



OPEN ACCESS

EDITED BY

Mingjie Wang,
Shanghai Jiao Tong University, China

REVIEWED BY

Chenghai Liu,
Shanghai University of Traditional Chinese
Medicine, China
Xiaohua Chen,
Shanghai Jiao Tong University, China

*CORRESPONDENCE

Zhili Tan,
✉ lwfdtzi895@live.com
Lihong Qu,
✉ 1905365@tongji.edu.cn

[†]These authors have contributed equally
to this work

RECEIVED 29 October 2025

REVISED 11 November 2025

ACCEPTED 17 November 2025

PUBLISHED 26 November 2025

CITATION

Kong N, Wang K, Wang Y, Lou S, Zhou L,
Wang T, Tan Z and Qu L (2025) A machine
learning-based predictive model for 48-week
hepatitis B surface antigen seroclearance in
chronic hepatitis B patients treated with
pegylated interferon α -2b: prediction at week
24.

Front. Cell Dev. Biol. 13:1734654.
doi: 10.3389/fcell.2025.1734654

COPYRIGHT

© 2025 Kong, Wang, Wang, Lou, Zhou, Wang,
Tan and Qu. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A machine learning-based predictive model for 48-week hepatitis B surface antigen seroclearance in chronic hepatitis B patients treated with pegylated interferon α -2b: prediction at week 24

Nan Kong[†], Kaixia Wang[†], Yiling Wang[†], Shike Lou,
Luocheng Zhou, Tao Wang, Zhili Tan* and Lihong Qu*

Department of Infectious Diseases, Shanghai East Hospital, Tongji University School of Medicine,
Shanghai, China

Background: Chronic hepatitis B (CHB) is an infectious disease mainly affecting the liver, caused by the hepatitis B virus (HBV). In the treatment of CHB, pegylated interferon α -2b (PEG-IFN α -2b) is one of the important therapeutic options. However, there are significant individual differences in patients' responses to this treatment and only a few patients can achieve hepatitis B surface antigen (HBsAg) seroclearance. Therefore, an effective method to identify patients with a high likelihood of favorable response at an early stage is urgently needed.

Methods: In this study, we analyzed data from CHB patients who received antiviral treatment with PEG-IFN α -2b and completed 48 weeks of follow-up in the "OASIS" Project. Patients were divided into the seroclearance group and the non-seroclearance group based on whether HBsAg seroclearance was achieved at week 48. Five distinct machine learning feature selection algorithms were used to identify the optimal predictive variables for HBsAg seroclearance. These key variables were then incorporated into 12 machine learning algorithms to build predictive models for HBsAg seroclearance. The best-performing model was selected, and its performance was evaluated.

Results: A total of 680 subjects were included in this study, comprising 165 in the 48-week seroclearance group and 515 in the 48-week non-seroclearance group. Through five different machine learning feature selection algorithms, 11 variables were identified and used to construct 12 distinct machine learning models. Comparative analysis of these models, based on the Area Under the Receiver Operating Characteristic Curve (AUC) and Decision Curve Analysis (DCA) results from the training set, indicated that the Random Forest model was the optimal model for predicting HBsAg seroclearance.

Conclusion: The Random Forest model effectively predicted the 48-week HBsAg seroclearance rate using indicators measured at 24 weeks of PEG-IFN α -2b therapy. This model can provide a reliable reference for optimizing clinical treatment strategies.

KEYWORDS

chronic hepatitis B, HBsAg seroclearance, predictive model, machine learning, clinical utility

1 Introduction

CHB is an infectious liver disease caused by persistent hepatitis B virus (HBV) infection for more than 6 months. According to the WHO's 2024 Global Hepatitis Report, it affected 254 million people globally in 2022, with approximately 1.2 million new infections and 1.1 million deaths annually, primarily due to cirrhosis and hepatocellular carcinoma (Burki, 2024). In China, the CHB epidemic remains a serious concern. A nationwide study revealed approximately 75 million individuals living with chronic HBV infection. Despite this progress from 120 million in 1992, the prevention and control situation remains challenging. Among these 75 million, an estimated 30 million are unaware of their status, and 17 million require antiviral treatment, yet only three million are receiving it (Hui et al., 2024). These critical gaps in diagnosis and treatment expose a vast number of individuals to a high risk of disease progression to severe complications, including cirrhosis and hepatocellular carcinoma (HCC), which severely diminish quality of life and create a heavy socioeconomic burden (Jian et al., 2025).

The achievement of HBsAg seroclearance, regarded as a “functional cure” for CHB, significantly reduces the risk of cirrhosis and HCC, thereby establishing it as a primary therapeutic goal. As a key drug for immunomodulatory treatment of CHB, PEG-IFN α -2b is widely used in eligible patients due to its advantages of limited treatment duration and potential curability. However, treatment response exhibits substantial inter-individual variation, and the rate of HBsAg seroclearance remains low. Moreover, the current clinical paradigm requires awaiting 48-week outcomes to assess final response, which can result in ineffective treatment for non-responders, escalating both healthcare costs and the risk of adverse events (Hong et al., 2022). Early identification of patients with a high likelihood of HBsAg seroclearance during PEG-IFN α -2b treatment is crucial for optimizing individualized CHB treatment. Currently, most domestic and international studies continue to focus on baseline or early-treatment (e.g., week 12) predictors (Lee et al., 2025; Ren et al., 2021; Rijckborst et al., 2010; Sonneveld et al., 2010; Ye et al., 2024). However, week 24 of therapy is a well-established key timepoint for assessing response to interferon-based regimens and determining whether to adjust treatment strategies (Hong et al., 2022). Developing a robust predictive model at this juncture, integrating multidimensional data such as dynamic liver function indices and serological virological markers, can provide clinicians with a practical tool at this critical decision-making window. This enables early warning for potential non-responders, allowing timely strategy modifications, while reinforcing treatment confidence for potential responders, ultimately advancing personalized and precision management of CHB.

In the era of big data and artificial intelligence, machine learning (ML)—a concept pioneered by Arthur Samuel—enables computers to simulate human learning and predict outcomes through mathematical modeling (Deo, 2015). ML encompasses a wide range of mathematical models, such as Decision Trees, Random Forests, Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Neural Networks, and Naive Bayes (NB) (Asnicar et al., 2024; Kotsiantis et al., 2006; Sim et al., 2023). ML techniques are particularly adept at processing high-dimensional clinical data and uncovering complex relationships to build reliable ML models. (Fan et al., 2023; Lee et al., 2023; Xue et al., 2025).

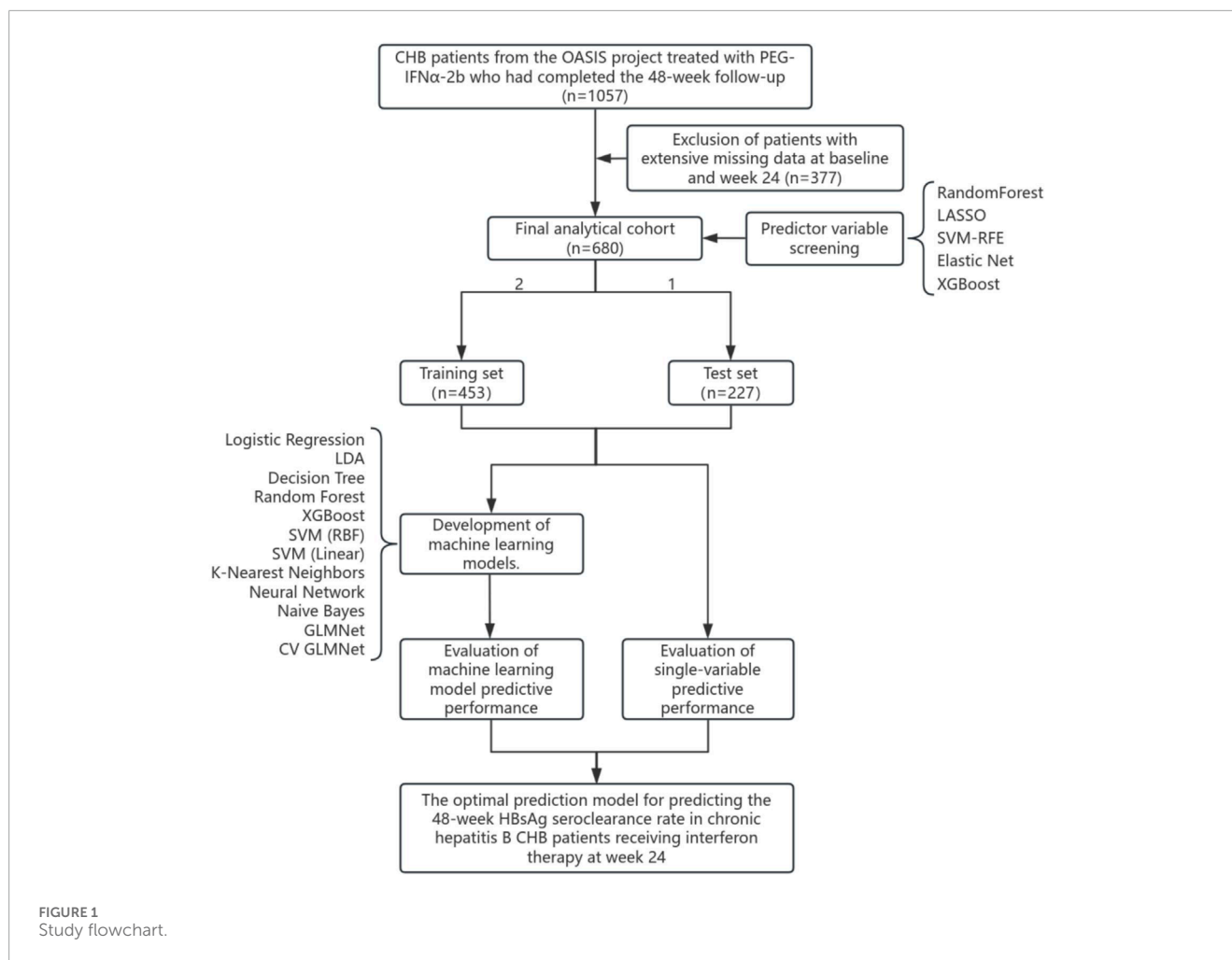
Based on the multi-center, prospective, real-world OASIS project, this study collected clinical data from CHB patients treated with PEG-IFN α -2b. The aim was to develop and validate a ML-based prediction model that utilizes accessible indicators at 24 weeks of treatment—such as quantitative HBsAg and virological markers (e.g., HBV DNA load)—to accurately predict the HBsAg seroclearance rate at 48 weeks. The resulting model is expected to provide clinicians with a decision-support tool for the early identification of patients who are most likely to respond favorably to PEG-IFN α -2b therapy, thereby reducing ineffective treatments and advancing the optimization of individualized CHB management.

2 Materials and methods

2.1 Participants

This study utilized data from the “China Hepatitis Prevention and Treatment Foundation - Project to Reduce Hepatocellular Carcinoma Incidence in Chronic Hepatitis B Patients” (the OASIS project, NCT04896255). This large, multicenter, prospective, real-world study enrolled CHB patients from 32 provinces across China, including treatment-naïve individuals, those previously treated with PEG-IFN α -2b, and those with nucleos(t)ide analogues (NAs) experience. Participants received either a PEG-IFN α -2b-based regimen (as monotherapy or in combination with NAs) or NAs monotherapy and are being followed for a planned 5-year period. This study was approved by the Ethics Committee of Shanghai East Hospital, Tongji University (Approval Number: [2020] Pre-review No. (157)). Written informed consent was obtained from all participants. The data were sourced from the OASIS project, with data cleaning and quality control handled by National Medical Center of Infectious Diseases (NMCID) Liver Diseases Research Group.

