Using machine learning models to predict HBeAg seroconversion in CHB patients receiving pegylated interferon-α monotherapy

Hongyan Shang1,2 | Yuhai Hu3 | Hongyan Guo4 | Ruimin Lai5 | Ya Fu1,2 | Siyi Xu1,2 | Yongbin Zeng1,2 | Zhen Xun1,2 | Can Liu1,2 | Wennan Wu1,2 | Jianhui Guo1,2 | Qishui Ou1,2 | Tianbin Chen1,2

1Department of Laboratory Medicine, Gene Diagnosis Research Center, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China
2Fujian Key Laboratory of Laboratory Medicine, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China
3Department of Hepatopancreatobiliary Surgery, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China
4The School of Public Health, Fujian Medical University, Fuzhou, China
5Department of the Center of Liver Diseases, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China

Correspondence
Qishui Ou and Tianbin Chen, The First Affiliated Hospital, Fujian Medical University, NO.20 Chazhong Road, Fuzhou 350005, China. Email: ouqishui@fjmu.edu and nihaocht@gmail.com

Funding information
Education Scientific Research Projects for Middle-aged Young Teachers from the Education Department Fujian Province, Grant/Award Number: JAT200130; Fujian Provincial Natural Science Foundation General Project, Grant/Award Number: 2020J01955; National Natural Science Foundation of China, Grant/Award Number: 81702073, 81971996, 82030063 and 82172338; Open Research Fund of Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Grant/Award Number: FMUGCIC-202103

Abstract
Background and objective: Though there are many advantages of pegylated interferon-α (PegIFN-α) treatment to chronic hepatitis B (CHB) patients, the response rate of PegIFN-α is only 30~40%. Therefore, it is important to explore predictors at baseline and establish models to improve the response rate of PegIFN-α.

Methods: We randomly divided 260 HBeAg-positive CHB patients who were not previously treated and received PegIFN-α monotherapy (180 μg/week) into a training dataset (70%) and testing dataset (30%). The intersect features were extracted from 50 routine laboratory variables using the recursive feature elimination method algorithm, Boruta algorithm, and Least Absolute Shrinkage and Selection Operator Regression algorithm in the training dataset. After that, based on the intersect features, eight machine learning models including Logistic Regression, k-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Naïve Bayes were applied to evaluate HBeAg seroconversion in HBeAg-positive CHB patients receiving PegIFN-α monotherapy in the training dataset and testing dataset.

Results: XGBoost model showed the best performance, which had largest AUROC (0.900, 95% CI: 0.85~0.95 and 0.910, 95% CI: 0.84~0.98, in training dataset and testing dataset, respectively), and the best calibration curve performance to predict HBeAg seroconversion. The importance of XGBoost model indicated that treatment time contributed greatest to HBeAg seroconversion, followed by HBV DNA(log), HBeAg, HBeAb, HBcAb, ALT, triglyceride, and ALP.

Conclusions: XGBoost model based on common laboratory variables had good performance in predicting HBeAg seroconversion in HBeAg-positive CHB patients receiving PegIFN-α monotherapy.
INTRODUCTION

Up till now, the rate of hepatitis B virus (HBV) infection is still high around the world and it is estimated that approximately 70 million HBsAg carriers (5–6% prevalence) in China. After acute infection, a large number of patients gradually progress into chronic HBV infection (CHB), which often leads to liver cirrhosis and hepatocellular carcinoma (HCC). At present, there are two main drugs for CHB patients including pegylated interferon-α (PegIFN-α) and nucleotide(s)ide analogs (NAs). Compared with NAs treatment, the PegIFN-α therapy has more advantages due to exerting antiviral effects without drug resistance and long-term immunomodulatory effects within a finite duration treatment. However, the therapeutic efficacy of PegIFN-α remains limited due to its low response rate and side effects such as Influenza-like syndrome and transient peripheral cytopenia. In addition, the PegIFN-α therapy compliance is poor because of the uncomfortable reactions or high medical expenses. As a result, when HBeAg-positive CHB patients are treated with PegIFN-α for 48 weeks (180μg/week), the HBeAg seroconversion rate is less than 35% at week 24 later after drug withdrawal. Therefore, it is important to explore predictors to improve the therapeutic efficacy of PegIFN-α in HBeAg-positive CHB patients.

