



ORIGINAL ARTICLE OPEN ACCESS

AI-Based Platelet-Independent Noninvasive Test for Liver Fibrosis in MASLD Patients

Shun-ichi Wakabayashi¹ | Takefumi Kimura^{1,2}  | Nobuharu Tamaki³ | Takanobu Iwadare¹ | Taiki Okumura¹ | Hiroyuki Kobayashi¹ | Yuki Yamashita¹  | Naoki Tanaka^{4,5,6} | Masayuki Kurosaki³ | Takeji Umemura^{1,2}

¹Department of Medicine, Division of Gastroenterology, Shinshu University School of Medicine, Matsumoto, Japan | ²Consultation Center for Liver Diseases, Shinshu University Hospital, Matsumoto, Japan | ³Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan | ⁴Department of Global Medical Research Promotion, Shinshu University Graduate School of Medicine, Matsumoto, Japan | ⁵International Relations Office, Shinshu University School of Medicine, Matsumoto, Japan | ⁶Research Center for Social Systems, Shinshu University, Matsumoto, Japan

Correspondence: Takefumi Kimura (kimuratakefumii@yahoo.co.jp; t_kimura@shinshu-u.ac.jp)

Received: 6 September 2024 | **Revised:** 14 March 2025 | **Accepted:** 21 March 2025

Funding: This work was supported by AMED under grant number JP23fk0210125 and JP24fk0210125, and by JSPS KAKENHI (grant numbers JP22K20884 and JP24K11087). Nobuharu Tamaki and Masayuki Kurosaki receives funding support from Japan Agency for Medical Research and Development (grant numbers: JP23fk0210111h0002, JP23fk0210104s0202, JP23fk0210123h0001) and Japanese Ministry of Health, Welfare and Labor (grant numbers: 23HC2003, 23HC2002).

Keywords: AI | liver fibrosis | MASH | MASLD | noninvasive test | platelet-independent

ABSTRACT

Background and Aim: Noninvasive tests (NITs), such as platelet-based indices and ultrasound/MRI elastography, are widely used to assess liver fibrosis in metabolic dysfunction-associated steatotic liver disease (MASLD). However, platelet counts are not routinely included in Japanese health check-ups, limiting their utility in large-scale screenings. Additionally, elastography, while effective, is costly and less accessible in routine practice. Most existing AI-based models incorporate these markers, restricting their applicability. This study aimed to develop a simple yet accurate AI model for liver fibrosis staging using only routine demographic and biochemical markers.

Methods: This retrospective study analyzed biopsy-proven data from 463 Japanese MASLD patients. Patients were randomly assigned to training ($N=370$, 80%) and test ($N=93$, 20%) cohorts. The AI model incorporated age, sex, BMI, diabetes, hypertension, hyperlipidemia, and routine blood markers (AST, ALT, γ -GTP, HbA1c, glucose, triglycerides, cholesterol).

Results: The Support Vector Machine model demonstrated high diagnostic performance, with an area under the curve (AUC) of 0.886 for detecting significant fibrosis ($\geq F2$). The AUCs for advanced fibrosis ($\geq F3$) and cirrhosis (F4) were 0.882 and 0.916, respectively. Compared to FIB-4, APRI, and FAST score (0.80–0.96), SVM achieved comparable accuracy while eliminating the need for platelet count or elastography.

Conclusion: This AI model accurately assesses liver fibrosis in MASLD patients without requiring platelet count or elastography. Its simplicity, cost-effectiveness, and strong diagnostic performance make it well-suited for large-scale health screenings and routine clinical use.

Abbreviations: γ -GTP, gamma-glutamyl transferase; AI, artificial intelligence; ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio Index; AST, aspartate aminotransferase; AUC, area under the curve; BMI, body mass index; DM, diabetes mellitus; F, fibrosis stage; FBS, fasting blood sugar; FIB-4, fibrosis 4; FPR, false positive rate; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; HL, hyperlipidemia; HT, hypertension; IQR, interquartile range; KNN imputer, k-nearest neighbors imputer; LDL-C, low-density lipoprotein cholesterol; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatosis liver disease; n.s., not significant; NITs, noninvasive tests; ROC, receiver operating characteristic; TG, triglyceride; TPR, true positive rate; US, ultrasound.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *JGH Open* published by Journal of Gastroenterology and Hepatology Foundation and John Wiley & Sons Australia, Ltd.

1 | Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) has emerged as one of the most common chronic liver diseases worldwide [1]. MASLD is a heterogeneous condition that is strongly associated with diabetes mellitus (DM), hypertension (HT), hyperlipidemia (HL), obesity, and metabolic syndrome [1]. Although ethnic differences in the incidence of MASLD have been reported, this condition is found in approximately 30% of the world's population and is expected to remain the most prevalent form of the disease globally [2]. Metabolic dysfunction-associated steatohepatitis (MASH) includes inflammation/fibrosis and is a more progressed form of MASLD. In MASLD patients, chronic inflammation leads to liver fibrosis, which may progress to cirrhosis and hepatocellular carcinoma [3]. Beyond liver-related complications, MASLD is also associated with an increased risk of cardiovascular disease and extrahepatic malignancies [4, 5].