The specific inclusion criteria were as follows: (1) Chronic HBV infection, defined as: positivity for HBsAg for more than 6 months,



or positivity for HBsAg for less than 6 months but with a liver biopsy within the past year meeting the pathological criteria for chronic hepatitis B, with other liver diseases excluded. (2) Age between 18 and 80 years, inclusive, irrespective of gender. (3) Treatment with PEG-IFNα-2b-based antiviral therapy. (4) Completion of the 48-week follow-up.

Of the 1,057 eligible CHB patients, 377 were excluded from the final analysis due to extensive missing data at baseline or the 24-week time point. Consequently, the final analytical cohort consisted of 680 patients. A flowchart detailing the study design and patient selection process is presented in [Figure 1](#).

2.2 Data preprocessing and feature engineering

The clinical dataset for this study comprised demographic and clinical characteristics—including age, gender, height, weight, body mass index (BMI), cirrhosis status, and family history—of CHB patients treated with PEG-IFNα-2b. It also included serial serological parameters measured at both baseline and 24 weeks (6 months) of treatment. These parameters covered virological markers (HBV DNA, HBsAg, HBeAg, HBeAb, HBcAb), liver function

indices (ALT, AST, TBIL, DBIL, ALP, GGT, ALB, GLO), and routine blood test results (PLT, NEU, Hb). The initial data processing steps included: handling missing values, applying logarithmic transformation to viral load markers (HBV DNA, HBsAg, HBeAg, HBeAb, HBcAb), and deriving kinetic variables (HBsAg_desc = baseline HBsAg level - week 24 HBsAg level; HBeAg_desc = baseline HBeAg level - week 24 HBeAg level; HBeAb_desc = baseline HBeAb level - week 24 HBeAb level; HBcAb_desc = baseline HBcAb level - week 24 HBcAb level). Categorical variables (gender, cirrhosis, family history) were factor-encoded. After excluding patients with excessive missing data, the final dataset contained 31 predictor variables. The quantitative results for HBV DNA and HBsAg were $\log_{10}$ -transformed during data preprocessing.

2.3 Feature selection using machine learning

To identify robust predictors of treatment response, we employed five distinct feature selection algorithms. (1) Random Forest (RF): we calculated mean decrease in Gini impurity using the randomForest package to rank feature importance ([Chen et al., 2024](#); [Hosseini Sarkhosh et al., 2022](#)). (2) Least Absolute Shrinkage

and Selection Operator (LASSO): using 5-fold cross-validation via glmnet, we selected non-zero coefficients at $\lambda_{\min}$ to retain features with strongest associations (Friedman et al., 2010; Simon et al., 2011). (3) Support Vector Machine-Recursive Feature Elimination (SVM-RFE): we implemented backward feature elimination with linear SVM kernel and 5-fold cross-validation (Huang et al., 2018). (4) Elastic Net: combining L1 and L2 regularization ($\alpha = 0.5$), we identified features with persistent coefficients across cross-validation folds (Hughey and Butte, 2015). (5) XGBoost: we utilized gradient boosting with xgboost package, ranking features by gain importance metric which quantifies the total reduction of loss contributed by each feature. The model was configured with multi:softmax objective, maximum depth of 3, learning rate of 0.1, and 100 boosting rounds with early stopping (Chen and Guestrin, 2016; Liang et al., 2025).

Features consistently ranked highly across multiple methods were prioritized for model construction.

2.4 Predictive model development

We trained 12 diverse classifiers using the mlr3 framework, including (1) Linear models: Logistic Regression, Linear Discriminant Analysis; (2) Tree-based models: Decision Tree, Random Forest, XGBoost; (3) Kernel methods: SVM (linear/RBF kernels); (4) Distance-based: k-Nearest Neighbors; (5) Neural networks: Single-hidden-layer Network; (6) Probabilistic: Naïve Bayes; (7) Regularized regression: GLMNet, CV GLMNet. All models were configured for probability prediction and trained on a stratified 2:1 train-test split. Hyperparameters used default values from respective mlr3learners implementations. Model performance was assessed using: Discrimination metrics: Area under ROC curve (AUC), accuracy, sensitivity, specificity, precision; F1-score; Calibration: Precision-Recall curves and net benefit analysis via Decision Curve Analysis (DCA); Stability: Comparative training-test performance and ranking consistency across metrics.

We implemented k-fold cross-validation during feature selection phases and held-out test set validation for final model assessment. Additionally, for optimal models, we performed SHapley Additive exPlanations (SHAP) analysis using DALEXtra to compute Shapley values for feature contribution quantification. Besides, We also performed univariate ROC analysis of single features and compared with ML models. This comprehensive framework ensured robust feature selection, validated model performance, and facilitated clinical interpretation of predictive factors for interferon response.

2.5 Statistical analyses

Based on HBsAg seroclearance status at week 48, patients were categorized into “Seroclearance” and “Non-Seroclearance” groups. Continuous variables with a normal distribution are presented as mean (standard deviation, SD) and compared using the independent samples t-test. Non-normally distributed variables are expressed as median (interquartile range, IQR) and compared using the Mann-Whitney U test. Categorical variables are summarized as percentages and compared using the chi-square test. All statistical analyses were

performed with R software (version 4.5.1). The significance level (α) was set at 0.05, and a result with $P < 0.05$ was considered statistically significant.

3 Results

3.1 Characteristics of the study

Of the 680 CHB patients included in this study, 165 (24.3%) were classified into the seroclearance group and 515 (75.7%) into the non-seroclearance group. The data preprocessing pipeline began with the enumeration of missing values across all observations. Subsequent analysis was confined to clinical data from baseline (0W) and the 6-month (24W) time points. After excluding patients with a substantial amount of missing data, the final dataset comprised the following 31 predictor variables: Age, Gender, BMI, Liver Cirrhosis (LC), Family history, HBsAg (0W), HBsAg (24W), HBsAg_desc, HBeAg(0W), HBeAg(24W), HBeAg_desc, HBeAb (0W), HBeAb (24W), HBeAb_desc, HBcAb (0W), HBcAb (24W), HBcAb_desc, ALT (0W), ALT (24W), AST (0W), AST (24W), GGT (0W), GGT (24W), TBIL (0W), TBIL (24W), DBIL (0W), DBIL (24W), ALP (0W), ALP (24W), and PLT (0W), PLT (24W) (Supplementary Table S1).

The baseline characteristics of two groups are presented in Table 1. Differences between groups were assessed using the Chi-square test and the Mann-Whitney U test, with a statistical significance level set at $\alpha = 0.05$. Significant baseline differences (all $P < 0.05$) between groups were observed in: LC (4.2% vs. 9.7%, $P = 0.041$), Family history (22.4% vs. 31.8%, $P = 0.027$) HBV DNA (median 0.00 vs. 1.62 $\log_{10}$ IU/mL, $P = 0.034$), HBsAg (median 1.86 vs. 3.08 $\log_{10}$ IU/mL, $P < 0.001$), HBeAg (median 0.40 vs. 0.52 COI, $P < 0.001$), HBeAb (median 0.03 vs. 0.18 COI, $P = 0.001$), HBcAb (median 3.29 vs. 3.28 COI, $P = 0.012$), ALT (median 26.70 vs. 31.00 U/L, $P = 0.004$), AST (median 23.70 vs. 27.00 U/L, $P = 0.006$), GGT (median 19.00 vs. 24.00 U/L, $P = 0.001$), ALB (median 46.00 vs. 45.10 g/L, $P < 0.001$), and Hb (median 148.00 vs. 153.00 g/L, $P = 0.005$). The laboratory parameters of two groups after 24 weeks of PEG-IFN α -2b treatment are presented in Table 2. Statistical comparisons were performed using the Chi-square test and Mann-Whitney U test, with the significance level set at $\alpha = 0.05$. The analysis revealed statistically significant differences (all $P < 0.05$) in the following variables at week 24: HBsAg (median 0.00 vs. 2.47 $\log_{10}$ IU/mL, $P < 0.001$), HBeAg (median 0.37 vs. 0.45 COI, $P < 0.001$), HBeAb (median 0.04 vs. 0.26 COI, $P < 0.001$), and Hb (median 131.00 vs. 135.00 g/L, $P = 0.006$).

Using the mlr3 framework in R, we implemented a 2:1 random split of patients into training ($n = 453$) and validation ($n = 227$) sets, followed by 500 bootstrap iterations to enhance the robustness of model assessment. Comparative analysis of baseline and 24-week treatment characteristics between the training and validation sets (Tables 3, 4) showed that, aside from family history (35.2% vs. 26.7%, $P = 0.027$), ALT at week 24 (median 48.80 vs. 44.00 U/L, $P = 0.023$), AST at week 24 (median 46.20 vs. 42.20 U/L, $P = 0.048$), and GGT at week 24 (median 55.00 vs. 48.00 U/L, $P = 0.013$), no other indicators exhibited significant differences ($P > 0.05$). This indicates that the baseline characteristics of the two sets were overall well-balanced, supporting their use for model training and validation.

TABLE 1 Baseline characteristics and intergroup comparisons of the 680 patients.

Items	Total patient (n = 680)	Seroclearance group (n = 165)	Non-seroclearance group (n = 515)	p
Age (years)	38.00 [32.00, 46.00]	38.00 [32.00, 45.00]	38.00 [32.00, 46.00]	0.681
Height (cm)	170.00 [163.00, 174.00]	170.00 [162.00, 174.00]	170.00 [163.50, 175.00]	0.282
Weight (kg)	69.00 [59.00, 75.00]	67.00 [56.00, 74.00]	70.00 [60.00, 75.00]	0.06
BMI(kg/m ²)	23.53 [21.25, 25.12]	23.15 [20.96, 24.91]	23.66 [21.47, 25.26]	0.088
Gender (%)				
Male	467 (68.7)	104 (63.0)	363 (70.5)	0.089
Female	213 (31.3)	61 (37.0)	152 (29.5)	
LC (%)				
Yes	57 (8.4)	7 (4.2)	50 (9.7)	0.041
No	623 (91.6)	158 (95.8)	465 (90.3)	
Family history (%)				
Yes	201 (29.6)	37 (22.4)	164 (31.8)	0.027
No	479 (70.4)	128 (77.6)	351 (68.2)	
HBVDNA (log ₁₀ IU/ml) (0W)	1.34 [0.00, 3.89]	0.00 [0.00, 3.08]	1.62 [0.00, 4.26]	0.034
HBsAg (log ₁₀ IU/ml) (0W)	2.81 [1.90, 3.47]	1.86 [0.86, 2.58]	3.08 [2.30, 3.65]	<0.001
HBeAg (COI) (0W)	0.49 [0.10, 1.30]	0.40 [0.09, 0.56]	0.52 [0.14, 1.99]	<0.001
HBeAb (COI) (0W)	0.11 [0.01, 1.55]	0.03 [0.01, 1.20]	0.18 [0.01, 1.76]	0.001
HBcAb (COI) (0W)	3.28 [3.08, 3.44]	3.29 [3.18, 3.49]	3.28 [3.04, 3.42]	0.012
ALT (U/L) (0W)	30.00 [19.00, 50.40]	26.70 [15.70, 42.30]	31.00 [20.00, 51.90]	0.004
AST (U/L) (0W)	26.00 [21.00, 34.19]	23.70 [19.80, 30.60]	27.00 [21.00, 35.30]	0.006
Tbil (μmol/L) (0W)	14.18 [10.90, 18.90]	13.50 [10.80, 19.20]	14.20 [10.94, 18.70]	0.948
Dbil (μmol/L) (0W)	3.00 [2.20, 4.20]	3.10 [2.30, 4.22]	3.00 [2.20, 4.16]	0.472
ALP (μmol/L) (0W)	80.00 [65.15, 97.45]	77.00 [64.00, 90.70]	80.85 [66.85, 98.08]	0.111
GGT (U/L) (0W)	23.00 [15.00, 36.54]	19.00 [13.00, 31.00]	24.00 [16.00, 38.00]	0.001
ALB (g/L) (0W)	45.50 [43.48, 47.32]	46.00 [44.00, 48.00]	45.10 [43.20, 47.00]	<0.001
GLO (g/L) (0W)	29.00 [26.35, 31.00]	28.10 [26.00, 30.85]	29.00 [26.40, 31.35]	0.163
NEU(10 ⁹ /L) (0W)	2.85 [2.16, 3.71]	2.66 [2.06, 3.39]	2.90 [2.20, 3.73]	0.084
PLT (10 ⁹ /L) (0W)	204.00 [164.00, 245.00]	206.00 [161.00, 251.00]	203.00 [165.50, 243.50]	0.984
Hb(g/L) (0W)	152.00 [137.00, 162.00]	148.00 [133.00, 159.00]	153.00 [139.00, 162.50]	0.005