Several predictors have been identified to predict PegIFN-α treatment response in HBeAg-positive CHB patients: (1) HBV DNA levels less than 2×10^6IU/ml; (2) high ALT level (above 2–5 times ULN); (3) Hepatitis B genotype A or B; (4) higher baseline anti-HBc and low baseline HBsAg level; (5) necroinflammatory score of liver biopsy above G2. However, in the real clinical world, patients would not have all the predictors results, especially the necroinflammatory score of liver biopsy, the predicting models based on the above predictors are seldom. Therefore, developing a predicting model based on common clinical or laboratory baseline variables, which are easier retrieved in the health or LIS database is urgent and necessary to improve the response rate of PegIFN-α in HBeAg-positive CHB patients.

Currently, machine learning has been an important branch of artificial intelligence and is widely applied in clinical research. As a powerful classification method, machine learning has successfully applied to make accurate diagnoses and predictions from the high-dimensional, nonlinear, correlated, and/or imbalanced clinical data. Indeed, it has been proved that compared with conventional methods, like Logistic regression or Cox proportional hazard regression, machine learning has better performance in a variety of clinical scenarios. However, up till now, there are few machine learning models have been proposed to predict the response for HBeAg-positive CHB patients treated by PegIFN-α. Therefore, in this study, we intended to dig out a machine learning model based on laboratory variables to well predict the response to PegIFN-α monotherapy in HBeAg-positive CHB patients.

METHODS

Patients and basic data

In total, 357 HBeAg-positive CHB patients who were untreated previously and underwent PegIFN-α monotherapy (180μg/week) were enrolled at the Liver Diseases Center of the First Affiliated Hospital of Fujian Medical University from January 2008 to December 2018. The inclusion criteria in this study were as follows: (1) the presence of serum HBsAg for 6 or more months; (2) age between 16 and 60 years old; (3) a previous lack of antiviral treatment; (4) the PegIFN-α monotherapy time greater than 28 days (4 weeks). The exclusion criteria were as follows: combination with Hepatitis A, C, or D infection or HIV, a history of liver cirrhosis, hyperthyroidism, thyroiditis, autoimmune hepatitis, pregnancy, or any kind of tumors. Finally, 260 HBeAg-positive CHB patients were included and further analyzed in this study. Among these patients, there were 177 (68.07%) males and 83 (31.93%) females and the patients’ average age was 27.88±5.93 years old and body mass index (BMI) was 22.08±2.01. According to the EASL treatment guideline, patients who had HBeAg loss and HBeAg seroconversion at the terminal time of PegIFN-α therapy were defined as the response in this study.

Inform consent was obtained from each patient for their data to be used for research purposes, and all private information of the included patients was erased. This single-center retrospective study was approved by the Research Ethics Committee of the First Affiliated Hospital of Fujian Medical University and followed the principles of the Declaration of Helsinki.

Serological and biochemical assays

The baseline HBsAg, HBeAg, and HbcAb levels were tested by Chemiluminescence Microparticle Immuno Assay, and HBV DNA content was determined by the real-time PCR method. Baseline serum biomarkers of liver, kidney, and lipid function were measured by biochemical methods using a chemical analyzer (Siemens ADVIA 2400, Germany). Baseline routine complete blood cell counts and parameters were measured by an automated analyzer (Siemens ADVIA 2120, Germany) using EDTA-2K anticoagulated blood.