Liver fibrosis is the strongest predictor of liver-related and cardiovascular events, underscoring the need for early detection and timely intervention [6–8]. Although liver biopsy remains the gold standard for fibrosis assessment, its invasive nature and associated risks limit its widespread use. Consequently, noninvasive tests (NITs) such as FIB-4 and APRI have become widely adopted, alongside imaging-based methods like ultrasound (US) and MRI elastography [9, 10].

In Japan, the nationwide health check-up program (Tokutei-Kenshin) plays a crucial role in preventive medicine, targeting individuals aged 40 and older to detect metabolic diseases and associated risk factors [11]. In Fiscal Year 2021 alone, 30.39 million people underwent Tokutei-Kenshin, representing approximately a quarter of the Japanese population [12, 13]. Despite its large-scale implementation, the program lacks an effective liver fibrosis screening component. One major limitation is the absence of platelet count in routine health check-ups, making conventional platelet-based NITs such as FIB-4 and APRI impractical. While alternative biomarkers like type 4 collagen 7S, M2BPGi, autotaxin, and ELF scores, as well as elastography, have been proposed, their high cost limits their feasibility for large-scale screenings [14–17].

Recent advancements in artificial intelligence (AI) have facilitated the development of diagnostic models for MASLD [18–20]. However, most existing AI models incorporate platelet count or elastography, limiting their applicability in population-based screenings where these parameters are not routinely available. To address this gap, we developed an AI-based model that predicts fibrosis solely using routinely available demographic and biochemical markers, ensuring high diagnostic accuracy while enhancing accessibility and scalability.

2 | Methods

2.1 | Patients

This retrospective, cross-sectional study was approved by the Committee for Medical Ethics of Shinshu University School of Medicine (ID number: 2802) and was performed following the

Helsinki declaration of 1975, 1983 revision. Informed consent was obtained from all patients. We enrolled 463 biopsy-proven Japanese MASLD patients who were admitted to Shinshu University Hospital (Matsumoto, Japan) between 2003 and 2022 [21]. Other causes of liver disease were ruled out, including alcohol intake (>20 g/day), viral hepatitis, drug-induced liver injury, autoimmune liver disease, Wilson's disease, hereditary hemochromatosis, and citrine deficiency [22, 23]. Patients were considered to have DM with a fasting glucose level of ≥ 126 mg/dL or hemoglobin A1c (HbA1c) level of $\geq 6.5\%$, or if they were taking insulin or oral hypoglycemic agents [24]. Patients were considered to have HT if their systolic/diastolic pressure was $>140/90$ mmHg or if they were taking anti-hypertensive drugs [25]. Patients were judged as having HL if their fasting serum levels of total cholesterol, low-density lipoprotein cholesterol (LDL-C), or triglyceride (TG) were ≥ 220 mg/dL, ≥ 140 mg/dL, or ≥ 150 mg/dL, respectively, or if they were taking lipid-lowering drugs [26]. Body weight and height were measured before liver biopsy in an overnight fasting condition. Biochemical tests were obtained in an overnight fasting state on the day of liver biopsy, including liver tests, serum lipids, fasting blood sugar (FBS), and HbA1c. APRI was calculated as $100 \times (\text{aspartate aminotransferase [AST]}/\text{upper normal limit of AST [U/L]})/\text{platelet count} [\times 10^9/\text{L}]$. FIB-4 index was determined as $(\text{age [years]} \times \text{AST [U/L]})/(\text{platelet count} [\times 10^9/\text{L}] \times \sqrt{\text{alanine aminotransferase [ALT] [U/L]}})$ [27]. In this study, the FibroScan-AST (FAST) score was applied to 48 patients. The FAST score was calculated using the formula: $\text{FAST} = (0.395 \times \text{liver stiffness measurement [LSM]}) + (0.025 \times \text{controlled attenuation parameter [CAP]}) - (0.014 \times \text{AST}) - 1.264$.

2.2 | Liver Biopsy and Histological Assessment

Liver specimens of at least 1.5 cm in length were obtained from segment 5 or 8 using 14-gauge needles, as described previously, and immediately fixed in 10% neutral formalin. Sections of $4 \mu\text{m}$ in thickness were cut and stained using the hematoxylin and eosin and Azan-Mallory methods. The histological activity of MASLD was assessed by an independent expert pathologist in a blinded manner according to the non-alcoholic fatty liver disease scoring system proposed by Kleiner et al. [28]. Steatosis was graded as 1 to 3 based on the rate of steatotic hepatocytes (5%–33%, $>33\%$ –66%, and $>66\%$, respectively). Lobular inflammation was graded as 0 to 3 based on the overall assessment of all inflammatory foci (no foci, <2 foci/200 \times field, 2–4 foci/200 \times field, and >4 foci/200 \times field, respectively). Ballooning grade was scored as 0 to 2 by the frequency of ballooned hepatocytes (none, few, and many, respectively). Fibrosis stage was scored as follows: F0, none; F1, perisinusoidal or periportal; F2, perisinusoidal and portal/periportal; F3, bridging fibrosis; and F4, cirrhosis. Significant fibrosis was defined as \geq F2 [29]. Patients with uncorrectable coagulopathy, severe thrombocytopenia, massive ascites, or an inability to cooperate during the procedure were excluded from liver biopsy.