TABLE 2 Comparison between the two groups after 24 weeks of treatment in the 680 patients.

Items	Total patient (n = 680)	Seroclearance group (n = 165)	Non-seroclearance group (n = 515)	p
HBVDNA (log ₁₀ IU/ml) (24W)	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	N/A
HBsAg (log ₁₀ IU/ml) (24W)	2.10 [0.43, 3.14]	0.00 [0.00, 0.40]	2.47 [1.57, 3.37]	<0.001
HBeAg(COI) (24W)	0.44 [0.09, 0.91]	0.37 [0.08, 0.49]	0.45 [0.09, 1.52]	<0.001
HBeAb (COI) (24W)	0.12 [0.01, 1.40]	0.04 [0.01, 0.93]	0.26 [0.02, 1.50]	<0.001
HBcAb (COI) (24W)	3.23 [3.03, 3.42]	3.23 [3.07, 3.44]	3.23 [3.01, 3.41]	0.328
ALT (U/L) (24W)	45.20 [30.00, 71.55]	48.80 [29.42, 80.20]	45.00 [30.20, 68.00]	0.416
AST (U/L) (24W)	44.00 [32.27, 65.00]	46.20 [33.70, 70.00]	43.00 [32.00, 63.00]	0.192
Tbil (μmol/L) (24W)	12.80 [10.30, 15.97]	12.50 [9.94, 15.45]	13.06 [10.41, 16.20]	0.107
Dbil (μmol/L) (24W)	3.20 [2.50, 4.22]	3.20 [2.53, 4.39]	3.20 [2.50, 4.20]	0.512
ALP (μmol/L) (24W)	83.00 [70.85, 98.83]	81.05 [70.92, 97.07]	84.00 [70.75, 99.00]	0.337
GGT (U/L) (24W)	50.10 [32.00, 91.00]	49.00 [32.00, 110.60]	51.00 [32.00, 87.17]	0.886
ALB (g/L) (24W)	44.00 [42.00, 45.41]	44.00 [42.00, 45.60]	44.00 [42.00, 45.39]	0.683
GLO (g/L) (24W)	29.00 [26.60, 32.00]	29.00 [26.40, 32.10]	29.00 [26.73, 32.00]	0.65
NEU(10 ⁹ /L) (24W)	1.63 [1.25, 2.21]	1.61 [1.21, 2.03]	1.63 [1.27, 2.24]	0.215
PLT (10 ⁹ /L) (24W)	116.00 [92.00, 152.00]	121.00 [96.00, 152.00]	114.00 [91.00, 152.00]	0.282
Hb(g/L) (24W)	134.00 [122.00, 146.00]	131.00 [119.00, 143.00]	135.00 [124.00, 147.00]	0.006

3.2 Variable selection

To identify robust predictors of response to interferon-based therapy, this study employed five distinct machine learning-based feature selection algorithms: Random Forest: The results identified the following ten top-ranking features: HBsAg (24W), HBsAg (0W), HBsAg_desc, ALT (24W), GGT (24W), DBIL (24W), HBcAb_desc, HBeAb (24W), HBeAg_desc, and HBcAb (0W), as detailed in [Supplementary Table S2](#). LASSO Regression: The final model retained the following variables: HBsAg (24W), HBeAb (24W), ALT (24W), TBIL (24W), ALP (24W), Age, and Gender, as detailed in [Supplementary Table S3](#). SVM-RFE: The results identified the following top features: HBsAg (24W), HBsAg (0W), HBsAg_desc, HBeAg_desc, HBcAb (0W), HBeAg(24W), GGT (0W), HBeAb (24W),HBeAb_desc, and HBeAg(0W), as presented in [Supplementary Table S4](#). Elastic Net: The algorithm identified the following key variables: HBsAg (0W), HBsAg (24W), and HBsAg_desc, as detailed in [Supplementary Table S5](#). XGBoost: The top 10 variables were: HBsAg (24W), HBsAg (0W), HBsAg_desc, ALT (24W), DBIL (24W), GGT (24W), HBeAb (24W), ALP (24W), HBcAb (0W), and TBIL (24W), as detailed in [Supplementary Table S6](#).

To enhance model generalizability and clinical interpretability, we defined the optimal predictor set as variables selected by at least

two out of the five feature selection algorithms, which yielded 11 key predictors ([Supplementary Table S7](#)). This consensus approach is visually summarized in a Venn diagram ([Supplementary Figure S1](#)). The final predictors were: ALP (24W), ALT (24W), DBIL (24W), GGT (0W), HBcAb (0W), HBeAb (24W), HBeAg_desc, HBsAg (0W), HBsAg_desc, HBsAg (24W), and TBIL (24W). Collectively, this set captures baseline status, on-treatment levels at mid-therapy, and early kinetic changes, thereby providing multi-dimensional predictive information.

3.3 Comparative performance of multiple machine learning models

Based on the selected set of 11 key predictors, we constructed 12 distinct machine learning models, including Logistic Regression, LDA, Decision Tree, Random Forest, XGBoost, SVM (RBF), SVM (Linear), k-Nearest Neighbors, Neural Network, Naive Bayes, GLMNet, and CV GLMNet. The performance of all 12 models was rigorously evaluated on both the training and validation sets. The evaluation metrics included the Area Under the Curve (AUC), accuracy, sensitivity, specificity, precision, and F1-score. Results for each metric are reported as the mean values derived from 500 effective bootstrap samples ([Supplementary Table S8](#)). The Receiver

TABLE 3 Comparison of baseline characteristics between the training and test sets.

Items	Test set (n = 227)	Training set (n = 453)	<i>p</i>
Age	38.00 [32.00, 45.00]	38.00 [32.00, 46.00]	0.872
Height	170.00 [163.00, 175.00]	170.00 [163.00, 174.00]	0.764
Weight	69.00 [60.00, 75.00]	70.00 [58.00, 75.00]	0.894
BMI	23.39 [21.51, 25.08]	23.66 [21.03, 25.26]	0.967
Gender (%)			0.916
Male	157 (69.2)	310 (68.4)	
Female	70 (30.8)	143 (31.6)	
LC (%)			0.458
Yes	16 (7.0)	41 (9.1)	
No	211 (93.0)	412 (90.9)	
Family.history (%)			0.027
Yes	80 (35.2)	121 (26.7)	
No	147 (64.8)	332 (73.3)	
HBVDNA (log ₁₀ IU/ml) (0W)	1.36 [0.00, 3.65]	1.32 [0.00, 4.08]	0.698
HBsAg (log ₁₀ IU/ml) (0W)	2.92 [2.03, 3.38]	2.76 [1.81, 3.54]	0.586
HBeAg(COI) (0W)	0.51 [0.10, 1.06]	0.48 [0.11, 1.40]	0.882
HBeAb (COI) (0W)	0.17 [0.02, 1.55]	0.08 [0.01, 1.53]	0.266
HbAb (COI) (0W)	3.31 [3.15, 3.46]	3.26 [3.05, 3.42]	0.057
ALT (U/L) (0W)	29.90 [19.00, 47.76]	30.00 [19.00, 51.30]	0.86
AST (U/L) (0W)	26.00 [21.00, 33.98]	26.00 [21.00, 35.00]	0.7
Tbil1 (μmol/L) (0W)	14.50 [10.95, 19.57]	14.08 [10.90, 18.60]	0.266
Dbil1 (μmol/L) (0W)	3.17 [2.20, 4.27]	2.93 [2.20, 4.16]	0.226
ALP (μmol/L) (0W)	79.90 [66.07, 94.00]	80.00 [65.03, 98.07]	0.737
GGT (U/L) (0W)	24.00 [15.00, 36.70]	22.40 [15.00, 36.12]	0.589
ALB (g/L) (0W)	46.00 [43.50, 47.45]	45.39 [43.50, 47.00]	0.368
GLO (g/L) (0W)	28.70 [26.60, 31.15]	29.00 [26.17, 31.00]	0.835
NEU(10 ⁹ /L) (0W)	2.90 [2.15, 3.71]	2.84 [2.18, 3.68]	0.812
PLT (10 ⁹ /L) (0W)	201.00 [164.00, 240.50]	205.00 [164.00, 251.00]	0.51
Hb(g/L) (0W)	154.00 [137.50, 162.00]	151.00 [137.00, 161.00]	0.352

Operating Characteristic (ROC) curve was used to assess the discriminative ability of the predictive models. These curves were plotted with 1-specificity on the x-axis and sensitivity on the y-axis, illustrating the performance for both the training and validation sets. In the training set, all models except Neural Network ($AUC = 0.685 \pm 0.168$) demonstrated strong performance ($AUC > 0.85$), as shown in [Figure 2A](#). A comprehensive analysis of the ROC curves from the test set ([Figure 2B](#)) and the multi-metric radar chart

TABLE 4 Comparison of patient characteristics at 24 Weeks of treatment between the training and test sets.