Since not all patients had liver biopsies to assess the severity of necroinflammatory score, APRI (aspartate aminotransferase to platelet ratio index), FIB-4 index, and GPR (glutamyl transpeptidase
to platelet ratio) were adopted to assess the severity of fibrosis in this retrospective study.¹⁶,¹⁷

2.3 | Performance of machine learning algorithms

2.3.1 | Data collection

We extracted all CHB patients’ data generated in the daily clinical practice from June 2008 to December 2018 from the laboratory information management system (LIS). In the data cleaning process, variables that had missing values of more than ninety percent were removed. Categorical variables that had a single category made up more than ninety percent were removed and continuous variables that had a small standard deviation (nearly constants) or had a coefficient of variation (CV) < 0.1 were also excluded from this study. For the continuous variable, the missing values were imputed by the mean values while the missing values of the categorical variable were imputed by the most frequently occurring value. The collected data were randomly divided into 70% of training datasets that were used for developing machine learning models and 30% of the testing datasets that were used for validating and comparing the performance of the models developed by the training dataset.

2.3.2 | Feature selection

A total of 50 routine medical and laboratory variables were extracted from the LIS system and the important features were selected by using the recursive feature elimination method (RFE) algorithm, Boruta algorithm, and least absolute shrinkage and selection operator regression (LASSO regression) algorithm in the training dataset using the caret, Boruta and glmnet R package, respectively. RFE algorithm implemented a backward selection of features based on feature importance ranking which to find a subset of features that could be used to produce an accurate model. Boruta algorithm was a feature ranking and selection algorithm based on random forest and the advantage of the Boruta algorithm was that it decided whether a variable was important or not. LASSO regression algorithm was applied to minimize the potential collinearity of variables and avoid overfitting of variables. The intersection features of the RFE, Boruta, and LASSO regression algorithms were considered as the most important variables by using VennDiagram R package.

2.3.3 | Model development and validation

With the intersection variables, eight common machine learning methods including Logistic Regression (LN), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GBM), Extrem Gradient Boosting (XGBoost), and Naïve Bayes (NB) were performed and compared to find the best prediction model using Caret R package. To avoid overfitting, the optimal hyperparameters for machine learning models were tuned by random search and 10-fold cross-validation with five repeats. After finding the optimal hyperparameters, prediction models are trained using the training dataset to predict HBeAg seroconversion. Further, we evaluated the predictive ability of each machine learning model, with the same hyperparameters, in testing datasets where the area under the receiver operating characteristic (AUROC) value, and the corresponding sensitivity, specificity, and F1 score of each machine learning model were all calculated.¹⁸

MLevel, an evaluation package for R, was used to establish AUROC curves and calibration curves. ROC curves were plotted to describe the variance in numbers of correctly classified abnormal cases and those of incorrectly classified normal cases as an abnormality. A calibration curve that compared the predicted probabilities placed into bins with the actual probabilities was important in determining whether a model produced sensible probability estimates. Therefore, the AUROC and calibration curve were both established to comprehensively evaluate the machine learning model prediction ability.

3 | STATISTICAL ANALYSIS

The normal distribution variables were expressed as the mean and standard deviation and compared using the Student’s t-test while the non-normal distribution variables were expressed as median and interquartile and compared by Mann–Whitney U-test. The categorical variable was described in terms of frequency and compared using the Fisher’s test. All statistical analyses were performed by using R software, version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). The R packages “caret,” “Boruta,” “glmnet,” “pROC,” “VennDiagram,” and “MLevel” were used for performed machine learning. The level of statistical significance was set at p < 0.05 with two-sided.

4 | RESULTS

4.1 | Characteristics of the subjects

In total, there were 260 HBeAg-positive CHB patients whose ages were between 16 and 60 years old and whose treatment time was more than 4 weeks were selected from 2008 to 2018 (Figure 1), and the summary of patients’ baseline characteristics was presented in Table 1. At the endpoint time of PegIFN-α monotherapy, 44 (16.92%) patients who had HBeAg seroconversion were defined as the response. As shown in Table 1, response patients had lower HBeAg baseline levels [410.33 [25.74−753.99] vs 845.97 [320.24−1238.05], p < 0.001] and HBeAb baseline levels [16.74 [1.16−40.19] vs 35.77 [10.59−51.03], p < 0.001]. Meanwhile, it is found that response patients were more likely to be younger (25.89 ± 5.43 vs 28.29 ± 5.96, p = 0.014), had lower TCHO (4.16 ± 0.79 vs 4.5 ± 0.79, p = 0.01), Triglyceride(TG) (0.83 ± 0.24 vs 1.05 ± 0.47, p = 0.003), CREA
(63.84 ± 13.07 vs 68.81 ± 12.97, p = 0.022), UA (316.03 ± 57.58 vs 339.82 ± 68.62, p = 0.032), and APOB (0.74 ± 0.18 vs 0.83 ± 0.18, p = 0.005) baseline levels while had longer PegIFN-α treatment time (42.57 ± 19.83 vs 24.98 ± 17.59, p = 0.001).