2.3 | Conventional Statistical Techniques

Continuous variables are expressed as the median and interquartile range and compared with Student's *t*-test or the

Mann–Whitney U test, as appropriate. The chi-squared test was used for the comparison of categorical variables. Delong's test was employed to compare the significance of receiver operating characteristic (ROC) curves. A p value of <0.05 (2-tailed test) was considered statistically significant.

2.4 | Data Analysis

Machine learning was performed using Python 3.9 software (Scipy version 1.9.3, Scikit-learn version 1.2.1) [30]. Data analysis was conducted with the R (version 4.3.1) and pROC package (version 1.18.4) [31].

3 | Machine Learning Techniques

3.1 | Dataset

Random splits were made using the API of scikit-learn, a Python machine learning package. Subject data were randomly divided into two groups, the training dataset ($N=370$) and the test dataset ($N=93$), using the stratified split method (test size = 20%) in scikit-learn to homogenize the ratio of fibrosis in both groups (Figure 1). Models were trained using data from the training cohort, and test data were used only to evaluate the final model's performance.

3.2 | Parameters

The parameters of age, sex, body mass index (BMI), presence of DM, HT, and HL, and levels of AST, ALT, γ -glutamyl transpeptidase (γ -GTP), HbA1c, TG, HDL-cholesterol (HDL-C), LDL-C, and FBS were considered. These data are routinely collected in the Japanese health checkup system.

3.3 | Training and Evaluation of Models

The models trained included Logistic regression, Support Vector Machine (SVM) [32], Random Forest [33], XGBoost [34], and

LightGBM [35]. Logistic Regression is commonly used for binary classification, modeling the relationship between inputs and the target variable to predict probabilities. Its linear nature makes it interpretable and effective for linearly separable data, serving as a baseline for complex models. SVM is a classification and regression algorithm that finds an optimal hyperplane to separate classes. Effective for high-dimensional and small sample datasets, SVM can handle non-linear data through the kernel trick, making it useful in fields like image and text classification. Random Forest is an ensemble method that combines multiple decision trees to enhance accuracy and reduce overfitting, making it robust and effective for both classification and regression tasks. Its feature importance evaluation is valuable in exploratory analysis. XGBoost, a high-performance gradient boosting algorithm, enhances prediction accuracy through boosting and regularization and is optimized for speed and sparse data, excelling in structured data tasks like classification and ranking. LightGBM, another gradient boosting framework, focuses on speed and memory efficiency, using a leaf-wise growth strategy and histogram-based decision rules, making it ideal for large datasets and high-dimensional, sparse data where resource efficiency is key.

Before model training, missing value processing was conducted with the k-nearest neighbors imputer (Figure S1) [36]. Details of the missing values in the data set are shown in Table S1. The modeling flow for this study is summarized in Figure 1. To assess the models' generalization performance, we employed the cross-validation technique. Specifically, we used 10-fold cross-validation in the training dataset and performed hyperparameter tuning [37]. In 10-fold cross-validation, the training dataset was randomly divided into 10 subsets. The models were trained using nine of these subsets (training subsets) and evaluated on the remaining subset (validation subset). This process was repeated across all 10 subsets, and the average area under the curve (AUC) over the 10 folds was used as the performance metric.

Model performance was evaluated and hyperparameters were tuned based on the average AUC values from 10-fold cross-validation. The search range for hyperparameters in each model

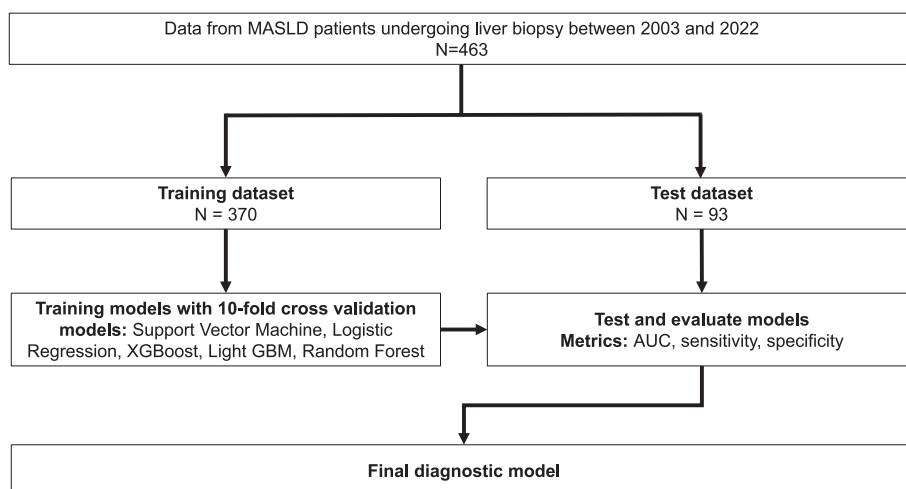


FIGURE 1 | Flow chart of this study. AUC: area under the curve, LGBM: LightGBM, LR: logistic regression, MASLD: metabolic dysfunction-associated steatosis liver disease, n.s.: not significant, RF: Random Forest, SVM: Support Vector Machine, TPR: true positive rate, XGB: XGBoost.