Items	Test set (n = 227)	Training set (n = 453)	p
HBVDNA (log ₁₀ IU/ml) (24W)	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	N/A
HBsAg (log ₁₀ IU/ml) (24W)	2.18 [0.42, 3.07]	2.06 [0.44, 3.21]	0.818
HBeAg(COI) (24W)	0.41 [0.08, 0.84]	0.44 [0.09, 1.00]	0.291
HBeAb (COI) (24W)	0.21 [0.02, 1.45]	0.10 [0.01, 1.37]	0.248
HbAb (COI) (24W)	3.26 [3.05, 3.44]	3.22 [3.01, 3.41]	0.192
ALT (U/L) (24W)	48.80 [32.80, 78.65]	44.00 [29.00, 66.90]	0.023
AST (U/L) (24W)	46.20 [34.00, 68.25]	42.20 [31.60, 63.00]	0.048
Tbil1 (μmol/L) (24W)	13.30 [10.65, 15.97]	12.71 [10.10, 15.97]	0.455
Dbil1 (μmol/L) (24W)	3.20 [2.51, 4.30]	3.20 [2.50, 4.20]	0.621
ALP (μmol/L) (24W)	82.50 [71.50, 99.50]	83.40 [69.25, 98.46]	0.904
GGT (U/L) (24W)	55.00 [35.50, 105.50]	48.00 [30.00, 87.07]	0.013
ALB (g/L) (24W)	44.00 [42.00, 45.30]	43.90 [42.00, 45.60]	0.94
GLO (g/L) (24W)	29.00 [26.83, 32.00]	29.00 [26.50, 32.00]	0.771
NEU(10 ⁹ /L) (24W)	1.58 [1.21, 2.17]	1.64 [1.27, 2.23]	0.242
PLT (10 ⁹ /L) (24W)	114.00 [91.00, 152.50]	117.00 [94.00, 152.00]	0.649
Hb(g/L) (24W)	135.00 [122.50, 146.50]	134.00 [121.00, 146.00]	0.671

Categorical data are presented as frequency (percentage). Normally distributed continuous variables are expressed as mean (standard deviation) and compared using the independent samples t-test. Non-normally distributed variables are summarized as median (interquartile range) and compared using the Mann-Whitney U test.

(Figure 2E) identified three top-performing models: Random Forest (AUC = 0.915 ± 0.020), XGBoost (AUC = 0.891 ± 0.026), and CV GLMNet (AUC = 0.862 ± 0.029) (Figures 2C,D). In contrast, models such as k-Nearest Neighbors and Neural Network showed marked performance degradation on the validation set, with their ROC curves deviating substantially from the upper-left corner, indicating poorer generalizability.

3.4 Analysis of top-performing models

Among the 12 machine learning models evaluated, Random Forest, XGBoost, and CV GLMNet consistently outperformed the others. The discriminative performance of Random Forest on the validation set was as follows: AUC = 0.915 ± 0.020, accuracy = 0.887 ± 0.019, precision = 0.853 ± 0.062, sensitivity = 0.707 ± 0.066, specificity = 0.954 ± 0.021, and F1-score = 0.770 ± 0.042. XGBoost achieved: AUC = 0.891 ± 0.026, accuracy = 0.878 ± 0.020, precision = 0.820 ± 0.064, sensitivity = 0.705 ± 0.063, specificity = 0.942 ± 0.023, and F1-score = 0.755 ± 0.043. CV GLMNet yielded: AUC = 0.862 ± 0.029, accuracy = 0.836 ± 0.063, precision = 0.783 ± 0.089, sensitivity = 0.567 ± 0.271, specificity = 0.940 ± 0.041, and F1-score = 0.724 ± 0.061 (Figures 3A,B). The Precision-Recall (PR) curves, plotted for both training and validation sets, demonstrated

that CV GLMNet, Random Forest, and XGBoost all maintained high and stable predictive performance. Their PR curves were positioned close to the upper-right corner and exhibited narrow 95% confidence intervals, indicating a strong and reliable balance between precision and recall (Figures 4A–D). Notably, the Random Forest model achieved the highest AUC on the validation set (Figures 3A,B) and showed a balanced performance across accuracy, sensitivity, specificity, and F1-score (Figure 2F). Its performance on the validation set was comparable to that on the training set, with no significant overfitting observed. This demonstrates superior generalizability, establishing Random Forest as the optimal model for predicting HBsAg seroclearance.

3.5 Clinical utility assessment of the machine learning models

This study evaluated the clinical utility of the twelve machine learning models using Decision Curve Analysis (DCA), with 95% confidence intervals calculated to reflect the reproducibility of the results. As summarized in Table 5, the threshold corresponding to the maximum net benefit was 0.01 for all models. The Random Forest model demonstrated the highest Average Net Benefit Index (ANDI) value of 0.547 and the widest range of positive threshold probabilities

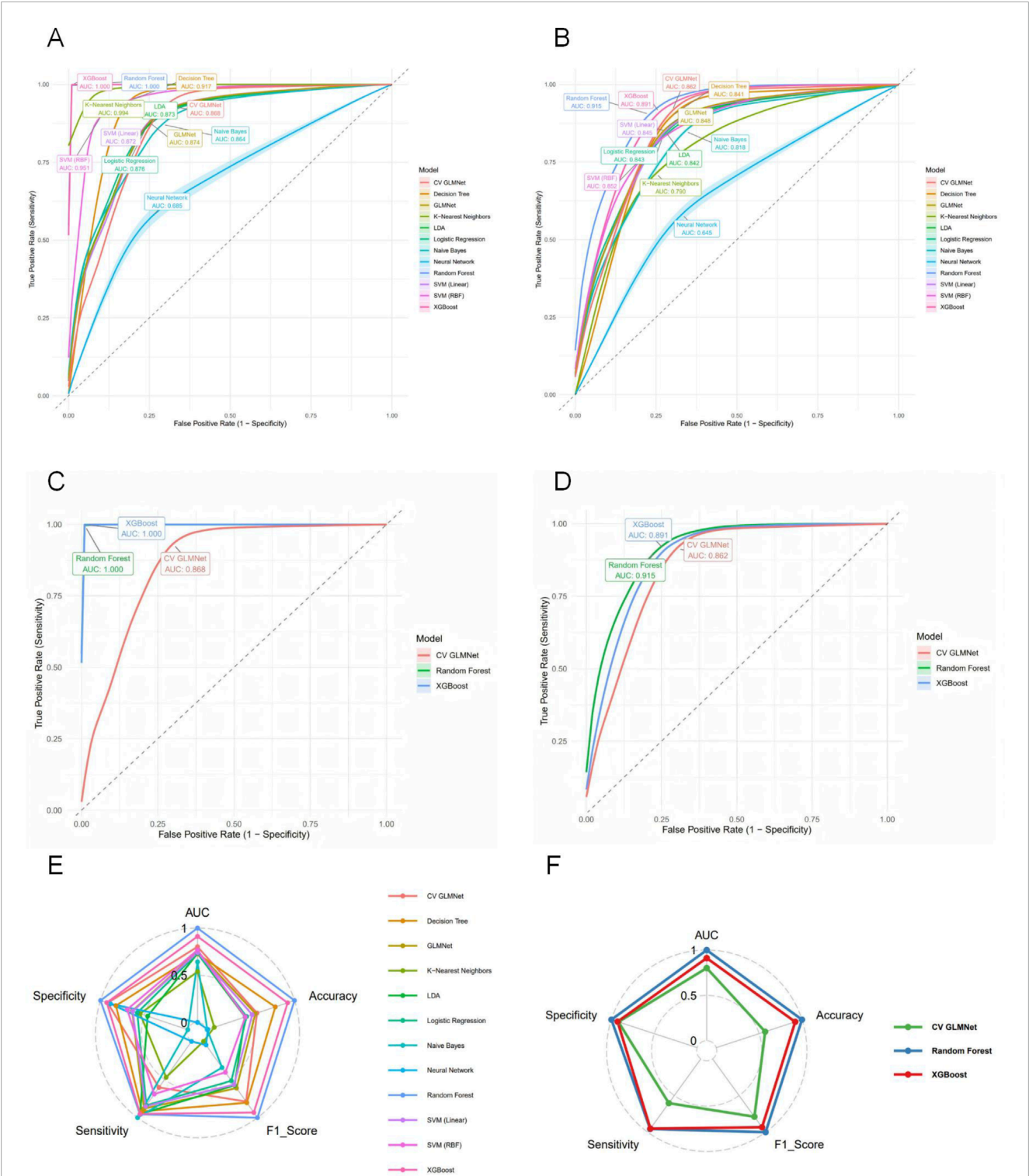


FIGURE 2 Evaluation of predictive performance for the twelve machine learning models. **(A)** ROC curves with 95% confidence intervals (shaded areas) for all models on the training set. **(B)** ROC curves for all models on the test set. **(C)** ROC curves for the top three performing models on the training set. **(D)** ROC curves for the top three models on the test set. **(E)** Radar plot comparing six performance metrics across all models on the test set. **(F)** Radar plot highlighting the balanced multi-metric performance of the top three models (ranked by test set AUC). The models are distinguished by color in all panels. The top three models, Random Forest, XGBoost, and CV GLMNet, consistently demonstrated superior and stable discriminative ability across both.