A total of 260 patients were randomly assigned to the training dataset (70%) and testing dataset (30%). Differences in the demographic characteristics and laboratory measurements between the training dataset and testing dataset were listed in Table S1, which were not significant implying that the segmentation process was random and balanced.

Three feature selecting algorithms including RFE, Boruta, and LASSO regression algorithms were used in the training dataset to choose the best variables, respectively. As shown in Figure 2A, the RFE algorithm using an accuracy measure of importance by Repeated Cross-Validation with a repeated 10-Fold cross-validation and 5 repeats indicated that it was possible to obtain high Accuracy (0.8875) using 12 variables (Treatment time, HBeAg, TG, HBeAb, ALT, HBV DNA(log), AST, CHE, APRI, CHE, ALP, and APOB).

The importance of variables calculated by the Boruta algorithm was shown in Figure 2B and 13 variables (Treatment time, HBeAg, HBeAb, TG, ALT, HBV DNA(log), AST, APRI, CHE, ALP, ALB, and HBeAb) in green columns or yellow were “confirmed” important variables while the ones in red were not.

As shown in Figure 2C,D, 50 variables were included in the LASSO regression algorithm analysis to avoid overfitting problems in the training dataset and 18 variables (Age, ALB, ALT, ALP, LDH, BUN, UA, GLU, APOB, TG, TCHO, LYMPH, HBeAg, HBeAb, HBeAb, GPR, Treatment time and HBV DNA(log)) what LASSO considered important were selected as the lambda with the lowest mean squared error.

As shown in Figure 2E, the intersection variables (8 variables: Treatment time, HBV DNA(log), HBeAg, HBeAb, TG, ALP, ALT, and HBcAb) of the RFE, Boruta, and LASSO regression algorithms were considered as the optimal features to construct machine learning models further.

## 4.2 Models for predicting HBeAg seroconversion in HBeAg-positive CHB patients receiving PegIFN-α monotherapy

In this study, eight machine learning models including Logistic Regression (LN), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GBM), Extreme Gradient Boosting (XGBoost), and Naïve Bayes (NB) were performed to predict HBeAg seroconversion in HBeAg-positive CHB patients receiving PEG-IFN monotherapy. Sensitivity, specificity, F1 score, AUROC, and calibration curve, which were calculated by 10-fold cross-validation with five repeats in the training dataset and testing dataset were used to evaluate the efficacy of each model.

The optimal hyperparameters, AUROC curves, and calibration curves of the final models in the training dataset were presented in Table S2, Figure 3A and Figure 4A and XGBoost model had the highest F1 score (0.94), AUROC (0.900, 95% CI: 0.85–0.95) and best calibration curve.