is as shown in Table S2. Using these optimal hyperparameters, final models for each algorithm were constructed and trained on the entire training dataset. To address the risk of overfitting, models were assessed on the training dataset to evaluate their training adequacy (model performances are shown in Figure S2), and the final performance was assessed using AUC, sensitivity, and specificity with the test dataset. The model with the best diagnostic performance was selected. This best model's performance was compared to the established indices of FIB-4 and APRI in terms of AUC, sensitivity, and specificity. To improve the interpretability of the final models, coefficients for linear models (Logistic regression and Support Vector Machine) and feature importances for tree-based models (Random Forest, XGBoost, LightGBM) were used to identify the factors contributing to the diagnosis.

4 | Results

4.1 | Patient Characteristics

Patient characteristics are summarized in Table 1. Overall, the median age was 56 years, with a female predominance (43.8% male). The median BMI of 26.6 kg/m² indicated that the patients were overweight as an Asian population [38]. The concurrence of DM was seen in 172 cases (37.1%), HT in 187 cases (40.4%), and HL in 287 cases (62.0%). F0, F1, F2, F3, and F4 were judged in 83 cases (17.9%), 198 cases (42.8%), 55 cases (11.9%), 99 cases (21.4%), and 28 cases (6.0%), respectively (Table S3). Comparisons between the significant fibrosis (\geq F2) and non-significant fibrosis

(F0-1) group are presented in Table 2 and Table S4. Age, BMI, AST, ALT, and γ -GTP were significantly higher, while LDL-C was significantly lower in the significant fibrosis group (\geq F2). Patients with significant fibrosis (\geq F2) also had a significantly higher prevalence of DM and HT as well as a significantly lower prevalence of HL.

4.2 | Significant Fibrosis Diagnostic Ability

Modeling was performed as shown in Figure 1. Figure 2 displays the significant fibrosis (\geq F2) diagnostic performance of the test dataset. The model with the highest AUC was Support Vector Machine (AUC: 0.886), followed next by Logistic regression (AUC: 0.877), XGBoost (AUC: 0.852), and Random Forest (AUC: 0.810). Although there were no significant differences between the models, the Support Vector Machine model had a relatively higher AUC (0.886), sensitivity (0.857), specificity (0.785), and negative value (0.927) than the other models (Table 3).

4.3 | Comparison With Existing NITs

Comparisons of diagnostic performance in the test dataset at each fibrosis stage among the Support Vector Machine model, FIB-4 index, and APRI are shown in Figure 3A–C and Table 4. For the diagnosis of significant fibrosis (\geq F2), the Support Vector Machine model achieved AUC values that were nearly equivalent to those of FIB-4 index and APRI, with a higher sensitivity and

TABLE 1 | Characteristics of patients.

N = 463	Median/N	IQR/%	Training dataset (N = 370)	Test dataset (N = 93)	p
Age (years)	56	42–65	56 (42–65)	52 (39–61)	0.0902
Male	203	43.8%	155 (41.9%)	48 (51.6%)	0.1159
Diabetes mellitus	172	37.1%	140 (37.8%)	32 (34.4%)	0.6229
Hypertension	187	40.4%	152 (41.1%)	35 (37.6%)	0.6260
Hyperlipidemia	287	62.0%	236 (63.8%)	51 (54.8%)	0.1418
BMI (kg/m ²)	26.6	24.0–30.3	26.8 (24.1–30.5)	26.0 (23.9–29.8)	0.5672
AST (U/L)	48	31–72	48 (31–72)	47 (31–71)	0.9209
ALT (U/L)	69	41–109	69 (41–111)	72 (47–107)	0.4452
γ -GTP (U/L)	54	37–90	54 (35–90)	55 (39–89)	0.4959
TG (mg/dL)	124	91–163	124 (90–166)	121 (95–148)	0.8511
HDL-C (mg/dL)	51	44–59	51 (44–59)	52 (45–60)	0.2479
LDL-C (mg/dL)	126	106–146	127 (106–147)	126 (100–141)	0.4553
HbA1c (%)	5.7	5.3–6.2	5.7 (5.3–6.2)	5.7 (5.2–6.0)	0.2014
FBS (mg/dL)	106	96–119	108 (97–120)	103 (93–117)	0.0307
APRI	0.6	0.4–1.0	0.6 (0.4–1.0)	0.6 (0.4–1.1)	0.9305
FIB-4 index	1.4	0.8–2.4	1.5 (0.8–2.4)	1.3 (0.8–2.3)	0.416

Abbreviations: γ -GTP: gamma-glutamyl transferase, ALT: alanine aminotransferase, APRI: aspartate aminotransferase-to-platelet ratio Index, AST: aspartate aminotransferase, BMI: body mass index, FBS: fasting blood sugar, FIB-4: fibrosis 4, HbA1c: hemoglobin A1c, HDL-C: high-density lipoprotein cholesterol, IQR: interquartile range, LDL-C: low-density lipoprotein cholesterol, TG: triglyceride.