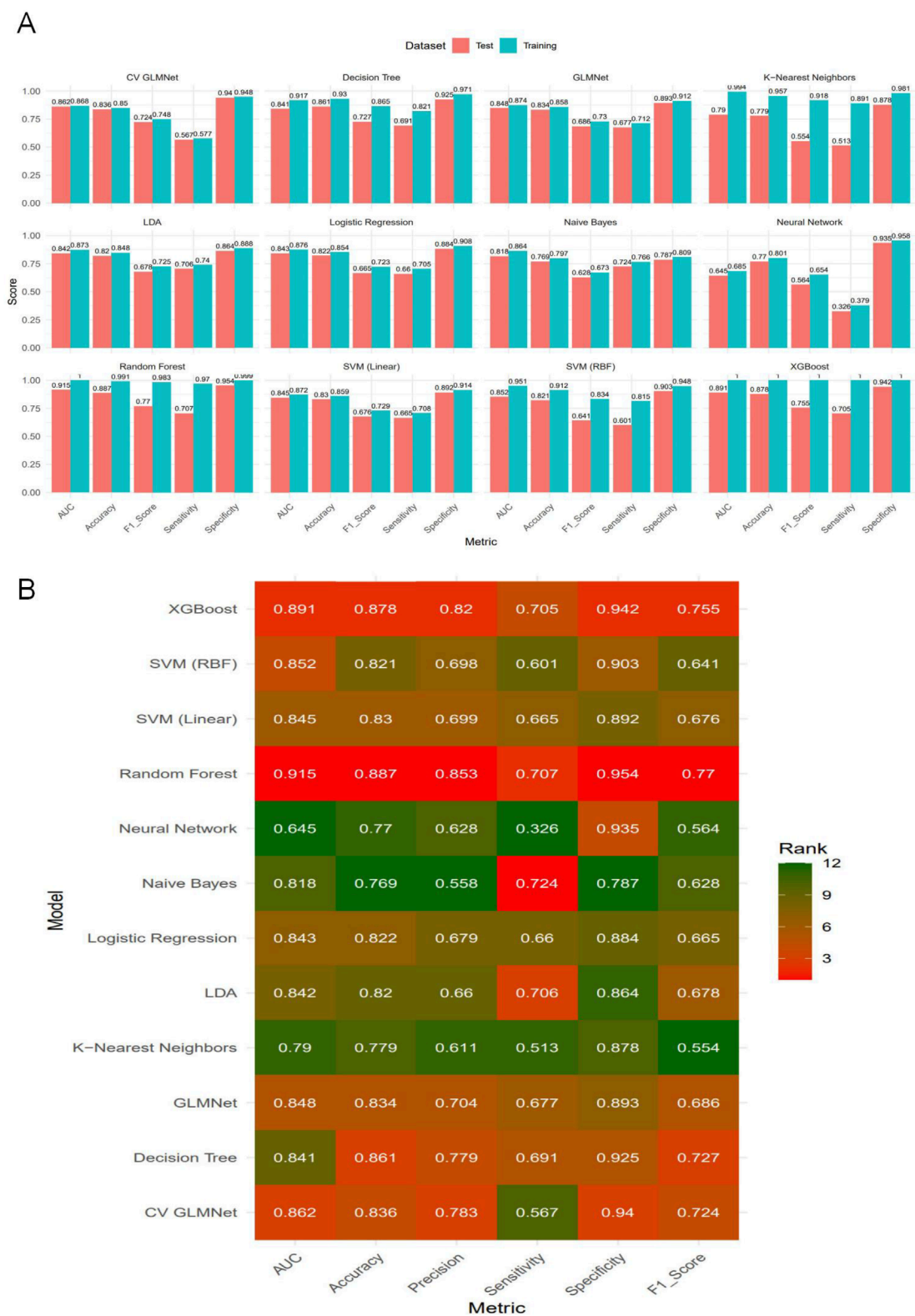
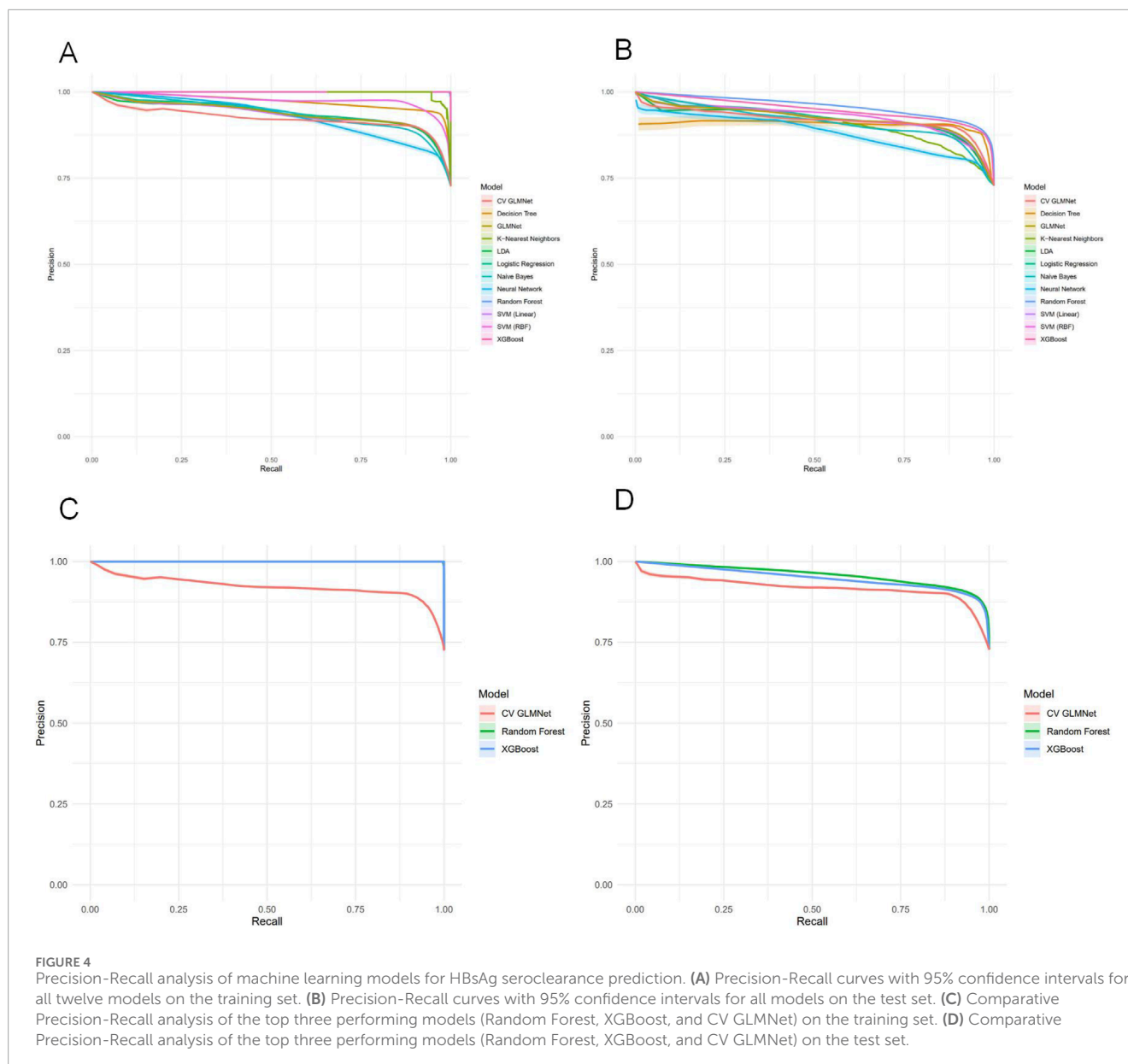


FIGURE 3 Comparative performance evaluation of the twelve machine learning models. **(A)** Bar plot displaying the bootstrap mean estimates of five performance metrics (AUC, Accuracy, F1-Score, Sensitivity, Specificity) for each model on both the training and test sets. **(B)** Heatmap summarizing the model rankings based on bootstrap mean performance on the test set. Color intensity corresponds to the metric score, with darker shades indicating higher values. The top-performing models—Random Forest, CV GLMNet, and XGBoost—consistently achieve superior and balanced results across most evaluation metrics in the test set.



(0.01–0.99). These results indicate its high clinical application value, supporting its use in guiding decisions on whether to continue PEG-IFN α -2b therapy in CHB patients. Decision curves were plotted to illustrate the relationship between decision thresholds (ranging from 0.00 to 1.00) and net benefit (ranging from –0.2–0.6) across the machine learning models. Over a wide range of threshold probabilities, the net benefit of the three top-performing models—Random Forest, XGBoost, and CV GLMNet—was consistently superior to that of the other models and remained higher than the “treat-all” or “treat-none” strategies, demonstrating robust clinical discriminative ability. Among them, the Random Forest model exhibited greater adaptability to threshold variations, as evidenced by its smoother curve and stable performance across different intervention scenarios. This suggests that Random Forest is particularly well-suited for balancing benefits and risks in clinical decision-making regarding antiviral therapy for hepatitis B (Figures 5A,B).

To quantify the contribution of each predictor to the performance of the Random Forest model, we employed both the Permutation Importance method and SHAP analysis to assess variable importance (Figures 5C,D). The Permutation Importance results revealed significant differences in the importance of the 11 predictive features. HBsAg (24W) was identified as the most influential feature for model predictions, indicating that shuffling its values led to the greatest decrease in the model’s AUC. This establishes it as a core variable for maintaining the model’s discriminative ability. The following five features also provided substantial support for model performance: HBeAb (24W), DBIL (24W), GGT (0W), HBsAg_desc, and HBsAg (0W). Based on the SHAP analysis, using the shapviz package in R to compute Shapley values for features, a bar plot was generated to illustrate the contribution of the top 10 features to the model’s predictions in the Random Forest model. The ranking of feature contributions was as

TABLE 5 Decision curve analysis (DCA) of the twelve machine learning models.

Model	ANDI	Max_Net_Benefit	Threshold_at_Max	Positive_Threshold_Range	N_Valid_Bootstraps
Logistic regression	0.4732	0.7262	0.01	0.01–0.93	500
LDA	0.4712	0.7264	0.01	0.01–0.93	500
Decision tree	0.491	0.7221	0.01	0.01–0.90	500
Random forest	0.547	0.7271	0.01	0.01–0.99	500
XGBoost	0.4528	0.7238	0.01	0.01–0.91	500
SVM (RBF)	0.4881	0.7271	0.01	0.01–0.94	500
SVM (linear)	0.4772	0.7268	0.01	0.01–0.94	500
K-nearest neighbors	0.3568	0.7166	0.01	0.01–0.89	500
Neural network	0.4116	0.7265	0.01	0.01–0.88	500
Naive bayes	0.3954	0.7173	0.01	0.01–0.90	500
GLMNet	0.4799	0.7267	0.01	0.01–0.93	500
CV GLMNet	0.4755	0.7271	0.01	0.01–0.93	500

follows: HBsAg (24W) > HBeAb (24W) > HBsAg_desc > HBsAg (0W) > GGT (0W) > ALT (24W) > HBeAg_desc > DBIL (24W) > HBcAb (0W) > ALP (24W). Notably, HBcAb (0W) and ALP (24W) exhibited a negative predictive effect on HBsAg seroclearance, while the remaining variables showed positive predictive effects. Integrating the findings from both analytical methods, we conclude that HBsAg (24W), HBeAb (24W), GGT (0W), HBsAg_desc, and HBsAg (0W) are the key variables in the Random Forest model.

3.6 Predictive performance of individual variables

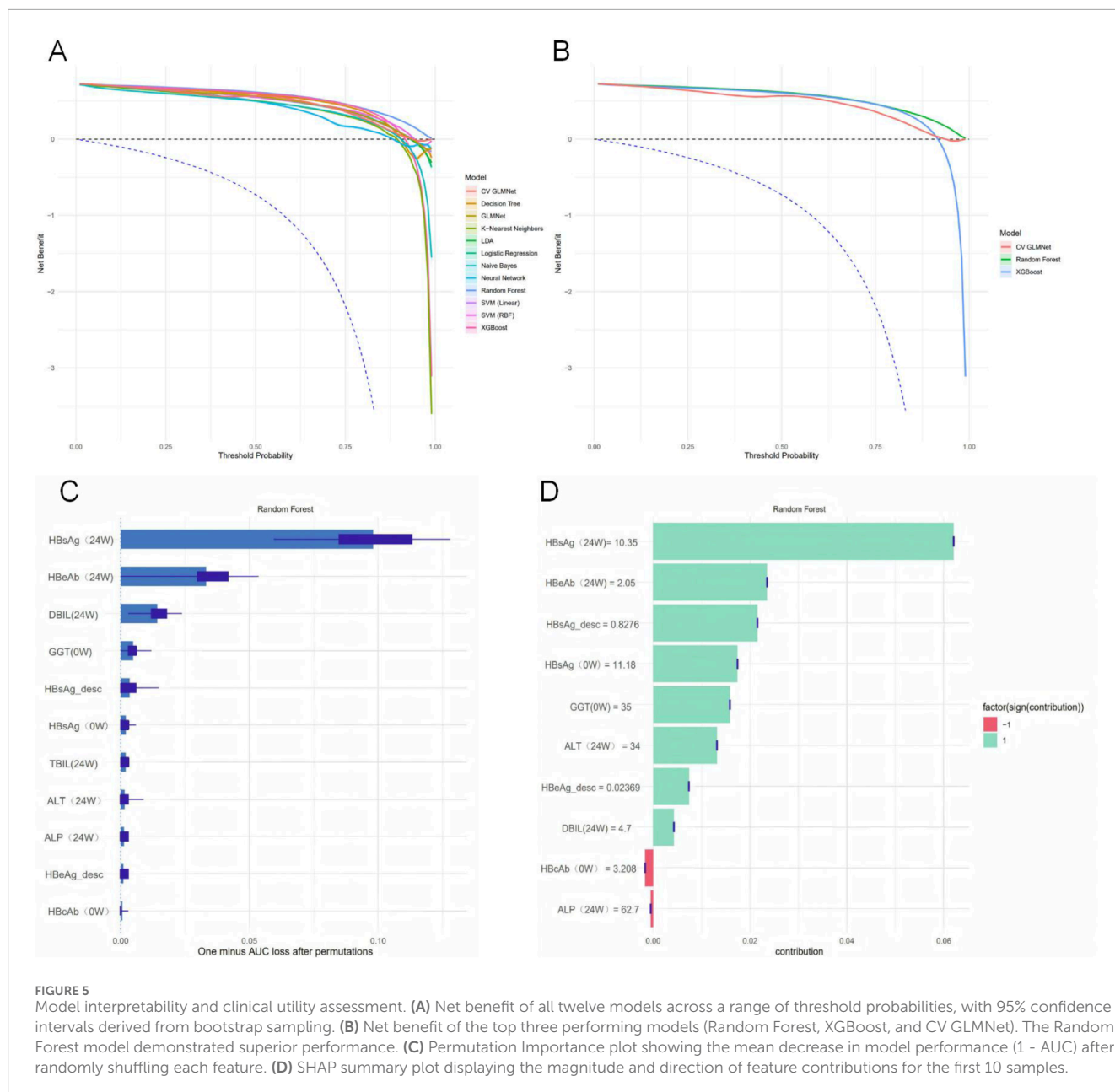
To compare the predictive performance between the machine learning models and individual variables, we conducted a univariate analysis of the 11 predictor variables on the test set, as detailed in [Supplementary Table S9](#). ROC curves were plotted for the top ten individual features ([Figures 6A,B](#)). HBsAg (24W) was identified as the single most predictive feature, achieving an AUC of 0.866 ± 0.030 , accuracy of 0.850 ± 0.025 , precision of 0.723 ± 0.071 , sensitivity of 0.736 ± 0.052 , specificity of 0.893 ± 0.031 , and an F1-score of 0.727 ± 0.046 on the validation set. Other valuable single features, in descending order of AUC, were HBsAg (0W) (AUC = 0.757 ± 0.032) and HBsAg_desc (AUC = 0.685 ± 0.036) ([Figures 6C,D](#)). In contrast, most liver function indicators—such as GGT (0W), TBIL (24W), ALP (24W), ALT (24W), and DBIL (24W)—demonstrated low predictive utility (AUC <0.6), indicating their limited value when used alone.