The performances of eight models predicting HBeAg seroconversion in the testing dataset, which was based on the optimal hyperparameters from the training dataset were exhibited in Table 2. The AUROC and calibration curves of each model were depicted in Figures 3B and 4B. It is also found that XGBoost model showed high sensitivity (0.894), specificity (0.81), F1 score (0.91), largest AUROC (0.910, 95% CI: 0.84–0.98), and best calibration curve while
|                                | Response (n = 44) | Nonresponse (n = 216) | p-value |
|--------------------------------|------------------|-----------------------|---------|
| Sex (male)                     | 27/44            | 150/216               | 0.384   |
| Age (years)                    | 25.89 ± 5.43     | 28.29 ± 5.96          | 0.014   |
| BMI (kg/m²)                    | 21.46 ± 3.56     | 22.62 ± 2.90          | 0.124   |
| Smoke (yes)                    | 3/44             | 18/216                | 0.973   |
| Drink (yes)                    | 3/44             | 19/216                | 0.894   |
| Treatment time (weeks)         | 42.57 ± 19.83    | 24.98 ± 17.59         | <0.001  |
| TBIL (μmol/L)                  | 18.27 ± 7.74     | 16.58 ± 8.48          | 0.222   |
| DBIL (μmol/L)                  | 6.81 ± 3.55      | 6.20 ± 5.10           | 0.446   |
| TP (g/L)                       | 71.56 ± 5.56     | 72.59 ± 4.81          | 0.208   |
| ALB (g/L)                      | 42.18 ± 3.37     | 43.11 ± 3.10          | 0.073   |
| ALT (U/L)                      | 304.73 (132.20–425.50) | 264.51 (99.75–314.50) | 0.368   |
| AST (U/L)                      | 5.39 (84.50–200.50) | 128.60 (58.0–155.50)  | 0.192   |
| GGT (U/L)                      | 71.18 (44.00–94.50) | 69.57 (29.00–85.00)   | 0.863   |
| ALP (U/L)                      | 79.91 ± 16.22    | 82.30 ± 25.63         | 0.553   |
| LDH (U/L)                      | 210.30 ± 61.75   | 201.00 ± 69.77        | 0.413   |
| CHE (U/L)                      | 7321.11 ± 2057.10 | 7841.4 ± 1881.36     | 0.101   |
| TBA (μmol/L)                   | 20.57 (6.30–20.70) | 19.34 (6.15–21.70)   | 0.795   |
| TCHO (mmol/L)                  | 4.16 ± 0.79      | 4.5 ± 0.79            | 0.01    |
| TG (mmol/L)                    | 0.83 ± 0.24      | 1.05 ± 0.47           | 0.003   |
| BUN (mmol/L)                   | 4.00 ± 0.86      | 4.14 ± 0.88           | 0.333   |
| CREA (μmol/L)                  | 63.84 ± 13.07    | 68.8 ± 12.97          | 0.022   |
| UA (μmol/L)                    | 316.03 ± 57.58   | 339.82 ± 68.62        | 0.032   |
| HDL (mmol/L)                   | 1.46 ± 0.31      | 1.45 ± 0.53           | 0.887   |
| LDL (mmol/L)                   | 2.39 ± 0.67      | 2.58 ± 0.62           | 0.068   |
| APOA1 (g/L)                    | 1.30 ± 0.25      | 1.31 ± 0.22           | 0.706   |
| APOB (g/L)                     | 0.74 ± 0.18      | 0.83 ± 0.18           | 0.005   |
| GLU (mmol/L)                   | 4.57 ± 0.43      | 4.68 ± 0.43           | 0.119   |
| PT (s)                         | 12.99 ± 0.94     | 12.86 ± 0.90          | 0.383   |
| INR                            | 1.04 ± 0.08      | 1.03 ± 0.07           | 0.339   |
| WBC (10⁹/L)                    | 5.23 ± 1.28      | 5.32 ± 1.30           | 0.698   |
| NEU (10⁹/L)                    | 2.63 ± 0.82      | 2.68 ± 0.92           | 0.777   |
| LYMPH (10⁹/L)                  | 1.88 ± 0.55      | 1.94 ± 0.58           | 0.551   |
| MONO (10⁹/L)                   | 0.38 ± 0.10      | 0.38 ± 0.30           | 0.89    |
| RBC (10¹²/L)                   | 4.62 ± 0.46      | 4.77 ± 0.77           | 0.213   |
| Hb (g/L)                       | 143.36 ± 13.38   | 145.12 ± 15.2         | 0.476   |
| HDW (g/L)                      | 23.84 ± 2.59     | 24.45 ± 6.62          | 0.551   |
| RDW (%)                        | 13.35 ± 1.12     | 13.46 ± 1.50          | 0.660   |
| HCT (L/L)                      | 0.43 ± 0.34      | 0.56 ± 0.48           | 0.649   |
| PLT (10⁹/L)                    | 195.39 ± 51.05   | 192.11 ± 49.75        | 0.692   |
| MCV (fl)                       | 92.07 ± 4.44     | 90.73 ± 6.58          | 0.198   |
| MPV (fl)                       | 8.64 ± 0.97      | 9.03 ± 2.67           | 0.338   |
| PDW (%)                        | 52.78 ± 6.34     | 52.95 ± 7.47          | 0.887   |
| HBaAg (log₁₀IU/ml)             | 3.77 (3.48–4.31) | 3.94 (3.48–4.52)      | 0.153   |
| HBeAg (S/CO)                   | 410.33 (25.74–753.99) | 845.97 (320.24–1238.05) | <0.001  |
| HBeAb (S/CO)                   | 16.74 (1.16–40.19) | 35.77 (10.59–51.03)   | <0.001  |