TABLE 2 | Comparisons of patients with significant fibrosis and non-significant fibrosis grade.

	Non-significant fibrosis, F0-1 (N=281)	Significant fibrosis, ≥ F2 (N=182)	p
Age (years)	50 (34–61)	61 (54–68)	<0.0001
Male	145 (51.6%)	58 (31.9%)	<0.0001
Diabetes mellitus	85 (30.2%)	87 (47.8%)	0.0002
Hypertension	95 (33.8%)	92 (50.5%)	0.0005
Hyperlipidemia	189 (67.3%)	98 (53.8%)	0.005
BMI (kg/m ²)	25.6 (23.5–29.4)	27.9 (25.0–31.2)	<0.0001
AST (U/L)	38 (27–61)	63 (46–102)	<0.0001
ALT (U/L)	64 (36–102)	80 (51–132)	0.0003
γ-GTP (U/L)	52 (34–85)	60 (44–98)	0.0076
TG (mg/dL)	127 (93–168)	116 (87–150)	0.0585
HDL-C (mg/dL)	50 (44–59)	52 (46–59)	0.2213
LDL-C (mg/dL)	130 (112–150)	120 (98–140)	0.0012
HbA1c (%)	5.6 (5.2–6.0)	5.8 (5.4–6.5)	0.0021
FBS (mg/dL)	103 (95–115)	110 (100–127)	<0.0001
APRI	0.4 (0.3–0.7)	1.0 (0.6–1.6)	<0.0001
FIB-4 index	1.0 (0.6–1.6)	2.5 (1.7–3.7)	<0.0001

Abbreviations: γ-GTP: gamma-glutamyl transferase, ALT: alanine aminotransferase, APRI: aspartate aminotransferase-to-platelet ratio Index, AST: aspartate aminotransferase, BMI: body mass index, FBS: fasting blood sugar, FIB-4: fibrosis 4, HbA1c: hemoglobin A1c, HDL-C: high-density lipoprotein cholesterol, IQR: interquartile range, LDL-C: low-density lipoprotein cholesterol, TG: triglyceride.

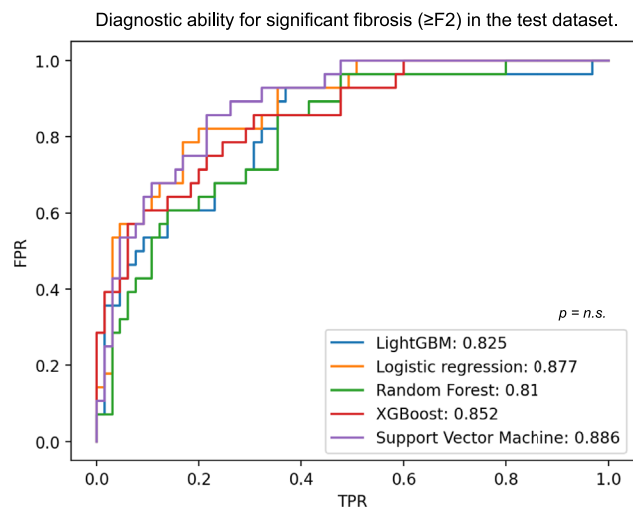


FIGURE 2 | Diagnostic ability for significant fibrosis (≥F2) in the final model. AUC curves. Support vector machine had the best AUC, followed by logistic regression and XGBoost, but no significant difference was observed. AUC: area under the curve, FPR: false positive rate, n.s.: not significant, TPR: true positive rate.

negative predictive value than the other indicators (Figure 3A, Table 4). For diagnosing ≥F3 and F4, this model yielded AUC values of 0.882 and 0.916, respectively (Figure 3B,C), which indicated high diagnostic performance on par with FIB-4 index and APRI. Additionally, Figure S3 presents a comparison of the AUC values for each fibrosis stage among the Support Vector

Machine model, FIB-4 index, APRI, and FAST score, limited to the 48 patients with available FAST score data. The results show that the Support Vector Machine model demonstrates equivalent performance to the other indices.

Figure 4 demonstrates the feature importance of the Support Vector Machine AI model. Age, BMI, AST, and the presence of DM were indicative of significant fibrosis (≥F2), whereas ALT, and the presence of HL did not. In other words, the absence of the latter variables supported the existence of significant fibrosis (≥F2). Whereas the presence of DM was supportive of significant fibrosis (≥F2), HbA1c showed an inverse trend. This could have potentially reflected a decrease in hemoglobin due to fibrosis progression.

