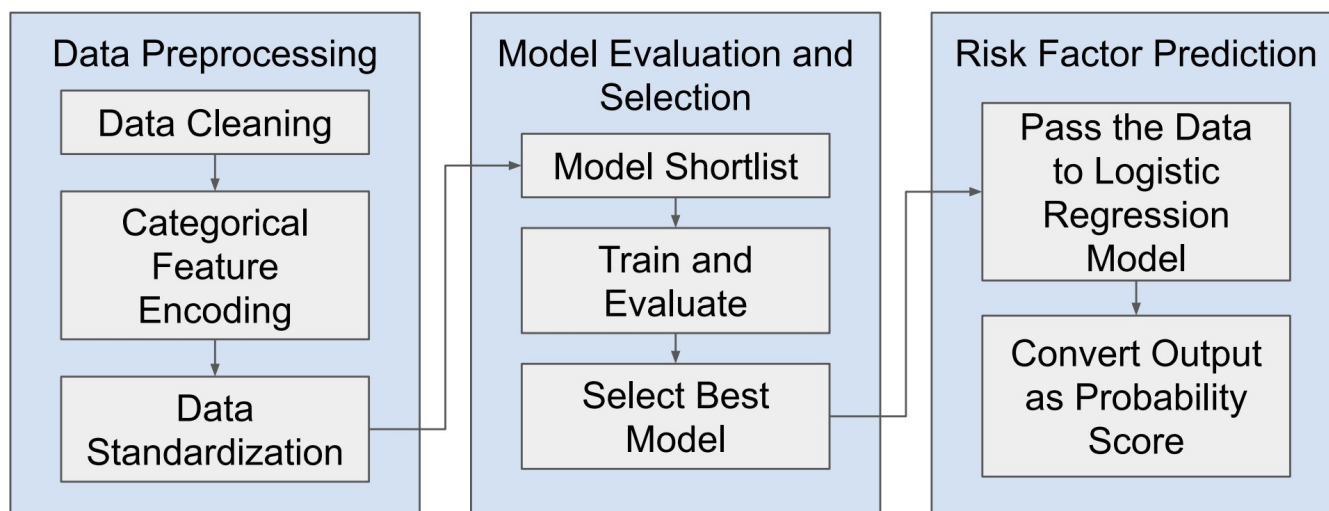


Figure 1. Overall methodological workflow for liver disease risk prediction. This figure illustrates the stepwise analytical pipeline adopted in the study, beginning with data preprocessing (including data cleaning, categorical feature encoding, and standardization), followed by model evaluation and selection (model shortlisting, training, and performance-based selection), and culminating in risk factor prediction using the selected Logistic Regression model. The final output is expressed as probability scores, which are subsequently interpreted as individualized risk estimates for liver disease.

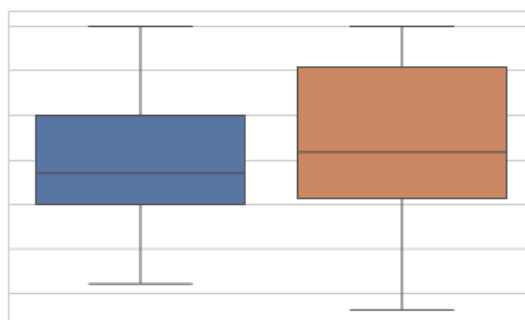


Figure 2. Distribution of predicted liver disease risk scores stratified by gender. This box plot compares the distribution of model-derived risk scores between male and female samples. The median risk, interquartile range, and overall spread appear higher among male individuals, suggesting a greater predicted susceptibility to liver disease in this group within the dataset. While the trend is noticeable, it should be interpreted with some caution, as it may reflect underlying sample imbalance or population-specific characteristics rather than a universally generalizable pattern.

Table 1. Clinical and biochemical attributes included in the Indian Liver Patient Dataset (ILPD).

Attribute	Type	Annotations
Age	Numeric	Age of the patient in years (Ramana & Venkateswarlu, 2012).
Gender	Nominal	Male, Female (Ramana & Venkateswarlu, 2012).
Total Bilirubin	Continuous	Sum of direct and indirect bilirubin levels. Used in the diagnosis of jaundice and evaluation of liver metabolic processes. Normal range: 0.3–1.9 mg/dL (Ding et al., 2022; Weng et al., 2023).

Direct Bilirubin	Continuous	Conjugated bilirubin combined with glucuronic acid, forming a water-soluble molecule measurable in blood tests. Normal range: 0.0–0.4 mg/dL (Ding et al., 2022).
Alkaline Phosphotase	Continuous	Enzyme related to bile duct function; elevated levels may indicate liver or bile duct disorders (Dritsas & Trigka, 2023; Ding et al., 2022).
SGPT (Alanine Aminotransferase – ALT)	Continuous	Enzyme primarily found in the liver; increased levels indicate liver cell injury or inflammation (Wu et al., 2019; Chen et al., 2022).
SGOT (Aspartate Transaminase – AST)	Continuous	Enzyme found in liver, heart, and muscles; elevated levels may indicate liver damage or other tissue injury (Wu et al., 2019; Su et al., 2023).
Total Proteins	Continuous	A biochemical test for measuring the total amount of protein in blood plasma or serum. Concentrations below the reference range usually reflect low albumin concentration, for instance in liver disease or acute infection. (normal: 6.0 8.0 g/dl) (Ding et al., 2022; Weng et al., 2023).
Albumin	Continuous	The most common family of globular proteins is serum albumin. Albumin is the main protein of human plasma. (normal: 3.5 5.0 g/dL) (Ding et al., 2022; Chen et al., 2022).
A/G Ratio	Continuous	Albumin/globulin ratio. It provides information about the amount of albumin you have compared with globulin, a comparison called the A/G ratio. It is useful when to suspects Liver damage, Spleen problems, Thymus malfunction, Kidney disease/damage, Protein digestion and absorption, Protein intake and Autoimmune conditions. (normal: 1.2 1.5) (Ding et al., 2022; Weng et al., 2023).
Class	Nominal	Outcome label in the dataset: patient with liver disease = 1; normal/non-liver control = 2 (Ramana & Venkateswarlu, 2012).

aspartate aminotransferase (AST/SGOT), total proteins, albumin, and the albumin-to-globulin (A/G) ratio. These variables represent routinely measured biochemical markers in hepatology and are consistently associated with liver dysfunction and metabolic alterations (Ding et al., 2022; Weng et al., 2023). A detailed description of these attributes, including their clinical relevance and reference ranges, is provided in Table 1.

The dataset is structured as a binary classification problem, distinguishing between liver disease and non-liver cases, and includes both continuous and categorical variables.

### 2.3 Data Preprocessing

Data preprocessing was conducted carefully, as even relatively minor transformations can influence downstream model behavior.

First, the categorical variable (gender) was encoded using label encoding, converting categorical values into numerical form while avoiding the introduction of artificial ordinal relationships. This ensured compatibility with machine learning algorithms without distorting feature meaning.

Second, all continuous variables were standardized using z-score normalization, such that each feature had a mean of 0 and a standard deviation of 1. This step was particularly important for distance-based and margin-based algorithms—such as Support Vector Machines and K-Nearest Neighbors—which are sensitive to feature scale (Fathi et al., 2020).

Notably, the dataset did not contain missing values; therefore, no imputation procedures were required. While this simplifies preprocessing, it should be acknowledged that real-world clinical datasets often necessitate more complex strategies for handling incomplete data (Amann et al., 2020).

Following these steps, the dataset was considered suitable for both exploratory analysis and predictive modeling.

### 2.4 Exploratory Data Analysis

Exploratory data analysis was performed to better understand variable relationships, identify potential sources of bias, and inform model selection.

#### 2.4.1 Correlation Analysis

Pearson correlation coefficients were computed to assess linear relationships between continuous variables. This analysis helped identify potential multicollinearity, which can affect model stability and interpretability—particularly in regression-based approaches (Mostafa et al., 2021).

#### 2.4.2 Distribution Assessment

To evaluate the distributional characteristics of each feature, histograms, boxplots, and kernel density plots were generated. These visualizations revealed patterns such as skewness, outliers, and deviations from normality.

Interestingly, several biochemical markers—notably bilirubin levels and liver enzymes—exhibited non-normal distributions. This observation is consistent with clinical findings, where such markers often display skewed behavior in diseased populations (Chen et al., 2022; Su et al., 2023).

Overall, EDA provided a necessary foundation for interpreting model outputs and ensuring that subsequent analyses were contextually grounded.

### 2.5 Model Development and Evaluation

To avoid bias toward a single modeling paradigm, multiple supervised machine learning algorithms were evaluated, including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Extra Trees Classifier, AdaBoost Classifier, and Naïve Bayes. The inclusion of diverse models reflects differences in complexity, interpretability, and learning mechanisms (Dritsas & Trigka, 2023; Stenwig et al., 2022).

#### 2.5.1 Train–Test Split

The dataset was partitioned into training (80%) and testing (20%) subsets using stratified sampling to preserve class proportions. This step is essential for ensuring unbiased performance evaluation.

#### 2.5.2 Evaluation Metrics

Model performance was assessed using multiple complementary metrics: accuracy, precision, recall, and F1-score. Given the class imbalance present in the dataset, reliance on a single metric could be misleading; therefore, a multi-metric evaluation strategy was adopted (Khan et al., 2022).

#### 2.5.3 Model Selection

Among the evaluated models, Logistic Regression

demonstrated the most consistent overall performance, achieving an accuracy of approximately 71%. Although comparatively simple, its interpretability and probabilistic output make it particularly suitable for clinical applications (Mostafa et al., 2021; Wu et al., 2019).

## 2.6 Risk Factor Quantification

Following model selection, individual-level risk scores were derived from the probabilistic outputs of the Logistic Regression model. Specifically, the predicted probability of belonging to the “liver disease” class was transformed into a percentage-based risk score:

$$RF_i = P(\text{Liver Disease} | i) \times 100$$

This formulation provides an intuitive interpretation of model output. However, it is important to acknowledge that such probabilities represent statistical estimates rather than direct clinical risk. Even so, probabilistic outputs are commonly used in medical prediction models and can offer practical, albeit approximate, insights into disease likelihood (Ding et al., 2022; Weng et al., 2023).

## 2.7 Model Interpretability Using SHAP

To enhance transparency and clinical relevance, model interpretability was assessed using SHAP (SHapley Additive exPlanations). SHAP values, grounded in cooperative game theory, quantify the contribution of each feature to individual predictions (Amann et al., 2020; Stenwig et al., 2022).

The interpretability workflow included:

- Computation of SHAP values for all test samples
- Identification of features with the highest positive and negative contributions
- Generation of summary and local explanation plots

Features with positive SHAP values were associated with increased likelihood of liver disease prediction, whereas negative values indicated a protective or inverse relationship.

This step proved particularly informative, as it allowed translation of model outputs into clinically interpretable insights while also serving as a validation mechanism for model behavior (Amann et al., 2020).

## 2.8 Reproducibility Considerations

To ensure reproducibility, all preprocessing steps, model

configurations, and evaluation procedures were applied consistently across experiments. Deterministic train–test splitting and standardized machine learning libraries were used to minimize variability.

Importantly, the dataset is publicly available (Ramana & Venkateswarlu, 2012), allowing independent verification and replication of results. Where applicable, parameter settings and processing steps were documented explicitly to facilitate reproducibility in future studies.

## 3. Result and Discussion

In the present study, we undertook a detailed analytical exploration of liver disease prediction using a range of supervised machine learning models, complemented by interpretability techniques. While machine learning applications in hepatology are not entirely new, the consistency and transparency of their implementation often remain uneven (Khan et al., 2022; Dritsas & Trigka, 2023). Against this backdrop, our findings offer a somewhat grounded—if still evolving—perspective on how predictive modeling can meaningfully contribute to clinical insight. Across the evaluated models, Logistic Regression emerged as the most reliable performer, achieving a test accuracy of approximately 71% (Table 2). Although more complex ensemble methods, such as AdaBoost and Extra Trees, demonstrated competitive metrics, their performance gains were not sufficiently consistent to outweigh the interpretability advantages of Logistic Regression. This observation aligns with prior work suggesting that simpler, well-calibrated models can often perform comparably in structured clinical datasets while retaining greater transparency (Mostafa et al., 2021; Wu et al., 2019). Beyond model performance, the interpretability analysis provided perhaps the most clinically relevant insights. Using SHAP (SHapley Additive exPlanations), we were able to quantify the contribution of individual features to model predictions. The results indicated that Total Proteins, Age, and Albumin exerted the strongest influence on predicted outcomes (Figure 6). These findings are, in a sense, reassuring—they are consistent with established clinical understanding, where alterations in protein metabolism and liver enzyme balance often signal hepatic dysfunction (Ding et al., 2022; Weng et al., 2023). The SHAP summary plot (Figure 6) and individual-level explanation (Figure 5) further illustrate how these variables contribute both globally and locally to prediction behavior. Interestingly, the correlation structure of the dataset also revealed meaningful interdependencies

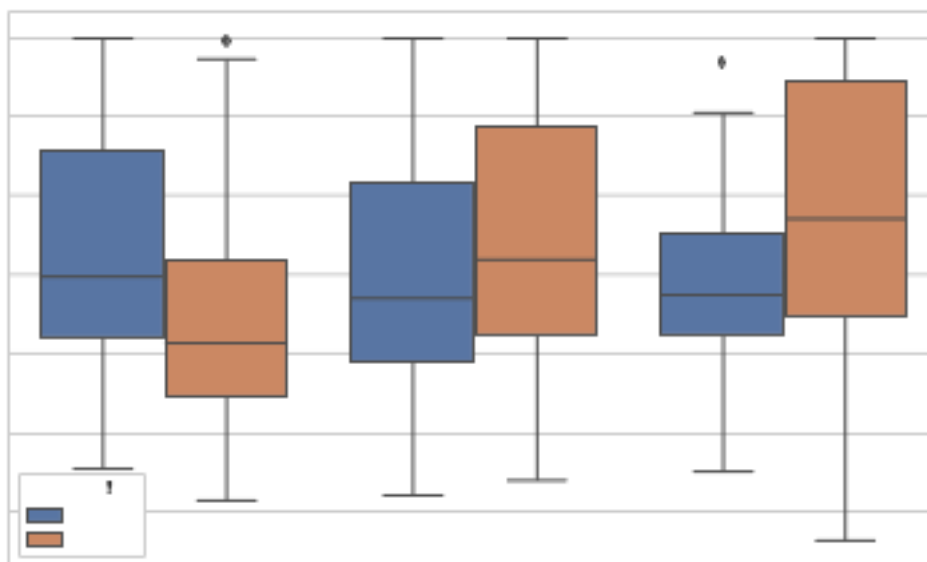


Figure 3. Distribution of predicted liver disease risk across age groups stratified by gender. This box plot illustrates the variation in model-derived risk scores across different age groups, with separate distributions shown for male and female samples. The figure highlights how risk patterns shift with age, revealing noticeable differences between genders within each age category. In general, male samples tend to exhibit higher median risk values across several age groups, while female risk appears comparatively lower and more variable. These trends, although suggestive, should be interpreted cautiously, as they may be influenced by sample size distribution and underlying demographic imbalances within the dataset.

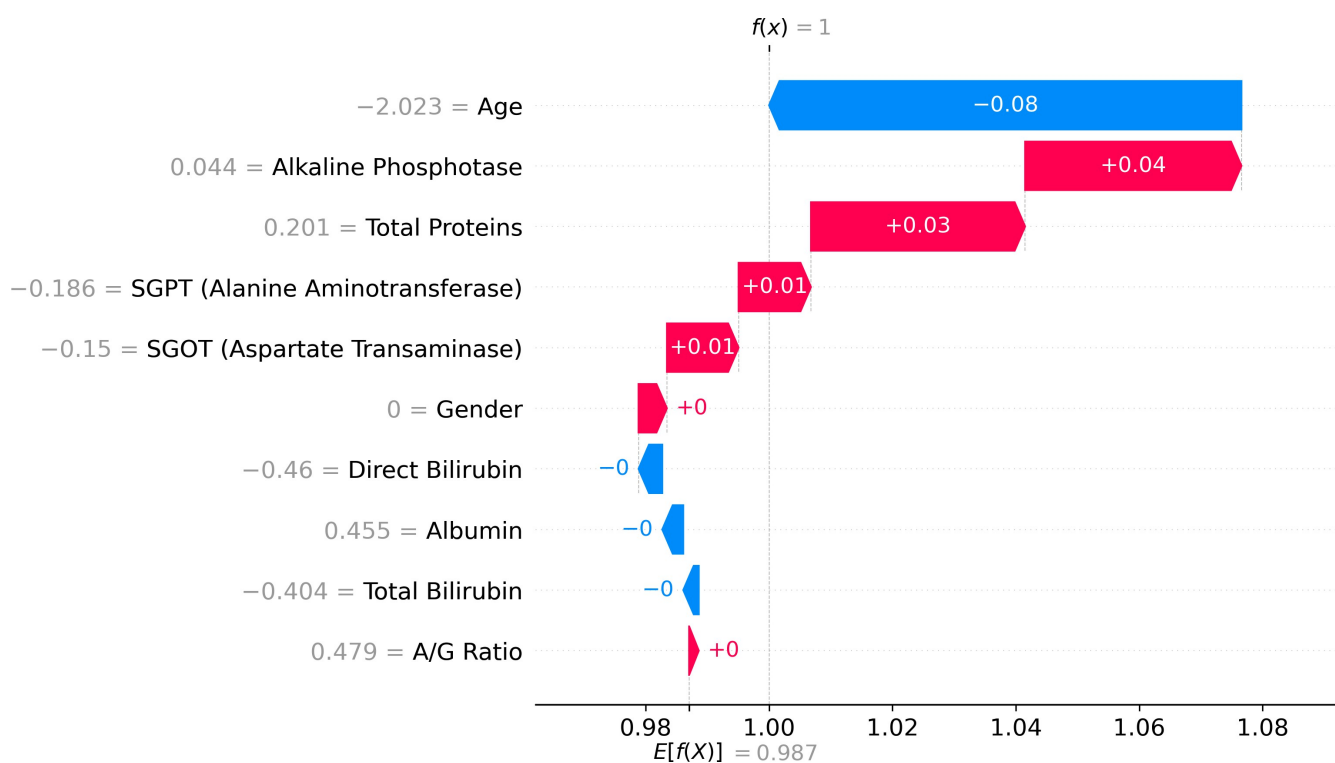


Figure 4. SHAP-based feature contribution plot for liver disease prediction. This figure illustrates the contribution of individual clinical features to the model's prediction for a representative case using SHAP values. Features shown in red increase the probability of liver disease, while those in blue decrease it. Among the variables, alkaline phosphatase, total proteins, and age contribute positively to risk, whereas albumin and bilirubin-related parameters exhibit negative or minimal influence. The horizontal axis represents the change in predicted probability relative to the baseline expectation ( $E[f(x)]$ ), highlighting how each feature shifts the final prediction.

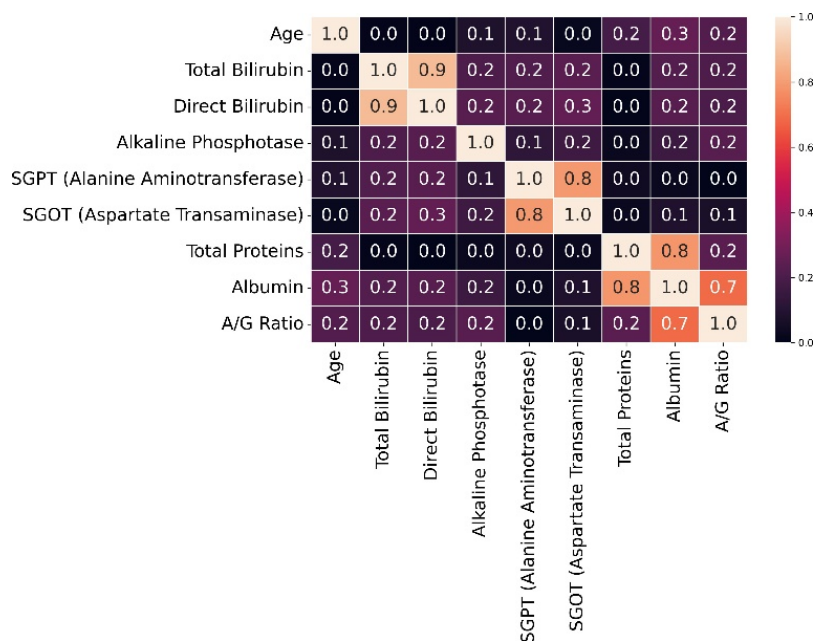


Figure 5. Correlation matrix of clinical variables in the liver patient dataset. This heatmap represents the pairwise Pearson correlation coefficients among key biochemical and demographic features. Warmer colors indicate stronger positive correlations, while lighter shades denote weaker associations. Notably, a strong positive correlation is observed between total bilirubin and direct bilirubin, as well as between SGPT and SGOT, reflecting related physiological pathways. Moderate associations are also evident between albumin and total proteins. Overall, the matrix highlights interdependencies among liver function biomarkers, which may influence model performance and feature interpretation.

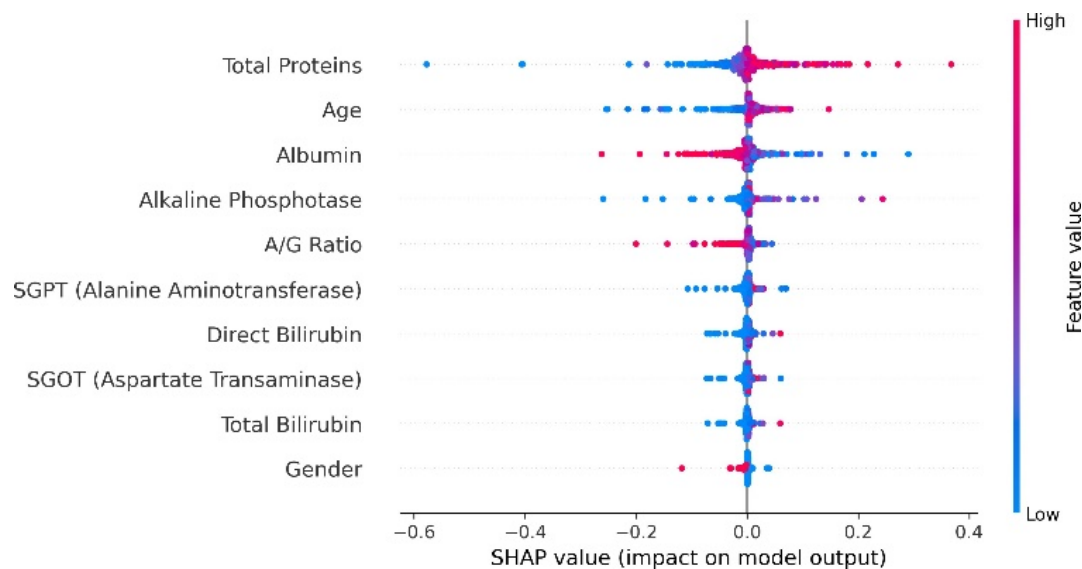


Figure 6. SHAP beeswarm plot illustrating global feature importance across the liver patient dataset. This plot summarizes the distribution of SHAP values for all features across the full dataset, highlighting their overall impact on model predictions. Each point represents an individual observation, with color indicating the feature value (blue = low, red = high). Features are ranked by their importance, with total proteins, age, and albumin showing the strongest influence on prediction outcomes. Positive SHAP values indicate increased likelihood of liver disease, whereas negative values suggest a reduced risk. The spread of points reflects both the magnitude and variability of each feature’s contribution across different samples.

Table 2. Comparative performance metrics of machine learning models for liver disease classification

Model	Accuracy	F1 Score	Precision	Recall	AUC
Logistic Regression	0.7133	0.2851	0.4688	0.2144	0.7263
Dummy Classifier	0.7133	0.0000	0.0000	0.0000	0.5000
Extra Trees Classifier	0.7109	0.3326	0.4951	0.2561	0.7232
AdaBoost Classifier	0.7107	0.4186	0.4792	0.3856	0.7245
Ridge Classifier	0.7084	0.0810	0.2000	0.0515	0.0000
Linear Discriminant Analysis	0.7035	0.1317	0.3000	0.0848	0.7020
Light Gradient Boosting Machine	0.7009	0.4020	0.4820	0.3591	0.6985
Random Forest Classifier	0.6888	0.2818	0.4330	0.2227	0.6992
SVM (Linear Kernel)	0.6865	0.1632	0.2006	0.1598	0.0000
Gradient Boosting Classifier	0.6813	0.3047	0.4298	0.2508	0.6995
K-Nearest Neighbors	0.6671	0.3137	0.3822	0.2765	0.6361
Decision Tree Classifier	0.6445	0.3850	0.3807	0.4091	0.5741
Quadratic Discriminant Analysis	0.5344	0.5265	0.3743	0.8977	0.7108
Naïve Bayes	0.5271	0.5341	0.3741	0.9402	0.7272

among biochemical markers (Figure 4). For instance, Total Bilirubin and Direct Bilirubin exhibited strong positive correlation, as did SGPT and SGOT, reflecting their shared physiological pathways. At the same time, Albumin showed a notable relationship with the albumin-to-globulin ratio, reinforcing its central role in liver function assessment. These relationships, visualized in the correlation matrix (Figure 4), underscore the interconnected nature of hepatic biomarkers and suggest that model predictions are influenced not by isolated variables, but by coordinated biological patterns. Another aspect worth noting—though it may require further validation—is the demographic variation observed in risk profiles. Male patients appeared to exhibit a higher overall predicted risk compared to female patients (Figure 2). Moreover, risk trajectories differed with age: while female risk showed a slight decline across age groups, male risk appeared to increase progressively (Figure 3). These patterns, while intriguing, should be interpreted cautiously, as they may reflect dataset-specific biases or underlying population characteristics rather than universal clinical trends. A key methodological contribution of this study lies in the transformation of model outputs into individualized risk scores. By converting predicted probabilities into percentage-based risk estimates, we attempted to bridge the gap between algorithmic output and clinical interpretability. Admittedly, this approach simplifies the complex nature of disease risk—probabilities derived from statistical models do not directly equate to real-world clinical outcomes. Still, such approximations are commonly used in predictive medicine and can provide a useful starting point for risk stratification (Ding et al., 2022). Equally important is the role of interpretability in fostering trust. The use of SHAP allowed us not only to validate model behavior but also to communicate results in a manner that is, at least to some extent, accessible to clinicians. This is particularly relevant in healthcare settings, where black-box models often face resistance due to lack of transparency (Amann et al., 2020; Stenwig et al., 2022). By contrast, interpretable models—supported by clear feature attribution—may be more readily integrated into decision-making workflows. Finally, it is worth situating this work within a broader technological context. The integration of interpretable machine learning in healthcare aligns with the emerging paradigm of Industry 5.0, which emphasizes human-centric, transparent, and sustainable technological solutions. In this sense, our approach—combining predictive modeling

with explainability—can be seen as a modest step toward more responsible and clinically aligned AI applications. Taken together, the findings of this study suggest that while predictive accuracy remains important, interpretability and clinical coherence are equally critical. Future work may benefit from expanding these analyses to larger, more diverse datasets and incorporating longitudinal or genetic data to refine risk prediction further.

#### 4. Limitations

While the present study provides useful insights into liver disease prediction, several limitations should be acknowledged—some of which, on reflection, are inherent to the data and methodological choices made. To begin with, the analysis relies exclusively on the Indian Liver Patient Dataset (ILPD), which, although widely used, represents a relatively small and geographically localized population. This raises some uncertainty regarding how well the findings might generalize to broader, more diverse clinical settings. The dataset also exhibits class imbalance and a noticeable gender skew, both of which may subtly influence model learning and prediction patterns. In addition, the study is based on cross-sectional data, meaning that temporal changes in disease progression are not captured. Liver disease typically develops over time, and without longitudinal information, it becomes difficult to infer how risk factors evolve or interact dynamically. The predictive performance of the selected model, while reasonable, remains moderate, suggesting that more complex or nonlinear relationships may not be fully captured. Moreover, although SHAP improves interpretability, it reflects associations rather than causal relationships. Finally, the absence of external validation limits the robustness of the findings, indicating that further multi-cohort and real-world clinical validation is needed.

#### 5. Conclusion

In summary, this study offers a structured yet cautiously interpreted application of machine learning to liver disease prediction, highlighting both predictive performance and interpretability. Among the evaluated models, Logistic Regression demonstrated the most consistent and clinically practical performance, achieving an accuracy of approximately 71%, while maintaining transparency—an aspect often undervalued in more complex models. The integration of SHAP-based interpretability provided additional depth, revealing that Total Proteins, Age, and Albumin play central roles in shaping predictive outcomes, findings that broadly align with established clinical

understanding. Perhaps more importantly, the proposed transformation of model probabilities into individualized risk scores presents a tentative but meaningful step toward patient-level risk stratification. While such estimates should not be interpreted as definitive clinical risk, they offer a useful approximation for further exploration. Overall, this work reinforces the value of interpretable artificial intelligence in healthcare, aligning with emerging Industry 5.0 principles that emphasize human-centric, transparent, and responsible technological integration.

## References

- Dritsas, E., & Trigka, M. (2023). Supervised machine learning models for liver disease risk prediction. *Computers*, 12(1), 19.
- Ding, H., Fawad, M., Xu, X., & Hu, B. (2022). A framework for identification and classification of liver diseases based on machine learning algorithms. *Frontiers in Oncology*, 12, 1048348.
- Khan, R. A., Luo, Y., & Wu, F.-X. (2022). Machine learning based liver disease diagnosis: A systematic review. *Neurocomputing*, 468, 492–509.
- Stenwig, E., Salvi, G., Rossi, P. S., & Skjærvold, N. K. (2022). Comparative analysis of explainable machine learning prediction models for hospital mortality. *BMC Medical Research Methodology*, 22(1), 1–14.
- Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 1–9.
- Singh, G., Agarwal, C., & Gupta, S. (2022). Detection of liver disease using machine learning techniques: A systematic survey. In *International Conference on Emerging Technologies in Computer Engineering* (pp. 39–51). Springer.
- Mostafa, F., Hasan, E., Williamson, M., & Khan, H. (2021). Statistical machine learning approaches to liver disease prediction. *Livers*, 1(4), 294–312.
- Fathi, M., Nemati, M., Mohammadi, S. M., & Abbasi-Kesbi, R. (2020). A machine learning approach based on SVM for classification of liver diseases. *Biomedical Engineering: Applications, Basis and Communications*, 32(3), 2050018.
- Jin, H., Kim, S., & Kim, J. (2014). Decision factors on effective liver patient data prediction. *International Journal of Bio-Science and Bio-Technology*, 6(4), 167–178.
- Weng, S., Hu, D., Chen, J., Yang, Y., & Peng, D. (2023). Prediction of fatty liver disease in a Chinese population using machine-learning algorithms. *Diagnostics*, 13(6), 1168.
- Chen, Y.-Y., Lin, C.-Y., Yen, H.-H., Su, P.-Y., Zeng, Y.-H., Huang, S.-P., & Liu, I.-L. (2022). Machine-learning algorithm for predicting fatty liver disease in a Taiwanese population. *Journal of Personalized Medicine*, 12(7), 1026.
- Su, P.-Y., Chen, Y.-Y., Lin, C.-Y., Su, W.-W., Huang, S.-P., & Yen, H.-H. (2023). Comparison of machine learning models and the fatty liver index in predicting lean fatty liver. *Diagnostics*, 13(8), 1407.
- Wu, C.-C., Yeh, W.-C., Hsu, W.-D., Islam, M. M., Nguyen, P. A. A., Poly, T. N., Wang, Y.-C., Yang, H.-C., & Li, Y.-C. J. (2019). Prediction of fatty liver disease using machine learning algorithms. *Computer Methods and Programs in Biomedicine*, 170, 23–29.
- Ramana, B., & Venkateswarlu, N. (2012). ILPD (Indian Liver Patient Dataset). UCI Machine Learning Repository. <https://doi.org/10.24432/C5D02C>