(Continues)
### TABLE 1 (Continued)

|                     | Response (n = 44) | Nonresponse (n = 216) | p-value |
|---------------------|------------------|-----------------------|---------|
| HBcAb (S/CO)        | 11.97 (8.54–13.29) | 11.77 (10.20–13.37) | 0.63    |
| HBVDNA (log_{10}IU/ml) | 6.84 (6.14–7.42)  | 6.88 (6.17–7.69)     | 0.829   |
| FIB4                | 1.40 ± 0.99       | 1.35 ± 1.74           | 0.833   |
| APRI                | 2.14 ± 1.56       | 1.85 ± 2.07           | 0.379   |
| GPR                 | 0.67 ± 0.47       | 0.69 ± 0.67           | 0.833   |
| RPR                 | 0.07 ± 0.03       | 0.08 ± 0.03           | 0.728   |

Bold indicates statistical significant value (p < 0.05).

**FIGURE 2** (A) Extraction of the optimum features from the total 50 variables was performed using an accuracy (repeated cross-validation) measure of importance in the RFE algorithm. (B) The importance of variables calculated by the Boruta algorithm. The green columns or yellow tones were “confirmed” important variables while the ones in red were not. (C) Selection of the optimal parameter (lambda) in the LASSO regression. (D) LASSO regression coefficient profiles of the 50 variables in the training dataset. (E) Venn diagram of variables in the RFE, Boruta, and LASSO regression algorithm datasets in which the common area represented the overlapping variables.

**FIGURE 3** AUROC of the machine learning models to predict HBeAg seroconversion receiving PegIFN-α monotherapy in the training dataset (A) and testing dataset (B).
The Decision Tree model showed the lowest AUROC (0.72, 95% CI: 0.58–0.86).

Further, variable importance scores in the XGBoost model were evaluated by the varImp function (caret package), and the variable importance plot was visualized in Figure 5, which suggested that Treatment time contributed the greatest to HBeAg seroconversion, followed by HBV DNA (log), HBeAg, HBeAb, HBeAb, ALT, TG, and ALP.

**5 | DISCUSSION**

There has been a great number of clinical studies focusing on the relationship between the baseline variables and PegIFN-α therapeutic efficacy, which is determined by the HBV DNA and HBeAg levels at 24 weeks or 48 weeks later. However, in real clinical practice, a great number of patients would receive PegIFN-α therapy for 24 weeks or 48 weeks for personal affairs or uncomfortable side effects. What’s more, some patients have no patience to wait for the HBV DNA to decline and often switch to another doctor or hospital. As a result, the real PegIFN-α therapy process is complex and a prediction model based on the result of PegIFN-α therapy at the regulated time point such as 24 weeks or 48 weeks is limited. What’s more, the liver biopsy is invasive and the necroinflammatory score is hard to obtain. Therefore, it is necessary to find a good model to predict the effect of PegIFN-α therapy based on the real therapy time and noninvasive baseline biomarkers.
Machine learning models have been used to predict disease stages or therapeutic effects in the medical and health field for a long time.\textsuperscript{14,19,20} Therefore, in this retrospective study, the prediction performance of eight common machine learning models based on the baseline noninvasive clinical or laboratory variables are performed and our results suggest that the best prediction model is XGBoost, which has the largest AUROC (0.900, 95% CI: 0.85-0.95 and 0.910, 95% CI: 0.84-0.98, in the training dataset and testing dataset, respectively) and best calibration curves. It shows that Treatment time act as the most important factor in the XGBoost model followed by HBV DNA(log), HBeAg, HBeAb, HBCAb, ALT, TG, and ALP. Most of the variables included in the XGBoost model have been mentioned by AASLD 2018 hepatitis B guidance or EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection or previous studies.\textsuperscript{4-7,21,22} Therefore, the XGBoost model is applicable in predicting HBeAg seroconversion in HBeAg-positive CHB patients receiving PegIFN-α therapy. Meanwhile, as we evaluate the efficiency of PegIFN-α therapy according to the exact therapy time point not just based on 24 weeks or 48 weeks, the whole response rate of PegIFN-α therapy was 16.92%, which is much lower than that at 24 weeks or 48 weeks. Even though, the XGBoost model established in this study may be more practical as we can dynamically evaluate the efficiency of PegIFN-α therapy at any time (>4 weeks) not just at 24 weeks or 48 weeks routinely.

Besides, it is found that some variables such as Age, CREA, or UA, which are significantly different between the response group and nonresponse group according to the Student’s t-test are excluded in the intersection variables of the RFE, Boruta, and LASSO regression algorithms, which may attribute to the different fundamental algorithm between traditional Student’s t-test and machine learning algorithms. Therefore, more research is needed to confirm and explain our machine learning models, which are helpful for applying the machine learning modes to predict the HBeAg seroconversion in clinical practice.

Though new insights on predicting HBeAg seroconversion using machine learning models have been shown in this study, some limitations should not be neglected including the small sample sizes and lack of multicenter validation data or longitudinal follow-up data. What's more, some variables like hepatitis B genotypes, which are important to predicting HBeAg seroconversion are not included in this study due to missing in most patients. Additionally, there are only eight common machine learning models analyzed in this study, therefore, more machine learning algorithms such as deep learning algorithms should be considered to improve the predicting accuracy in future research.

In conclusion, our study indicates that the XGBoost model with the largest AUROC and best calibration curve can be used to predict HBeAg seroconversion in the real PegIFN-α therapy process. To the best of our knowledge, we firstly utilize machine learning models based on the laboratory variables to predict HBeAg seroconversion receiving PegIFN-α monotherapy in HBeAg-positive CHB patients and what we found in this study may be helpful for the hepatologists to identify the suitable patients undergoing PegIFN-α monotherapy and dynamically evaluate the efficiency of PegIFN-α based on the noninvasive and common laboratory variables.

**AUTHOR CONTRIBUTIONS**
Hongyan Shang and Yuhai Hu contributed equally to this work; Hongyan Shang and Yuhai Hu conceived and designed the study; Hongyan Guo, Ruimin Lai, Yu Fu, Siyi Xu, and Yongbin Zen contributed to the statistical analysis; Zhen Xun, Can Liu, Wennan Wu and Jianhui Guo collected clinical data of the patients; Qishui Ou and Tianbin Chen wrote the article; all authors read and approved the final version of the article.

**FUNDING INFORMATION**
This study was supported by the National Natural Science Foundation of China (grant numbers: 82030063, 81702073, 81971996, 82172338), Fujian Provincial Natural Science Foundation General Project (2020J01955), Open Research Fund of Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University (FMUGIC-202103), Education Scientific Research Projects for Middle-aged Young Teachers from the Education Department Fujian Province (No.JAT200130).

**CONFLICT OF INTEREST**
The authors declared no conflict of interest.

**DATA AVAILABILITY STATEMENT**
The datasets and supporting conclusions of this article are available on reasonable request.

**ORCID**
Qishui Ou https://orcid.org/0000-0002-9923-3212

**REFERENCES**

1. Liu J, Liang W, Jing W, Liu M. Countdown to 2030: eliminating hepatitis B disease, China. Bull World Health Organ. 2019;97(3):230-238.
2. Fu Y, Zeng Y, Chen T, et al. Characterization and clinical significance of natural variability in hepatitis B virus reverse transcriptase in treatment-naive Chinese patients by sanger sequencing and next-generation sequencing. J Clin Microbiol. 2019;57(8):e00119-19.
3. Hou J, Wang G, Wang F, et al. Guideline of prevention and treatment for chronic hepatitis B (2015 update). J Clin Transl Hepatol. 2017;5(4):297-318.
4. Terrault NA, Lok ASF, McMahon BJ, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology. 2018;67(4):1560-1599.
5. Sun J, Ma H, Xie Q, et al. Response-guided peginterferon therapy in patients with HBeAg-positive chronic hepatitis B: a randomized controlled study. J Hepatol. 2016;65(4):674-682.
6. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the Liver. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. J Hepatol. 2017;67(2):370-398.
7. Terrault NA, Bzowej NH, Chang K-M, Hwang JP, Jonas MM, Murad MH. AASLD guidelines for treatment of chronic hepatitis B. Hepatology. 2016;63(1):261-283.
8. Tian X, Chong Y, Huang Y, et al. Using machine learning algorithms to predict hepatitis B surface antigen Seroclearance. *Comput Math Methods Med*. 2019;2019:1-7.

9. Wilkes EH, Rumsby G, Woodward GM. Using machine learning to aid the interpretation of urine steroid profiles. *Clin Chem*. 2018;64(11):1586-1595.

10. Hernandez-Suarez DF, Kim Y, Villablanca P, et al. Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *J Am Coll Cardiol Intv*. 2019;12(14):1328-1338.

11. Wang Y, Du Z, Lawrence WR, Huang Y, Deng Y, Hao Y. Predicting hepatitis B virus infection based on health examination data of community population. *Int J Environ Res Public Health*. 2019;16(23):4842.

12. Eaton JE, Vesterhus M, McCauley BM, et al. Primary sclerosing cholangitis risk estimate tool (PREsTo) predicts outcomes of the disease: a derivation and validation study using machine learning. *Hepatology*. 2018;71:214-224.

13. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett*. 2020;471:61-71.

14. Bhattacharai S, Klimov S, Aleskandarany MA, et al. Machine learning-based prediction of breast cancer growth rate in vivo. *Br J Cancer*. 2019;121(6):497-504.

15. Lee HC, Yoon SB, Yang SM, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. *J Clin Med*. 2018;7(11):428.

16. Castera L. Noninvasive methods to assess liver disease in patients with hepatitis B or C. *Gastroenterology*. 2012;142(6):1293-1302 e1294.

17. Xiao G, Yang J, Yan L. Comparison of diagnostic accuracy of aspartate aminotransferase to platelet ratio index and fibrosis-4 index for detecting liver fibrosis in adult patients with chronic hepatitis B virus infection: a systemic review and meta-analysis. *Hepatology*. 2015;61(1):292-302.

18. Yip TC, Ma AJ, Wong VW, et al. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther*. 2017;46(4):447-456.

19. ElHefnawi M, Abdalla M, Ahmed S, et al. Accurate prediction of response to interferon-based therapy in Egyptian patients with chronic hepatitis C using machine-learning approaches. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Institute of Electrical and Electronics Engineers; 2012:771-778.

20. Xie G, Wang X, Wei R, et al. Serum metabolite profiles are associated with the presence of advanced liver fibrosis in Chinese patients with chronic hepatitis B viral infection. *BMC Med*. 2020;18(1):144.

21. Brunetto MR, Oliveri F, Coco B, et al. Outcome of anti-HBe positive chronic hepatitis B in alpha-interferon treated and untreated patients: a long term cohort study. *J Hepatol*. 2002;36(2):263-270.

22. Sun Z, Lin JP, Liu C, et al. Association of serum total cholesterol with pegylated interferon-alpha treatment in HBeAg-positive chronic hepatitis B patients. *Antivir Ther*. 2019;24(2):85-93.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Shang H, Hu Y, Guo H, et al. Using machine learning models to predict HBeAg seroconversion in CHB patients receiving pegylated interferon-α monotherapy. *J Clin Lab Anal*. 2022;36:e24667. doi: [10.1002/jcla.24667](https://doi.org/10.1002/jcla.24667)