5 | Discussion

5.1 | Main Findings

This study presents an AI-based diagnostic model for liver fibrosis, specifically designed for Japan's health check-up system by eliminating the need for platelet count and elastography. Using clinical and biochemical data obtained from routine biopsy procedures, the Support Vector Machine model demonstrated diagnostic accuracy comparable to established NITs, including the FIB-4 index, APRI, and FAST score. Key predictive variables included age, sex, BMI, DM, HT, HL, and serum markers such as AST, ALT, γ-GTP, HbA1c, TG, HDL-C, LDL-C, and FBS.

TABLE 3 | Diagnostic performance of machine learning models for identifying significant fibrosis (\geq F2) in the Test Dataset.

	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value	Negative predictive value
Light GBM	0.825 (0.714–0.908)	0.441	0.929 (0.786–1.000)	0.631 (0.523–0.769)	0.520 (0.440–0.619)	0.954 (0.867–1.000)
Logistic regression	0.877 (0.806–0.950)	0.431	0.821 (0.679–0.964)	0.800 (0.692–0.892)	0.639 (0.534–0.767)	0.912 (0.849–0.980)
Random Forest	0.810 (0.738–0.921)	0.381	0.857 (0.714–0.964)	0.646 (0.477–0.708)	0.480 (0.400–0.574)	0.907 (0.829–0.977)
XGBoost	0.852 (0.767–0.933)	0.381	0.857 (0.714–0.964)	0.692 (0.584–0.800)	0.546 (0.457–0.657)	0.918 (0.849–0.980)
Support Vector Machine	0.886 (0.795–0.941)	0.389	0.857 (0.679–0.964)	0.785 (0.677–0.892)	0.631 (0.522–0.758)	0.927 (0.845–0.980)

Abbreviations: AUC: area under the curve, CI: confidence interval.

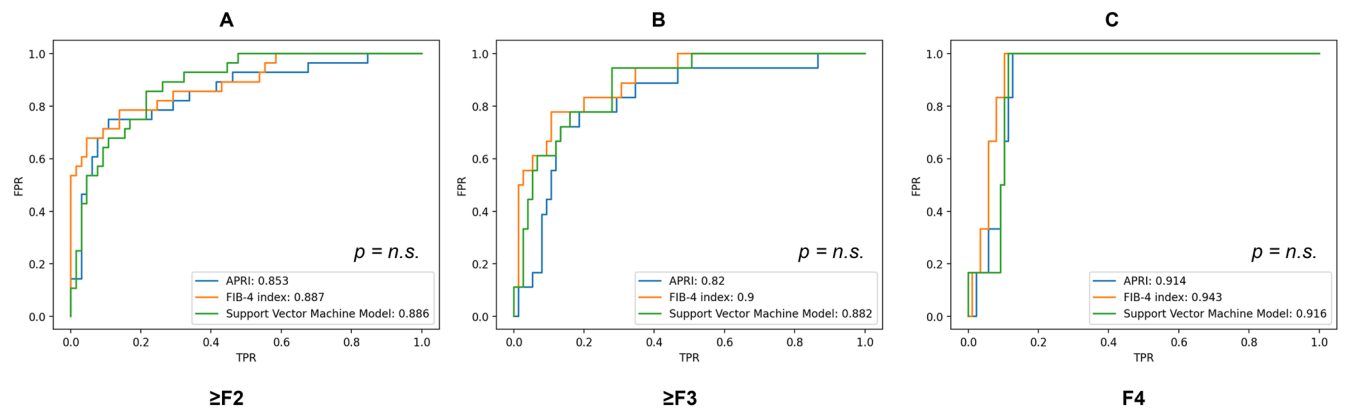


FIGURE 3 | Comparisons of diagnostic ability for fibrosis grade in the test dataset. AUC curves for diagnosing \geq F2 (A), \geq F3 (B), and F4 (C). There was no significant difference in the diagnostic performance of the FIB-4 index, Support Vector Machine, and APRI for diagnosing any stage of fibrosis. APRI: aspartate aminotransferase-to-platelet ratio Index, F: fibrosis stage, FIB-4: fibrosis 4, FPR: false positive rate, n.s.: not significant, TPR: true positive rate.

5.2 | Context With Published Literature

Liver fibrosis assessment has shifted from invasive procedures to noninvasive approaches. While liver biopsy remains the definitive method, NITs such as FIB-4 and APRI have gained clinical utility, while newer biomarkers (e.g., M2BPGI, thrombospondin 2, autotaxin) and imaging techniques (e.g., US and MRI elastography) have further improved diagnostic accuracy [14, 16, 39–45].

Recent advances in AI have further expanded the landscape of fibrosis diagnostics [20, 46–51]. Machine learning models have demonstrated high accuracy, yet most rely on platelet count and elastography, limiting their use in population-based screenings where these parameters are not routinely available [20, 46–51]. Additionally, the high cost and accessibility issues of imaging-based diagnostics restrict their feasibility for large-scale implementation.

Our model overcomes these limitations by eliminating the need for platelet count or elastography while maintaining comparable diagnostic performance. By utilizing only routinely available demographic and biochemical markers, such as age, BMI, AST,

ALT, glucose, HbA1c, and cholesterol, it is highly compatible with Japan’s health check-up system, where cost-effective and scalable diagnostic tools are essential.

Another critical aspect of AI-driven diagnostics is interpretability, which facilitates clinical acceptance [52]. Our model’s reliance on well-established metabolic risk factors—age, BMI, AST, and DM—aligns with their documented roles in fibrosis progression [27]. Conversely, the inverse relationship of ALT and HbA1c with advanced fibrosis, as previously reported [53, 54], reinforces its biological plausibility and clinical relevance in MASLD management.

5.3 | Strengths and Limitations

A key strength of this study is the development of an AI-based fibrosis assessment model independent of platelet count and imaging, making it particularly suitable for large-scale screenings. The model demonstrated diagnostic accuracy comparable to conventional NITs while relying solely on routinely available biochemical and demographic markers, ensuring cost-effectiveness and scalability.

TABLE 4 | Comparison between machine learning models and noninvasive tests in the test dataset.

	AUC	Cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
≥ F2						
Support Vector Machine	0.886	0.389	0.857	0.785	0.631	0.927
FIB-4 index	0.887	1.961	0.786	0.862	0.710	0.903
APRI	0.853	0.925	0.750	0.892	0.750	0.892
≥ F3						
Support Vector Machine	0.882	0.389	0.944	0.720	0.447	0.982
FIB-4 index	0.900	2.448	0.778	0.893	0.636	0.944
APRI	0.820	0.925	0.778	0.813	0.500	0.939
F4						
Support Vector Machine	0.916	0.665	1.000	0.885	0.375	1.000
FIB-4 index	0.943	3.123	1.000	0.897	0.400	1.000
APRI	0.914	1.358	1.000	0.874	0.353	1.000

Abbreviations: APRI: aspartate aminotransferase-to-platelet ratio Index, AUC: area under the curve, F: fibrosis stage, FIB-4: fibrosis 4.

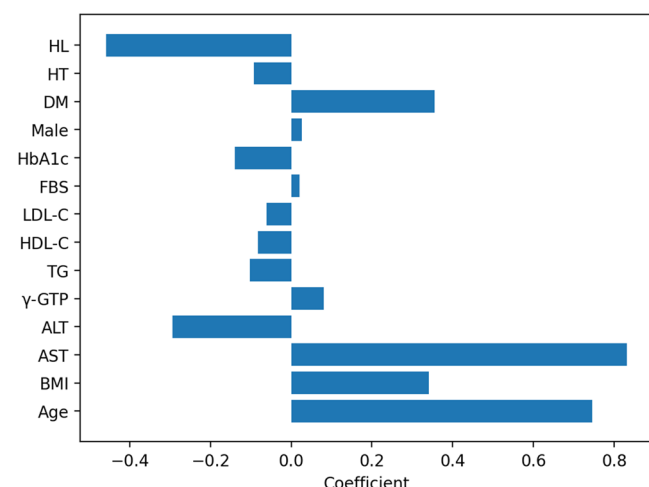


FIGURE 4 | Feature importance in the Support Vector Machine model. ALT: alanine aminotransferase, AST: aspartate aminotransferase, BMI: body mass index, DM: diabetes mellitus, FBS: fasting blood sugar, HbA1c: hemoglobin A1c, HDL-C: high-density lipoprotein cholesterol, HL: hyperlipidemia, HT: hypertension, LDL-C: low-density lipoprotein cholesterol, TG: triglyceride, γ -GTP: gamma-glutamyl transferase.

However, certain limitations should be noted. This study was retrospective and conducted at a single center, introducing potential selection bias. Additionally, the sample size was one-third of the 1202 cases estimated by Riley et al. as necessary for optimal predictive modeling [55]. Further validation across diverse populations is required to assess the model's generalizability.

5.4 | Future Implications

This study demonstrates that an AI-based fibrosis assessment model, utilizing only routinely available clinical markers, can

achieve diagnostic accuracy comparable to conventional NITs. Its integration into health check-up programs offers a practical, scalable solution for fibrosis screening, allowing for early identification and intervention in MASLD patients.

Acknowledgments

The authors thank Asami Yamazaki, Mie Karakida, and Yoshiaki Onda for their assistance in sample and database preparation. We also thank Trevor Ralph for his help in English proofreading.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available upon reasonable request from the corresponding author, T.K. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

References

1. Z. M. Younossi, "Non-Alcoholic Fatty Liver Disease—A Global Public Health Perspective," *Journal of Hepatology* 70 (2019): 531–544.
2. D. N. Amarapurkar, E. Hashimoto, L. A. Lesmana, J. D. Sollano, P. J. Chen, and K. L. Goh, "How Common Is Non-Alcoholic Fatty Liver Disease in the Asia-Pacific Region and Are There Local Differences?," *Journal of Gastroenterology and Hepatology* 22 (2007): 788–793.
3. A. M. Allen, T. M. Therneau, J. J. Larson, A. Coward, V. K. Somers, and P. S. Kamath, "Nonalcoholic Fatty Liver Disease Incidence and Impact on Metabolic Burden and Death: A 20 Year-Community Study," *Hepatology* 67 (2018): 1726–1736.
4. A. Mantovani, G. Petracca, G. Beatrice, et al., "Non-Alcoholic Fatty Liver Disease and Increased Risk of Incident Extrahepatic Cancers: A Meta-Analysis of Observational Cohort Studies," *Gut* 71, no. 4 (2022): 778–788, <https://doi.org/10.1136/gutjnl-2021-324191>.

5. T. Kimura, N. Tamaki, S. I. Wakabayashi, et al., "Colorectal Cancer Incidence in Steatotic Liver Disease (MASLD, MetALD, and ALD)," *Clinical Gastroenterology and Hepatology*, ahead of print, January 31 (2025), <https://doi.org/10.1016/j.cgh.2024.12.018>.
6. P. Angulo, D. E. Kleiner, S. Dam-Larsen, et al., "Liver Fibrosis, but no Other Histologic Features, Is Associated With Long-Term Outcomes of Patients With Nonalcoholic Fatty Liver Disease," *Gastroenterology* 149 (2015): 389–397.e10.
7. N. Tamaki, T. Kimura, S. I. Wakabayashi, et al., "Long-Term Clinical Outcomes in Steatotic Liver Disease and Incidence of Liver-Related Events, Cardiovascular Events and All-Cause Mortality," *Alimentary Pharmacology & Therapeutics* 60 (2024): 61–69.
8. N. Tamaki, T. Kimura, S. I. Wakabayashi, et al., "Cardiometabolic Criteria as Predictors and Treatment Targets of Liver-Related Events and Cardiovascular Events in Metabolic Dysfunction-Associated Steatotic Liver Disease," *Alimentary Pharmacology & Therapeutics* 60 (2024): 1033–1041.
9. J. Lee, Y. Vali, J. Boursier, et al., "Prognostic Accuracy of FIB-4, NAFLD Fibrosis Score and APRI for NAFLD-Related Events: A Systematic Review," *Liver International* 41 (2021): 261–270.
10. N. Tamaki, K. Imajo, S. R. Sharpton, et al., "Two-Step Strategy, FIB-4 Followed by Magnetic Resonance Elastography, for Detecting Advanced Fibrosis in NAFLD," *Clinical Gastroenterology and Hepatology* 21 (2023): 380–387.e3.
11. T. Kohro, Y. Furui, N. Mitsutake, et al., "The Japanese National Health Screening and Intervention Program Aimed at Preventing Worsening of the Metabolic Syndrome," *International Heart Journal* 49 (2008): 193–203.
12. D. Ichikawa, T. Saito, and H. Oyama, "Impact of Predicting Health-Guidance Candidates Using Massive Health Check-Up Data: A Data-Driven Analysis," *International Journal of Medical Informatics* 106 (2017): 32–36.
13. M. Suka, K. Yoshida, and S. Matsuda, "Effect of Annual Health Checkups on Medical Expenditures in Japanese Middle-Aged Workers," *Journal of Occupational and Environmental Medicine* 51 (2009): 456–461.
14. N. Tamaki, M. Kurosaki, D. Q. Huang, and R. Loomba, "Noninvasive Assessment of Liver Fibrosis and Its Clinical Significance in Non-alcoholic Fatty Liver Disease," *Hepatology Research* 52 (2022): 497–507.
15. T. Iwadare, T. Kimura, T. Okumura, et al., "Serum Autotaxin Is a Prognostic Indicator of Liver-Related Events in Patients With Non-Alcoholic Fatty Liver Disease," *Communications Medicine* 4 (2024): 73.
16. N. Fujimori, T. Umemura, T. Kimura, et al., "Serum Autotaxin Levels Are Correlated With Hepatic Fibrosis and Ballooning in Patients With Non-Alcoholic Fatty Liver Disease," *World Journal of Gastroenterology* 24 (2018): 1239–1249.
17. H. Uojima, K. Yamasaki, M. Sugiyama, et al., "Quantitative Measurements of M2BPGi Depend on Liver Fibrosis and Inflammation," *Journal of Gastroenterology* 59, no. 7 (2024): 598–608, <https://doi.org/10.1007/s00535-024-02100-3>.
18. D. Li, M. Zhang, S. Wu, H. Tan, and N. Li, "Risk Factors and Prediction Model for Nonalcoholic Fatty Liver Disease in Northwest China," *Scientific Reports* 12 (2022): 13877.
19. Y.-X. Liu, X. Liu, C. Cen, et al., "Comparison and Development of Advanced Machine Learning Tools to Predict Nonalcoholic Fatty Liver Disease: An Extended Study," *Hepatobiliary & Pancreatic Diseases International* 20 (2021): 409–415.
20. T. Okanoue, T. Shima, Y. Mitsumoto, et al., "Novel Artificial Intelligent/Neural Network System for Staging of Nonalcoholic Steatohepatitis," *Hepatology Research* 51 (2021): 1044–1057.
21. M. E. Rinella, J. V. Lazarus, V. