JOURNAL
Of Universal Applied Research

ISSN (Online): XXX
Vol. 01, Issue 01 (2026)
https://universalappliedresearch.com/index.php/JUAR/issue/archive

INTERPRETABLE PREDICTION OF GALLSTONE STATUS USING CLINICAL AND BIOCHEMICAL INDICATORS: A COMPARATIVE ANALYTICAL APPROACH

Minaketan Behera

Centre for Informal Sector and Labour Studies, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

Received:- 28/02/2026, Revised:- 11/04/2026, Accepted:- 19/04/2026, Published:- 28/04/2026

Abstract

Numerous metabolic, inflammatory, and biochemical factors influence gallstone disease, although much predictive research focuses on performance with models, but does not sufficiently consider interpretability and variable redundancy. This study examined the association between clinical and biochemical indicators and gallstone status and compared interpretable and non-linear predictive models within a structured analytical framework. A quantitative cross-sectional design based on secondary data analysis was employed. The dataset included 319 observations and 39 variables. Descriptive statistics, independent sample t-tests, correlation analysis, and multicollinearity assessment using the Variance Inflation Factor were conducted. Significant and non-collinear variables were retained for model development. Logistic Regression and Random Forest models were trained using an 80:20 train-test split and evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Eighteen variables showed significant differences between groups. C-reactive protein and Vitamin D emerged as the most influential predictors across both models. Logistic Regression achieved an accuracy of 0.781 and ROC-AUC of 0.829, while Random Forest achieved the same accuracy with a higher ROC-AUC of 0.860. These findings indicate that inflammatory, metabolic, and biochemical indicators jointly contribute to gallstone classification. A compact set of routinely measurable variables can reasonably differentiate gallstone status, and the integration of statistical analysis with machine learning supports both predictive performance and interpretability.

Keywords: gallstone disease, clinical biomarkers, interpretable machine learning, predictive modeling, C-reactive protein.

1. Introduction

Gallstone disease is one of the most widespread gastrointestinal diseases in the world, with significant impacts on the health of a person and the health care system. As global estimates show, the burden of gallstones has grown significantly in the 21st century, with the impact of lifestyle changes, eating habits, and aging of the population (Wang et al., 2024). The condition is linked with various complications such as biliary colic, cholecystitis and pancreatitis that may need medical or surgical intervention. Although the treatment has been improved, the question of the management of gallstone disease is still a critical clinical issue because of the recurrence and variability of the symptoms as well as comorbidity (Gutt et al., 2020). Simultaneously, the forecasts of the global burden of gallbladder and biliary diseases indicate that the risk is likely to increase, which is why it is necessary to better comprehend the risk factors and the methods of their early detection (Dai et al., 2025). The formation of gallstones is well known to be a multifactorial process that is influenced by metabolic, genetic and environmental factors. Genetic predisposition has been reported to be a contributor to the differences in susceptibility, and a number of loci are associated with cholesterol metabolism and bile composition (Costa et al., 2024). Simultaneously, clinical and biochemical clues, including lipid profiles, inflammatory indicators, and metabolic conditions, are significant factors in the development of the disease. This effect of lifestyle and metabolic conditions is specifically notable, with obesity and metabolic abnormalities having been greatly linked to gallstone development, particularly among younger patients (Su et al., 2019). Additionally, more general determinants like socioeconomic status and access to healthcare have also been identified to have an impact on disease management and prognosis, highlighting the complexity of gallstone disease beyond biological processes (Dupont et al., 2022).

The literature available has covered an extensive list of possible causes of gallstone formation, but specifically, biochemical and physiological pathways have been given focus. Literature reviews point to the importance of cholesterol supersaturation, bile stasis, and inflammation as the key processes in the development of gallstones (Sun et al., 2022). It is also suggested that biochemical studies can be used to understand the etiology and progression of diseases due to the changes in bile composition and metabolic parameters (Cevhertaş et al., 2025). The results highlight the significance of assessing gallstone risk with a combination of various clinical and biochemical parameters. Simultaneously, the development of data-driven approaches has created new opportunities to examine the intricate data in the healthcare field, making it possible to identify the patterns that could be unseen by traditional statistical methods (Freitas, 2023). The use of these methods in biomedical studies has been on the rise, and it has become more important to consider integrating the statistical rigor with the computational methods to increase predictive power (Dey, 2023). Within the framework of gallstone disease, machine learning techniques were recently investigated to predict the model and risk analysis. The research has shown that automated detection systems can be used to categorize the presence of gallstones based on clinical data, which is more efficient and scalable (Kanwal et al., 2025). Correspondingly, hybrid modeling methods have been suggested to combine probabilistic and machine learning methods for the estimation of risks in a more nuanced fashion (Chakraborty & Mukherjee, 2025). Although these developments suggest encouraging trends, much of the current research focuses on predictive performance without sufficiently discussing the issues of interpretability or the statistical connections between variables. Moreover, the use of big or highly-specialized data sets may restrict the generalizability of such models in clinical practice.

There is a significant gap in the combination of conventional statistical analysis and interpretable machine learning techniques, especially in cases of structured clinical data of moderate size. The fact that there is often multicollinearity between clinical variables is not adequately addressed in most cases, which may compromise the stability and interpretation of the model. Also, minimal attention is paid to the matching of the results of inferential and predictive modeling, which is necessary to make sure that the identified predictors are not only statistically significant but also meaningful. To fill these gaps, a solution that is both statistically inferential, model interpretable and predictive and at the same time applicable to real-world clinical data is needed.

The purpose of the study is to investigate the correlation between clinical and biochemical indicators and the presence of gallstones based on a structured model of analysis. It also aims at making and comparing predictive models that are both linear and non-linear in nature, and that are interpretable. Incorporating inferential analysis, multicollinearity, and machine learning, the study is a step toward a more detailed comprehension of the predictors of gallstones and their contribution to classification.

2. Methodology

2.1 Research Design

The research design in this study is a quantitative, cross-sectional research design using secondary data analysis. The method will combine statistical inference with supervised machine learning to analyze the relationships between clinical variables and gallstone status, and create predictive models. The design focuses on the interpretability and predictive performance and corresponds to the purpose of applied research.

2.2 Data Source

The data in this study were taken from an open-source clinical dataset that Esen et al. (2024) compiled. It includes patient-level data on the presence of gallstones and a variety of demographic, biochemical and body composition variables. The data set was created with the primary aim of assisting in clinical analysis of factors related to gallstone disease and contains structured measures appropriate to statistical and predictive modeling.

2.3 Data Description

The data set is 319 observations, and 39 variables. The dependent variable is categorical since it has the presence or absence of gallstones. Demographic (e.g., age, gender), clinical (e.g., hyperlipidemia, diabetes), anthropometric (e.g., BMI, body composition measurements), and biochemical (e.g., CRP, lipid profile, liver enzymes) variables are the independent variables. There are no non-numeric variables, and all the variables can be directly processed analytically.

2.4 Data Preprocessing

Validation and transformation of data were data preprocessing procedures to make the data analytically ready. The data was checked against missing values and duplicates, but no cases were found. All variables were then translated into the right numeric format, and the names of columns were made uniform. Since there were no missing data, imputation was not necessary. An 80:20 stratified split was then used to divide the dataset into a training and testing part to maintain the classes.

2.5 Statistical Analysis

In a bid to summarize the distribution of variables, descriptive statistics were calculated. Independent sample t-tests were used in inferential analysis to determine any significant differences between gallstone and non-gallstone groups. The variables having a p-value below 0.05 were considered significant and would be included in further analysis. Correlation was analyzed to identify correlations among the variables and the target outcome. Variance Inflation Factor (VIF) was used to determine multicollinearity, and high multicollinearity variables were discarded to maintain the model stability.

2.6 Feature Selection

The process of selecting features was twofold. To begin with, the shortlisting of variables that were statistically significant in the inferential analysis was done. Second, the variables with high multicollinearity (VIF > 10) were dropped. The last feature set was made of both statistically relevant and independent variables, which made it robust in future modeling.

2.7 Model Development

Two monitored classification models were created to predict performance and interpretability. Logistic Regression was used because it has the ability to give interpretable coefficients, which can be used to estimate the direction and magnitude of the relationships between predictors and the presence or absence of gallstones. Parallel to it, a Random Forest model was used to consider possible non-linear relationships and interactions between variables, thus increasing predictive strength. In the case of the Logistic Regression model, standardisation was used as a form of feature scaling to be able to compare variables. Random Forest, on the other hand, was applied without scaling because tree-based algorithms are not sensitive to variation in feature magnitudes in the first place.

2.8 Model Evaluation

A hold-out test set was used to test the performance of the model. It was evaluated using several complementary measures, such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). Overall correctness of classification was evaluated through accuracy, and precision and recall gave information about the performance at individual classes. F1-score was deemed to strike a balance between precision and recall especially in assessing consistency of classification. The model was tested with respect to the capability of discriminating between classes at various threshold settings using ROC-AUC. Collectively, these measures provide a holistic evaluation of predictive accuracy and classification performance.

2.9 Feature Importance Analysis

Model-specific methods were used to determine the importance of features. The direction and strength of the associations were determined using the Logistic Regression coefficients, whereas the relative contribution of each predictor was determined using the Random Forest feature importance scores. This two-fold method aids interpretability and predictive insight.

3. Results

3.1 Descriptive and Inferential Analysis

The research examined 319 observations, including an equal number of non-gallstone (n = 161) and gallstone (n = 158). Independent sample t-tests were used to perform inferential analysis, where several variables that are statistically significant across the two groups were found ($p < 0.05$). These variables cut across biochemical, anthropometric and clinical realms, suggesting that the gallstone disease is multifactorial.

The greatest differences were witnessed in Vitamin D levels, which were significantly lower in the gallstone group, and C-Reactive Protein (CRP), which was significantly higher in the affected population. Also, the body composition indicators, including lean mass and total body fat ratio, exhibited contrary trends across the groups, indicating a metabolism imbalance. A number of biochemical indicators, such as haemoglobin, AST, and creatinine, had also statistically significant changes. Table 1 gives the variables that have a significant difference among the groups in terms of mean value, difference and p-value.

Table 1. Significant Differences in Clinical and Biochemical Variables

Variable	Mean (No Gallstone)	Mean (Gallstone)	Difference	p-value
Vitamin D	24.905	17.831	-7.074	6.88×10^{-11}
C-Reactive Protein (CRP)	0.462	3.272	2.810	4.91×10^{-7}
Lean Mass (LM) (%)	73.522	69.718	-3.804	4.70×10^{-5}
Total Body Fat Ratio (TBFR) (%)	26.392	30.194	3.802	4.80×10^{-5}
Bone Mass (BM)	2.912	2.692	-0.220	9.66×10^{-5}
Hemoglobin (HGB)	14.764	14.066	-0.698	4.05×10^{-4}
Extracellular Water (ECW)	17.629	16.503	-1.127	0.00139
Total Fat Content (TFC)	21.871	25.135	3.265	0.00229
Extracellular Fluid/Total Body Water (ECF/TBW)	42.757	41.657	-1.100	0.00241
Hyperlipidemia	0.000	0.051	0.051	0.00435
High-Density Lipoprotein (HDL)	46.696	52.308	5.613	0.00480
Gender	0.416	0.570	0.153	0.00602
Visceral Fat Area (VFA)	11.441	12.916	1.475	0.01203
Aspartate Aminotransferase (AST)	23.913	19.415	-4.498	0.01583
Creatinine	0.824	0.777	-0.047	0.01806
Body Mass Index (BMI)	28.239	29.528	1.289	0.03002
Total Body Water (TBW)	41.460	39.699	-1.762	0.04718
Alkaline Phosphatase (ALP)	70.484	75.791	5.306	0.04972

Figure 1 below shows the top significant variables in order to visually compare their magnitude and direction of difference.

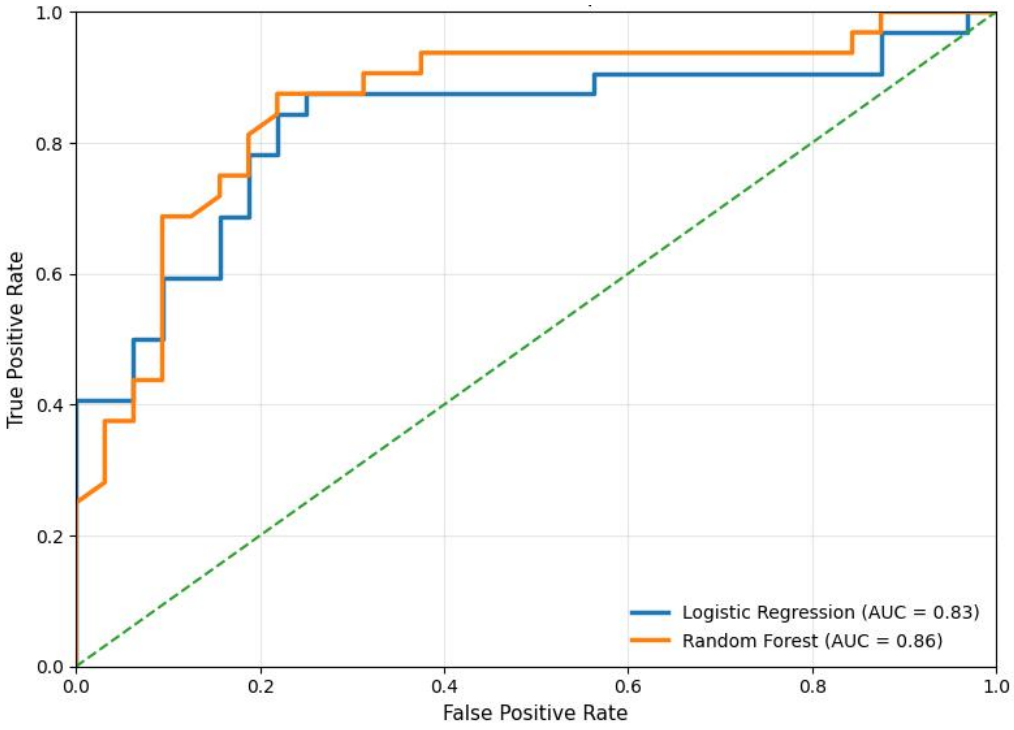

Figure 1. Top Significant Variables (Mean Comparison Between Groups)

3.2 Correlation and Model Performance

The correlation analysis showed that inflammatory and metabolic variables had the highest correlation with gallstone status. Specifically, CRP had the strongest positive correlation, whereas Vitamin D exhibited a significant negative relationship, which suggests that it has the potential to protect. Although these associations prevailed, intercorrelations between a number of body composition variables were high, which necessitated a reduction in features before the development of the model. After selecting the features, nine variables were selected to predict the model. The predictive performance of two classification models, Logistic regression and random forest, was compared. Table 2 shows the performance measures of the two models.

Table 2. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.781	0.800	0.750	0.774	0.829
Random Forest	0.781	0.821	0.719	0.767	0.860

The accuracy of both models was the same (78.1%), which means that they have the same classification ability. Nonetheless, Random Forest showed the best discriminative ability, with a higher ROC-AUC (0.860), indicating that it is more sensitive to the separation between classes. Figure 2 shows the comparative ROC curves of both models.

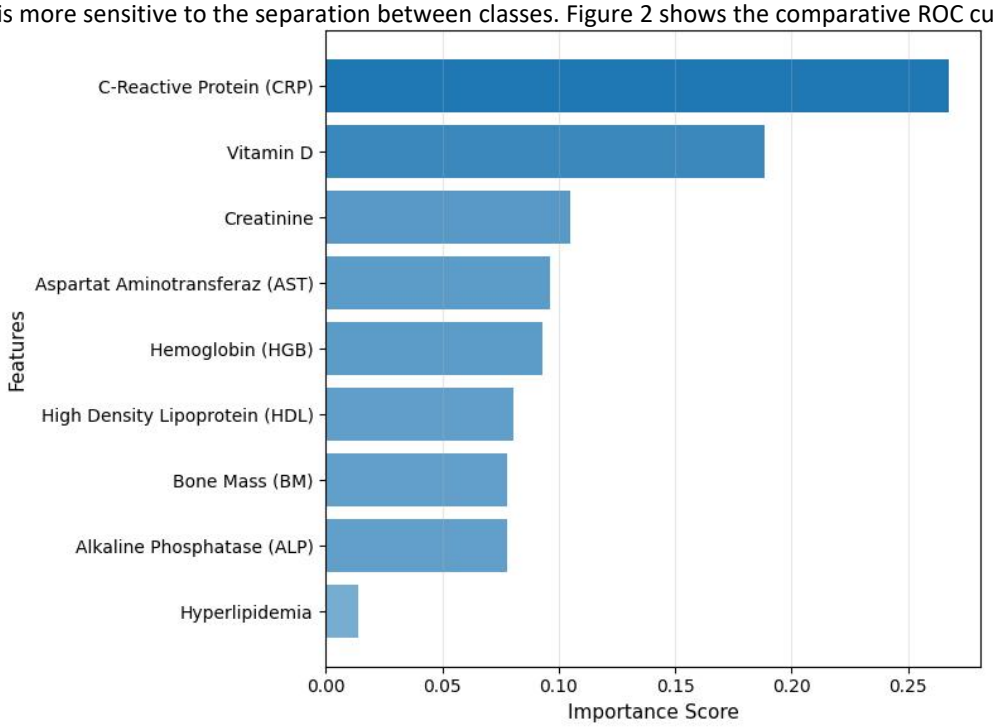

Figure 2. ROC Curve Comparison of Logistic Regression and Random Forest Models

3.3 Feature Importance and Predictive Factors

In order to further analyze the contribution of individual predictors, the importance of features was analyzed with the help of both Logistic Regression coefficients and Random Forest importance scores. Table 3 gives a comparative summary of these results.

Table 3. Comparative Feature Importance Across Models

Feature	Logistic Coefficient	Logistic Absolute Importance	Random Forest Importance
C-Reactive Protein (CRP)	1.750	1.750	0.268
Vitamin D	-0.830	0.830	0.188
Creatinine	0.070	0.070	0.105
Aspartate Aminotransferase (AST)	-0.497	0.497	0.096
Hemoglobin (HGB)	0.103	0.103	0.093
High-Density Lipoprotein (HDL)	0.231	0.231	0.081
Bone Mass (BM)	-0.319	0.319	0.078
Alkaline Phosphatase (ALP)	0.080	0.080	0.078
Hyperlipidemia	0.788	0.788	0.014

In both models, CRP was found to be the most significant predictor, and then Vitamin D. The fact that these variables are consistent across the methodologies of the different models strengthens their significance in the differentiation of gallstones. Figure 3 further visualizes the relative importance of the features based on the Random Forest model.

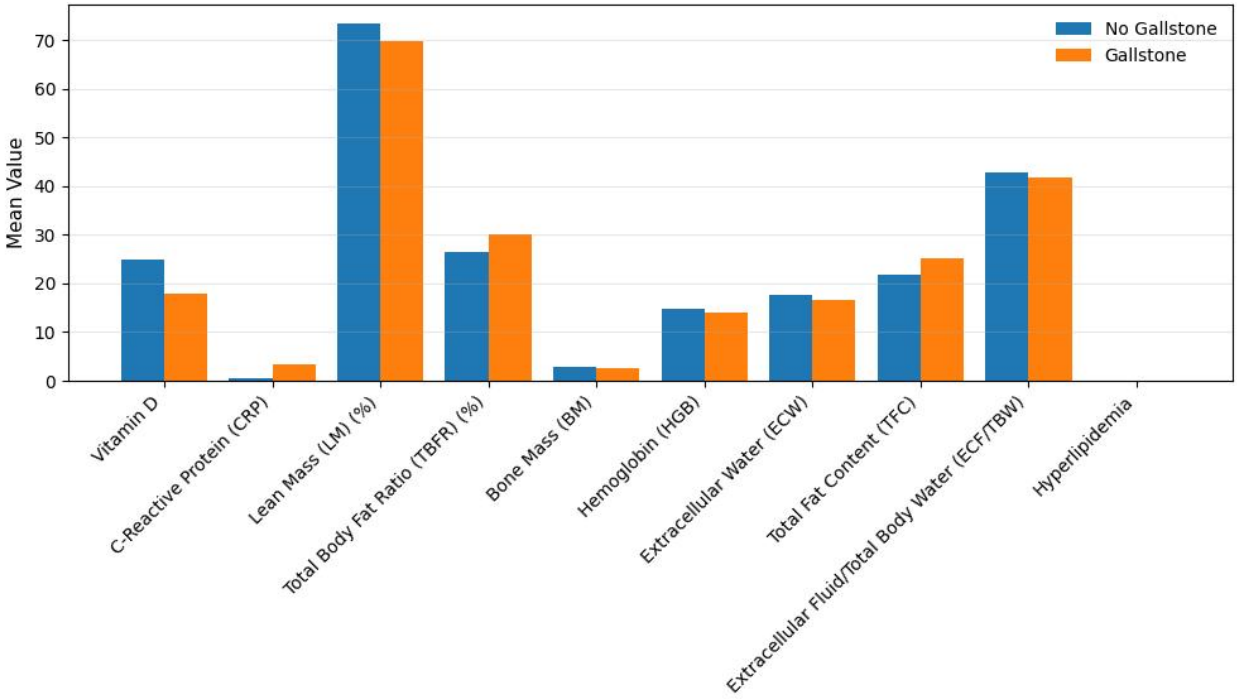

Figure 3. Top Feature Importance Based on Random Forest Model

The findings indicate that a set of inflammatory markers, metabolic indicators, and biochemical parameters can be successfully used in the distinction of gallstone and non-gallstone individuals, with the same results in both statistical and machine learning methods.

4. Discussion

The inflammatory, metabolic, and biochemical factors and not one predominant factor, were identified to relate to gallstone status in this study. The most significant predictor in both models was C-reactive protein (CRP), which suggests that inflammation in the body is closely associated with the presence of gallstones. This is indicated by its consistent significance following the elimination of features and multicollinearity, which implies that the inflammatory activity is a prominent distinguishing variable, rather than an incidental relationship. Vitamin D had the highest negative relationship with gallstone status, and was the second most important. The fact that the gallstone group has lower levels and a negative coefficient suggests that it may have a protective or regulatory effect. Combined, the dominance of CRP and Vitamin D points to a contrast between inflammatory load and metabolic equilibrium in the determination of the presence of gallstones. Other predictors such as hyperlipidemia, HDL, AST, alkaline phosphatase, creatinine, hemoglobin, and bone mass also indicate the presence of lipid metabolism, liver functions, and other extensive processes. The variables of body composition, including lean mass, total body fat ratio, BMI, and visceral fat area, were also found to differ based on descriptive analysis. Yet these variables were very collinear and reflected the same physiological dimensions. Their elimination enhanced model stability as well as predictors that were meaningful. Both models were equally accurate, yet Random Forest demonstrated greater ROC-AUC, which implies that it has better discrimination between classes and that there are possibly non-linear associations. Whereas Logistic Regression, though a little less discriminative, offered interpretable coefficients, so that the direction of associations could be explained. In general, the results indicate that a small number of indicators that can be measured routinely can be used to predict the status of gallstones fairly well, and a combination of the statistical and machine learning methods allows for a better understanding of the issue.

The key role of CRP in the current findings aligns with the previous literature that suggests a mediating effect of inflammatory processes between obesity and gallstone disease, which supports the idea that inflammation is a mechanism, and not merely a correlate (Zhen et al., 2025). The correlation between CRP and the presence of gallstones is also consistent with cohort-based evidence that high-sensitivity CRP is associated with a higher risk of gallstones (Liu et al., 2020). The adverse correlation between Vitamin D is consistent with recent findings that indicated that a lower Vitamin D intake is associated with gallstones in adults, which indicates that Vitamin D-related pathways could be one of the determinants of gallstones in adults (Bin & Zhang, 2025). The metabolic picture observed here is also in agreement with machine and deep learning studies that demonstrate that metabolites and clinical variables interact with each other to define gallstone disease, not as individual factors (Salem et al., 2023).

The average benefit of Random Forest over Logistic Regression is in line with comparative modeling research studies on biliary disorders, with findings indicating that machine learning tools can be used to offer better discrimination where the relationships among predictors are complex or non-linear (Cao et al., 2025). These tendencies have been noticed with predictive studies of gallstone disease and other complications, which computational models have shown

to be valuable in classification and risk stratification in diverse clinical contexts (Wenqian et al., 2024; Zhang et al., 2023). Meanwhile, the focus on interpretability can also be considered significant since the success of predictions is not enough to translate clinical results. The studies on the topic of gallbladder malignancy among patients with gallstones have demonstrated that risk modeling is best applicable in circumstances where predictor structure and clinical meaning stay clear (Zhu et al., 2023). Similar studies on acute gallstone pancreatitis, as well as predicting the composition of gallstones, also exemplify the increasing role of more complex models, yet these paradigms typically involve imaging or large-scale technical infrastructures that are not aligned with the organized clinical framework here (Ma et al., 2024; Yao et al., 2019).

These are both clinically and methodological relevant findings. The clinical findings indicate that routinely available measurements like CRP, Vitamin D, AST, ALP, creatinine, hemoglobin, and lipid-related measurements can help in identifying the patterns related to the presence of gallstones. This is not to suggest independent diagnostic substitution, but it endorses the worth of formatted clinical information in risk-based evaluation. Methodologically, the analysis shows the significance of using inferential screening together with multicollinearity evaluation prior to the development of the model. Reduction of predictors is necessary in datasets with numerous variables that characterize correlated physiological realms to obtain stable and interpretable models. The paper also reveals that interpretable techniques and machine learning can be used as complementary tools and not competitive approaches.

There are a number of limitations that ought to be taken into account when interpreting these findings. The sample was small, which can limit the extrapolation and lower sensitivity to less strong associations. The cross-sectional design also restricts the possibility of causal interpretation, as the relationships observed are based on correlation, not time-course. Moreover, external validation is not available, which implies that the performance of the models must be taken as tentative, not conclusive. Clinically rich data were still confined to those structured variables available, and perhaps interesting dietary, behavioral, imaging, or longitudinal variables were omitted.

Future studies need to assess these predictors on larger and more diverse populations, ideally by using external validation cohorts. Longitudinal designs would assist in establishing whether inflammatory and metabolic factors are predictive of future gallstone formation, as opposed to merely being comorbidities. Additional research might also investigate calibrated risk modeling, subgroup analysis, and clinically interpretable explainability methods to further streamline practice. These measures would enhance the trans-theoretical worth of predictive analytics in gallstone-related studies.

5. Conclusion

An organized statistical analysis and machine learning under supervision demonstrated that the status of the gallstones can be discriminated with a small number of clinical and biochemical indicators that are routinely measured. Of the predictors that were retained, the most informative ones were inflammatory and metabolic markers, and C-reactive protein and Vitamin D were the most consistently predictive in both interpretable and non-linear models. The similar performance of Logistic Regression and Random Forest, along with the superior discriminative power of the latter, implies that both linear and more complicated interactions between the predictors are useful in classification. The paper also reveals the usefulness of combining inferential tests, multicollinearity analysis, and feature modeling into a single analysis model. This method enhanced the stability of the models and retained clinically significant predictors. Despite the need to view the findings through the prism of the small sample size, cross-sectional study, and absence of external validation, they suggest that small-scale, easy-to-understand predictive models can deliver valuable information on the clinical presentation of gallstones. This method should be applied to larger and more diverse populations in the future to establish generalizability and enhance translational applicability.

References

1. Bin, C., & Zhang, C. (2025). The association between vitamin D consumption and gallstones in US adults: A cross-sectional study from the national health and nutrition examination survey. Journal of the Formosan Medical Association, 124(3), 212-217.
2. Cao, Y., Hu, X., Guo, J., & Fang, T. (2025). Machine learning-based prediction of recurrent extrahepatic bile duct stones after common bile duct exploration: a comparative study of models and SHAP-driven interpretability analysis. Frontiers in Medicine, 12, 1691519.
3. Cevhertaş, M., Çeltik, Ü., Keçeci, T., Çelik, A., & Ergün, M. O. (2025). Gallstone Biochemical Analysis: A Key to Unlocking Disease Etiology?. The Journal of Pediatric Research.
4. Chakraborty, C., & Mukherjee, N. (2025). Bayesian Hybrid Machine Learning of Gallstone Risk. arXiv preprint arXiv:2506.14561.
5. Costa, C. J., Nguyen, M. T. T., Vaziri, H., & Wu, G. Y. (2024). Genetics of gallstone disease and its clinical significance: a narrative review. Journal of clinical and translational hepatology, 12(3), 316.
6. Dai, F., Cai, Y., Yang, S., Zhang, J., & Dai, Y. (2025). Global burden of gallbladder and biliary diseases (1990–2021) with healthcare workforce analysis and projections to 2035. BMC gastroenterology, 25(1), 249.
7. Dey, N. (Ed.). (2023). Data-driven Approach for Bio-medical and Healthcare. Springer.

8. Dupont, B., Dejardin, O., Bouvier, V., Piquet, M. A., & Alves, A. (2022). Systematic review: impact of social determinants of health on the management and prognosis of gallstone disease. Health Equity, 6(1), heq-2022.
9. Esen, I., Arslan, H., Aktürk, S., Gülşen, M., Kültekin, N., & Özdemir, O. (2024). Gallstone [Dataset]. UCI Machine Learning Repository. https://doi.org/10.1097/md.0000000000037258.
10. Freitas, A. T. (2023). Data-driven approaches in healthcare: Challenges and emerging trends. Multidisciplinary perspectives on artificial intelligence and the law, 65-80.
11. Gutt, C., Schläfer, S., & Lammert, F. (2020). The treatment of gallstone disease. Deutsches Ärzteblatt International, 117(9), 148.
12. Kanwal, K., Batool, I., Naz, I., Shabbir, A., Fatima, M., & Gill, H. B. (2025). Automated Detection of Gallstone Disease Using Machine Learning Techniques: A Data-Driven Approach. STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 7(2), 157-176.
13. Liu, T., Siyin, S. T., Yao, N., Duan, N., Xu, G., Li, W., ... & Liu, S. (2020). Relationship between high-sensitivity C reactive protein and the risk of gallstone disease: results from the Kailuan cohort study. BMJ open, 10(9), e035880.
14. Ma, Y., Yue, P., Zhang, J., Yuan, J., Liu, Z., Chen, Z., ... & Meng, W. (2024). Early prediction of acute gallstone pancreatitis severity: a novel machine learning model based on CT features and open access online prediction platform. Annals of Medicine, 56(1), 2357354.
15. Salem, N. M., Jack, K. M., Gu, H., Kumar, A., Garcia, M., Yang, P., & Dinu, V. (2023). Machine and deep learning identified metabolites and clinical features associated with gallstone disease. Computer Methods and Programs in Biomedicine Update, 3, 100106.
16. Su, P. Y., Hsu, Y. C., Cheng, Y. F., Kor, C. T., & Su, W. W. (2019). Strong association between metabolically-abnormal obesity and gallstone disease in adults under 50 years. BMC gastroenterology, 19(1), 117.
17. Sun, H., Warren, J., Yip, J., Ji, Y., Hao, S., Han, W., & Ding, Y. (2022). Factors influencing gallstone formation: a review of the literature. Biomolecules, 12(4), 550.
18. Wang, X., Yu, W., Jiang, G., Li, H., Li, S., Xie, L., ... & Mao, M. (2024). Global epidemiology of gallstones in the 21st century: a systematic review and meta-analysis. Clinical Gastroenterology and Hepatology, 22(8), 1586-1595.
19. Wenqian, Y. U., Jing, X. I. A., Fangyuan, C. H. E. N., Peng, J. I. A. O., Ping, C. U. I., Yu, W. A. N. G., ... & Xin, W. A. N. G. (2024). Establishment and validation of a predictive model for gallstone disease in the general population: a multicenter study. Journal of Sichuan University (Medical Sciences), 55(3), 641.
20. Yao, C., Wu, S., Liu, Z., & Li, P. (2019). A deep learning model for predicting chemical composition of gallstones with big data in medical Internet of Things. Future Generation Computer Systems, 94, 140-147.
21. Zhang, H., Gao, J., Sun, Z., Zhang, Q., Qi, B., Jiang, X., ... & Shang, D. (2023). Diagnostic accuracy of updated risk assessment criteria and development of novel computational prediction models for patients with suspected choledocholithiasis. Surgical Endoscopy, 37(9), 7348-7357.
22. Zhen, X., Yisen, H., Weirong, J., Zhiwen, L., Rui, L., Wenhao, C., ... & Jianli, H. (2025). The mediating role of inflammatory factors in the relationship between obesity and gallstone disease: evidence from the NHANES 2017–2020. European Journal of Medical Research, 30(1), 650.
23. Zhu, Z., Luo, K., Zhang, B., Wang, G., Guo, K., Huang, P., & Liu, Q. (2023). Risk factor analysis and construction of prediction models of gallbladder carcinoma in patients with gallstones. Frontiers in oncology, 13, 1037194.