The univariate analysis in the test set was ranked by AUC in descending order ([Figure 7A](#)). Precision–Recall (PR) curves of the top 10 individual variables are plotted with Recall on

the x-axis and Precision on the y-axis, illustrating the precision of each variable at different recall levels ([Figure 7B](#)). The PR curve for HBsAg (24W) was positioned closest to the upper-right corner, confirming its status as the best-performing single-variable predictor. However, this curve also exhibited noticeable fluctuations, reflecting its sensitivity to variations in a single feature. A sharp decline in precision was frequently observed at medium recall levels (0.4–0.6). In contrast, the Random Forest model demonstrated markedly superior and stable performance: its precision remained consistently high (mostly above 0.75) across the entire recall range (0–1) and was maintained at a high level even at high recall values (0.7–0.8), resulting in a much smoother and more robust curve. The Random Forest model, by integrating multiple predictive variables, accurately identifies patients with a high likelihood of achieving HBsAg seroclearance. This precise stratification helps prevent two critical clinical scenarios: firstly, avoiding the premature discontinuation of therapy in potential responders due to misclassification as “non-responders,” which could otherwise lead to a missed cure opportunity; and secondly, preventing the continuation of ineffective treatment in true non-responders misclassified as “responders” thereby reducing unnecessary PEG-IFN α -2b-related side effects and economic burdens ([Table 6](#)).

4 Discussion

This study integrated dynamic serological indicators from CHB patients with machine learning algorithms to develop and validate a predictive model for HBsAg seroclearance following PEG-IFN α -2b therapy. The main findings are as follows:



4.1 Clinical significance of key predictor variables

In this study, following data cleaning and the exclusion of patients with excessive missing data, the final dataset comprised 31 predictor variables spanning demographic and clinical characteristics, virological markers, liver function indices, and routine blood test parameters. These variables were subsequently processed using five distinct machine learning feature selection methods, which identified 11 optimal predictors. These 11 key variables were used to construct 12 machine learning models. Among them, the Random Forest model demonstrated the best performance ($\text{AUC} = 0.915 \pm 0.020$) and surpassed the predictive capability of any single variable. Through Permutation Importance and SHAP analysis, the following variables were identified as the

most influential within the Random Forest model: HBsAg (24W), HBeAb (24W), GGT (0W), HBsAg_desc, and HBsAg (0W).

HBsAg (24W) emerged as the most pivotal predictor in this study, achieving an AUC of 0.866 in the univariate analysis on the validation set and consistently ranking as the primary feature across multiple machine learning models. The quantitative HBsAg level serves as a key surrogate marker for the transcriptional activity of hepatitis B virus cccDNA. Both a lower baseline HBsAg level and a lower level at week 24 are strongly associated with a higher probability of HBsAg seroclearance at week 48 of interferon-based therapy (Tan et al., 2025; Wen et al., 2024). Furthermore, the magnitude of HBsAg decline or its absolute value at week 24 has been well-established as a robust predictor of response to interferon treatment (Jiang et al., 2024). Furthermore, incorporating the dynamic change in HBsAg ($\log_{10}\text{IU/mL}$) from

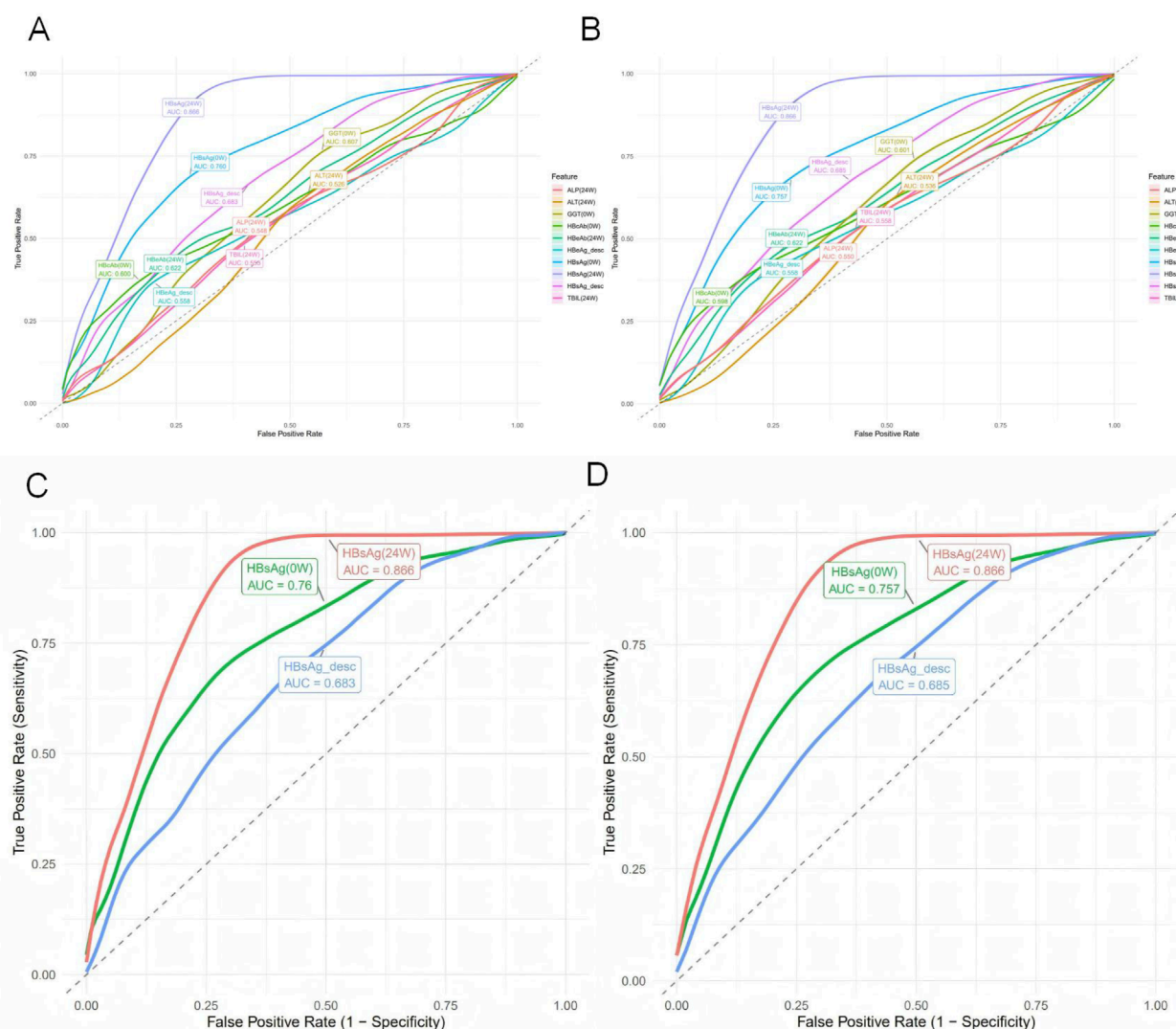


FIGURE 6

Univariate predictive performance of individual variables for HBsAg seroclearance. (A) Receiver operating characteristic (ROC) curves for the top 10 individual predictors on the training set, with area under the curve (AUC) values indicated. (B) ROC curves for the top 10 individual predictors on the test set. (C) ROC curves of the top three univariate predictors (HBsAg (24W), HBsAg (0W), and HBsAg_desc) on the training set. (D) ROC curves of the top three univariate predictors on the test set.

baseline to week 24 further enhanced the model's performance, underscoring that kinetic indicators provide a more reflective measure of treatment response trends than single time-point measurements (Reijnders et al., 2011).

During antiviral therapy, seroconversion of HBeAg at week 24—defined as the loss of HBeAg accompanied by the appearance of HBeAb—serves as a powerful predictor of favorable long-term outcomes. This serological change signifies the effective activation of the patient's immune system, enabling it to better recognize and clear HBV-infected hepatocytes. A robust and effective cellular immune response is a prerequisite for achieving HBsAg seroclearance (Liaw et al., 2012).

This study identified a significant association between baseline GGT levels and HBsAg seroclearance. In prior research, low

baseline GGT has been established as an independent predictor of sustained virological response in chronic hepatitis C treatment. A lower level of GGT indicates a significant reduction in liver inflammation and bile stasis, suggesting a relative preservation of liver cell function, which may better support the antiviral immune response induced by the combination of PEG-IFN and ribavirin (Akuta et al., 2007; Dogan et al., 2014). Furthermore, serum GGT has been proposed as a potential biomarker for predicting the activation of antiviral immunity and HBeAg seroconversion in HBeAg-positive chronic hepatitis B patients receiving nucleos(t)ide analogue (NAs) therapy (Huang, 2015). However, the role of GGT levels in predicting outcomes specifically for CHB patients treated with PEG-IFN α -2b requires further investigation and validation.

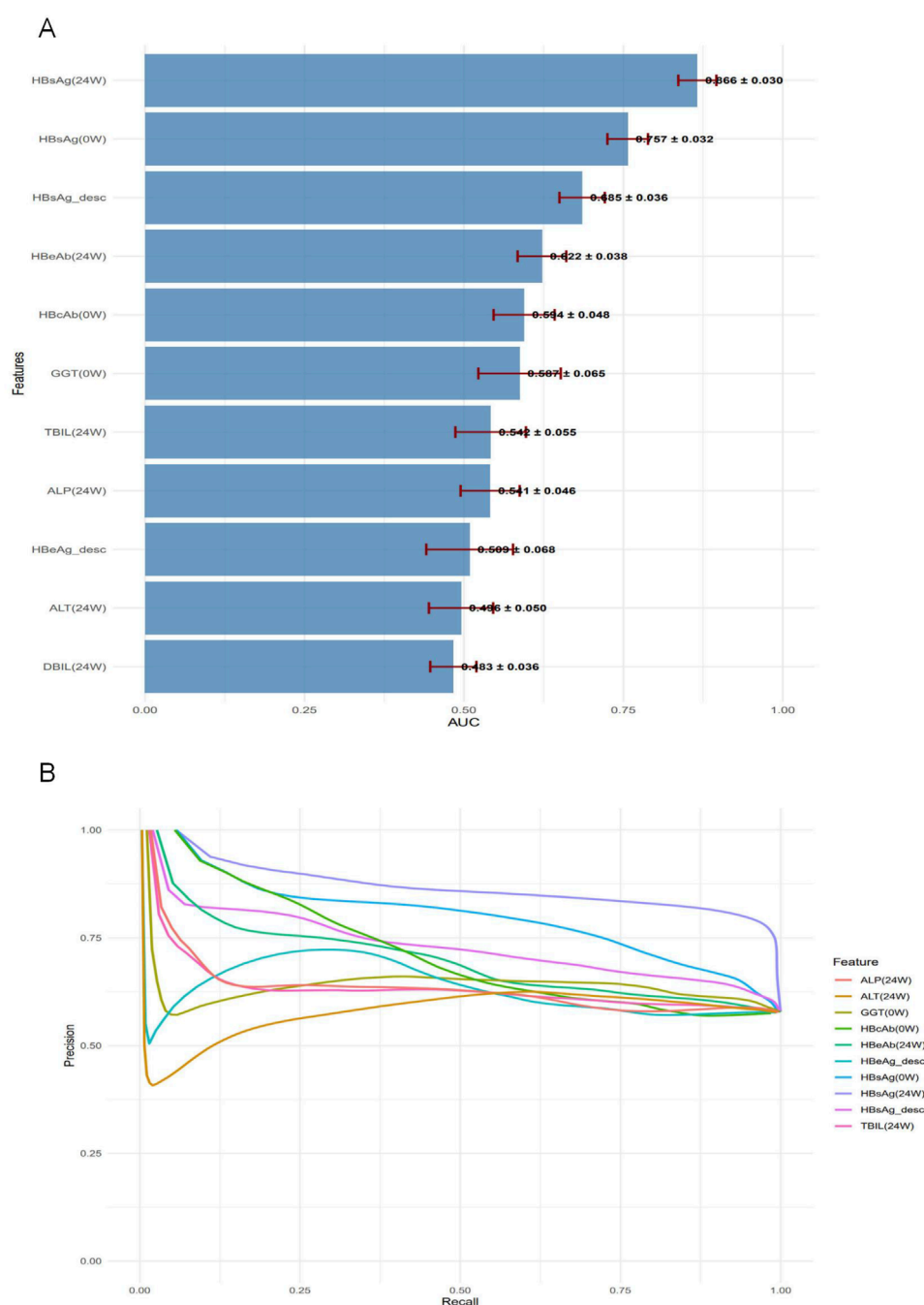


FIGURE 7

Performance assessment of individual predictors for HBsAg seroclearance. **(A)** Bar plot displaying the test set AUC values (mean ± standard deviation) of the univariate predictors, ranked in descending order. **(B)** Precision-Recall curves for the top 10 univariate predictors, illustrating their precision across the full range of recall values.

4.2 Superior performance of the machine learning model

In this study, the Random Forest (RF) model demonstrated superior performance. RF, an ensemble learning algorithm introduced by Breiman in 2001, operates on the core principle of “integrating votes from multiple decision trees” to model

and predict complex data (Breiman, 2001). Its advantages lie in its capability to handle non-linear relationships and feature interactions, making it particularly suitable for modeling the complex associations between serological markers and treatment response (Jiang et al., 2009). Furthermore, its robustness to outliers and missing values reduces the impact of clinical data quality on model performance (Kokla et al., 2019). Permutation Importance,

TABLE 6 Performance comparison between the random forest model and the univariate model (HBsAg at 24 Weeks).

Model Type	Name	AUC	Accuracy	F1 Score
Multivariable	Random forest	0.915 ± 0.020	0.887 ± 0.019	0.770 ± 0.042
Univariable	HBsAg (24W)	0.866 ± 0.030	0.850 ± 0.025	0.727 ± 0.046

a model-agnostic tool for assessing feature significance, evaluates the importance of a variable by randomly shuffling its values in the validation dataset and observing the resultant decline in model performance metrics (e.g., accuracy, AUC). A greater decrease in performance indicates a higher importance of the feature to the model's predictions. This method does not rely on the internal structure or parameters of any specific model, directly quantifying the feature's contribution to the actual predictive performance (Fisher et al., 2019). SHAP values quantify the contribution of each feature to individual predictions, thereby assisting clinicians in understanding the model's decision-making process and addressing the challenge of deploying traditional “black-box” models in clinical practice. By interpreting feature contributions through SHAP analysis, the model's clinical interpretability is significantly enhanced (Lundberg and Lee, 2017; Shu et al., 2025). The combination of Permutation Importance and SHAP analysis provides complementary insights—from global feature importance to local explanations—consistently identifying HBsAg (24W) as the core predictor, with HBeAb (24W), HBsAg_desc, HBsAg (0W), and GGT (0W) as important supporting variables. For future clinical translation, a user-friendly predictive tool (such as a nomogram or a mobile application) could be developed to transform the Random Forest model into an accessible decision-support instrument for clinicians. Furthermore, prospective interventional studies are warranted to validate whether “RF model-guided therapy for CHB” can ultimately improve the rate of HBsAg seroclearance and reduce overall treatment costs.

4.3 Clinical application value of the machine learning model

The Random Forest model developed in this study can predict the likelihood of eventual HBsAg seroclearance in CHB patients as early as 24 weeks after initiating PEG-IFN α -2b therapy. Its clinical application includes several key aspects: For patients predicted as “non-achievers,” clinicians can promptly adjust the treatment strategy (e.g., switching to nucleos(t)ide analogues), thereby avoiding ineffective therapy and unnecessary drug-related adverse events. For those predicted as “achievers,” the positive prediction can reinforce treatment confidence and improve adherence. Furthermore, since all required variables (such as HBsAg and GGT) are routinely measured in standard clinical practice, the model can be implemented without additional testing, facilitating its widespread adoption.

4.4 Clinical translation and future directions

To facilitate the clinical translation of our predictive model, we have outlined a clear roadmap for subsequent research. First, we will precisely calibrate the decision thresholds for predicted probabilities using a prospective cohort to establish a risk stratification framework categorizing patients as “high-, intermediate-, and low-probability responders.” Based on this, a user-friendly clinical decision support tool will be developed to enable instant input of patient indicators and automatic output of stratification-based recommendations. Ultimately, a multicenter randomized controlled trial will be conducted to validate the effectiveness of this decision-making workflow, comparing the HBsAg seroclearance rates and medical costs between the “model-guided strategy” group and the “standard management” group. This systematic effort aims to transition our model from prediction to clinical practice.

5 Study limitations

However, this study has several limitations. First, the predictive model did not incorporate several potentially significant variables, including HBV genotype, HBV RNA, and host genetic polymorphisms (e.g., IL28B genotype). Previous research has established that patients with HBV genotype B exhibit a more pronounced decline in HBsAg and less rebound during interferon therapy, suggesting a more sustained immune response (Brunetto et al., 2009). In contrast, patients with HBV genotype D demonstrate a lower probability of achieving a sustained virological response, regardless of baseline ALT levels or HBV DNA load, identifying genotype D as an independent predictor of poor treatment outcome (Buster et al., 2009). Furthermore, small interfering RNA (siRNA) therapies targeting HBV RNA represent a promising strategy for achieving functional cure, with the substantial HBsAg reduction they induce being a key predictor of subsequent HBsAg seroclearance (Yuen et al., 2024). Additionally, IL28B gene polymorphisms have been significantly associated with HBsAg clearance (OR = 15.534, 95% CI: 1.998–120.777, $P < 0.001$) (Geng et al., 2024; Seto et al., 2013). Other functional biomarkers, including miR-548c-3p (Lin et al., 2023) and CXCL13 (Luo et al., 2023), have also been linked to HBsAg clearance. Future studies should incorporate these features to further refine the model's accuracy. Second, the outcome measure was limited to “HBsAg seroclearance at 48 weeks of treatment” and was not linked to long-term clinical outcomes (e.g., the 5-year risk of hepatocellular carcinoma). Extending the follow-up period is necessary to comprehensively assess the prognostic implications. Furthermore, the model was only internally validated and lacked an independent external validation cohort. Both the training and validation data were sourced from the OASIS project, which may introduce data bias or overfitting. Consequently, the model's generalizability to other independent cohorts (e.g., from different regions or using different testing platforms) and specific subgroups (such as patients with HBV genotype D) remains unverified, potentially affecting the reliability of its clinical application.

6 Conclusion

The Random Forest model, constructed based on serological data from CHB patients, effectively predicts the HBsAg seroclearance rate at 48 weeks of PEG-IFN α -2b therapy. Key predictive variables identified include HBsAg (24W), HBsAb (24W), GGT (0W), HBsAg_desc, and HBsAg (0W). The model demonstrates strong discriminative ability and clinical utility, providing valuable data-driven support for clinicians to develop individualized interferon treatment strategies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by This study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the Institutional Review Board of Shanghai East Hospital, Tongji University (Approval Number: [2020] Pre-review No. (157)). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

NK: Writing – original draft, Formal Analysis, Writing – review and editing. KW: Writing – review and editing, Supervision, Conceptualization. YW: Supervision, Writing – review and editing. SL: Writing – review and editing, Visualization, Software. LZ: Resources, Visualization, Software, Writing – review and editing. TW: Visualization, Software, Writing – review and editing, Resources. ZT: Validation, Supervision, Conceptualization, Writing – review and editing. LQ: Funding acquisition, Writing – review and editing, Validation.

Funding

The authors declare that financial support was received for the research and/or publication of this article. The author(s) declare that

financial support was received for the research and/or publication of this article. This work was supported by the Shanghai Pudong New Area Health Committee Supervision Institute (2025-PWDL-02).

Acknowledgements

We acknowledged all the participating centers of OASIS project and their staff for their invaluable efforts in patients enrollment, follow-up and data collection for this study. We also thanked the members in NMCID Liver Diseases Research Group for their contribution in the data cleaning and quality control work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2025.1734654/full#supplementary-material>

References

- Akuta, N., Suzuki, F., Kawamura, Y., Yatsuji, H., Sezaki, H., Suzuki, Y., et al. (2007). Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis c virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J. Hepatol.* 46 (3), 403–410. doi:10.1016/j.jhep.2006.09.019
- Asnicar, F., Thomas, A. M., Passerini, A., Waldron, L., and Segata, N. (2024). Machine learning for microbiologists. *Nat. Rev. Microbiol.* 22 (4), 191–205. doi:10.1038/s41579-023-00984-1
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/a:1010933404324

- Brunetto, M. R., Moriconi, F., Bonino, F., Lau, G. K. K., Farci, P., Yurdaydin, C., et al. (2009). Hepatitis b virus surface antigen levels: a guide to sustained response to peginterferon alfa-2a in hbeag-negative chronic hepatitis b. *Hepatology* 49 (4), 1141–1150. doi:10.1002/hep.22760
- Burki, T. (2024). Who's 2024 global hepatitis report. *Lancet Infect. Dis.* 24 (6), e362–e363. doi:10.1016/S1473-3099(24)00307-4
- Buster, E. H. C. J., Hansen, B. E., Lau, G. K. K., Piratvisuth, T., Zeuzem, S., Steyerberg, E. W., et al. (2009). Factors that predict response of patients with hepatitis b e antigen-positive chronic hepatitis b to peginterferon-alfa. *Gastroenterology* 137 (6), 2002–2009. doi:10.1053/j.gastro.2009.08.061
- Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Paper presented at the*. New York, NY, USA.
- Chen, J., Wang, X., and Lei, F. (2024). Data-driven multinomial random forest: a new random forest variant with strong consistency. *J. Big Data* 11 (1), 34. doi:10.1186/s40537-023-00874-6
- Deo, R. C. (2015). Machine learning in medicine. *Circulation* 132 (20), 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593
- Dogan, U. B., Akin, M. S., and Yalaki, S. (2014). A low serum γ -glutamyltransferase level predicts a sustained virological response in patients with chronic hepatitis c genotype 1. *Gut Liver* 8 (1), 113–115. doi:10.5009/gnl.2014.8.1.113
- Fan, Z., Jiang, J., Xiao, C., Chen, Y., Xia, Q., Wang, J., et al. (2023). Construction and validation of prognostic models in critically ill patients with sepsis-associated acute kidney injury: interpretable machine learning approach. *J. Transl. Med.* 21 (1), 406. doi:10.1186/s12967-023-04205-4
- Fisher, A., Rudin, C., and Dominici, F. (2019). All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* 20, 177.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01
- Geng, N., Ma, L., Jin, Y., Lu, J., Zheng, Y., Wang, J., et al. (2024). Prediction model for the clearance of hepatitis b surface antigen in patients with chronic hepatitis b before interferon therapy: a prospective case-control study. *Diagnostics* 14 (1), 118. doi:10.3390/diagnostics14010118
- Hong, Y., Fusheng, W., Taisheng, L., Yameng, S., Xiaoyuan, X., Jidong, J., et al. (2022). Guidelines for the prevention and treatment of chronic hepatitis b (version 2022). *Chin. J. Hepatology* 30 (12), 1309–1331. doi:10.3760/cma.j.cn501113-20221204-00607
- Hosseini Sarkhosh, S. M., Esteghamati, A., Hemmatbadi, M., and Daraei, M. (2022). Predicting diabetic nephropathy in type 2 diabetic patients using machine learning algorithms. *J. Diabetes and Metabolic Disord.* 21 (2), 1433–1441. doi:10.1007/s40200-022-01076-2
- Huang, R., Yang, C. C., Liu, Y., Xia, J., Su, R., Xiong, Y. L., et al. (2015). Association of serum gamma-glutamyl transferase with treatment outcome in chronic hepatitis b patients. *World J. Gastroenterol.* 21 (34), 9957–9965. doi:10.3748/wjg.v21.i34.9957
- Huang, X., Zhang, L., Wang, B., Li, F., and Zhang, Z. (2018). Feature clustering based support vector machine recursive feature elimination for gene selection. *Appl. Intell.* 48 (3), 594–607. doi:10.1007/s10489-017-0992-2
- Hughey, J. J., and Butte, A. J. (2015). Robust meta-analysis of gene expression using the elastic net. *Nucleic Acids Res.* 43 (12), e79. doi:10.1093/nar/gkv229
- Hui, Z., Yu, W., Fuzhen, W., Liping, S., Guomin, Z., Jianhua, L., et al. (2024). New progress in hbv control and the Cascade of health care for people living with hbv in China: evidence from the fourth national serological survey, 2020. *Lancet Reg. Health-W. Pac.* 51, 101193. doi:10.1016/j.lanwpc.2024.101193
- Jian, L., Fuzhen, W., Zhongdan, C., Jinlei, Q., Ailing, W., Fanghui, Z., et al. (2025). Research advances in the disease burden of viral hepatitis in China. *J. Clin. Hepatology* 41 (2), 221–227. doi:10.12449/JCH250205
- Jiang, R., Tang, W., Wu, X., and Fu, W. (2009). A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinforma.* 10 (Suppl. 1), S65. doi:10.1186/1471-2105-10-S1-S65
- Jiang, S., Guo, S., Huang, Y., Yin, Y., Feng, J., Zhou, H., et al. (2024). Predictors of hbsag seroclearance in patients with chronic hbv infection treated with pegylated interferon- α : a systematic review and meta-analysis. *Hepatology* 18 (3), 892–903. doi:10.1007/s12072-024-10648-8
- Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., and Hanhineva, K. (2019). Random forest-based imputation outperforms other methods for imputing lc-ms metabolomics data: a comparative study. *BMC Bioinforma.* 20 (1), 492. doi:10.1186/s12859-019-3110-0
- Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* 26 (3), 159–190. doi:10.1007/s10462-007-9052-3
- Lee, H. W., Kim, H., Park, T., Park, S. Y., Chon, Y. E., Seo, Y. S., et al. (2023). A machine learning model for predicting hepatocellular carcinoma risk in patients with chronic hepatitis b. *Liver Int.* 43 (8), 1813–1821. doi:10.1111/liv.15597
- Lee, H. L., Lee, S. K., Han, J. W., Yang, H., Nam, H., Sung, P. S., et al. (2025). Prediction of long-term hbsag seroclearance in patients with hbeag-negative chronic hepatitis b. *JHEP Rep.* 7 (7), 101391. doi:10.1016/j.jhepr.2025.101391
- Liang, D., Wang, L., Zhong, P., Lin, J., Chen, L., Chen, Q., et al. (2025). Perspective: global burden of iodine deficiency: insights and projections to 2050 using xgboost and shap. *Adv. Nutr.* 16 (3), 100384. doi:10.1016/j.advnut.2025.100384
- Liaw, Y., Kao, J., Piratvisuth, T., Chan, H. L. Y., Chien, R., Liu, C., et al. (2012). Asian-pacific consensus statement on the management of chronic hepatitis b: a 2012 update. *Hepatology* 55 (3), 531–561. doi:10.1007/s12072-012-9365-4
- Lin, N., Wang, L., Guo, Z., Guo, S., Liu, C., Lin, J., et al. (2023). Mir-548c-3p targets trim22 to attenuate the peg-ifn- α therapeutic efficacy in hbeag-positive patients with chronic hepatitis b. *Antivir. Res.* 213, 105584. doi:10.1016/j.antiviral.2023.105584
- Lundberg, S., and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Adv. Neural Information Processing Systems* 30.
- Luo, M., Zhang, L., Yang, C., Zhou, B., Hou, J., and Jiang, D. K. (2023). Cxcl13 variant predicts pegylated-interferon α treatment response in hbeag-positive chronic hepatitis b patients. *J. Med. Virol.* 95 (7), e28963. doi:10.1002/jmv.28963
- Reijnders, J. G. P., Rijckborst, V., Sonneveld, M. J., Scherbeijn, S. M. J., Boucher, C. A. B., Hansen, B. E., et al. (2011). Kinetics of hepatitis b surface antigen differ between treatment with peginterferon and entecavir. *J. Hepatology* 54 (3), 449–454. doi:10.1016/j.jhep.2010.07.046
- Ren, P., Li, H., Huang, Y., Jiang, J., Guo, S., Cao, Z., et al. (2021). A simple-to-use tool for predicting response to peginterferon in hbv dna suppressed chronic hepatitis b patients in China. *Antivir. Res.* 194, 105163. doi:10.1016/j.antiviral.2021.105163
- Rijckborst, V., Hansen, B. E., Cakaloglu, Y., Ferenci, P., Tabak, F., Akdogan, M., et al. (2010). Early on-treatment prediction of response to peginterferon alfa-2a for hbeag-negative chronic hepatitis b using hbsag and hbv dna levels. *Hepatology* 52 (2), 454–461. doi:10.1002/hep.23722
- Seto, W., Wong, D. K., Kopaniszen, M., Proitsi, P., Sham, P., Hung, I. F., et al. (2013). Hla-dp and il28b polymorphisms: influence of host genome on hepatitis b surface antigen seroclearance in chronic hepatitis b. *Clin. Infect. Dis.* 56 (12), 1695–1703. doi:10.1093/cid/cit121
- Shu, P., Wang, X., Wen, Z., Chen, J., and Xu, F. (2025). Shap combined with machine learning to predict mortality risk in maintenance hemodialysis patients: a retrospective study. *Front. Med.* 12, 1615950. doi:10.3389/fmed.2025.1615950
- Sim, J. Z. T., Fong, Q. W., Huang, W., and Tan, C. H. (2023). Machine learning in medicine: what clinicians should know. *Singap. Med. J.* 64 (2), 91–97. doi:10.11622/smedj.2021054
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39 (5), 1–13. doi:10.18637/jss.v039.i05
- Sonneveld, M. J., Rijckborst, V., Boucher, C. A. B., Hansen, B. E., and Janssen, H. L. A. (2010). Prediction of sustained response to peginterferon alfa-2b for hepatitis b e antigen-positive chronic hepatitis b using on-treatment hepatitis b surface antigen decline. *Hepatology* 52 (4), 1251–1257. doi:10.1002/hep.23844
- Tan, Z., Kong, N., Zhang, Q., Gao, X., Shang, J., Geng, J., et al. (2025). Predictive model for hbsag clearance rate in chronic hepatitis b patients treated with pegylated interferon α -2b for 48 weeks. *Hepatology* 19 (2), 358–367. doi:10.1007/s12072-024-10764-5
- Wen, C., Wang, Y., Tian, H., Lei, Y., Wang, Z., Cai, D., et al. (2024). Clinical cure induced by pegylated interferon α -2b in the advantaged population of chronic hepatitis b virus infection: a retrospective cohort study. *Front. Cell. Infect. Microbiol.* 13, 1332232. doi:10.3389/fcimb.2023.1332232
- Xue, W., Zhang, G., Yang, C., Tan, T., Zhang, W., and Chen, H. (2025). Machine learning based immune evasion signature for predicting the prognosis and immunotherapy benefit in stomach adenocarcinoma. *Front. Cell. Dev. Biol.* 13, 1656367. doi:10.3389/fcell.2025.1656367
- Ye, Y., Lin, Y., Sun, F., Yang, W., Zhou, L., Lin, C., et al. (2024). A predictive model for functional cure in chronic hbv patients treated with pegylated interferon α : a comparative study of multiple algorithms based on clinical data. *Virol. J.* 21 (1), 333. doi:10.1186/s12985-024-02599-1
- Yuen, M., Lim, Y., Yoon, K. T., Lim, T., Heo, J., Tangkijvanich, P., et al. (2024). Vir-2218 (lebsiran) plus pegylated interferon- α -2a in participants with chronic hepatitis b virus infection: a phase 2 study. *Lancet Gastroenterology and Hepatology* 9 (12), 1121–1132. doi:10.1016/S2468-1253(24)00237-1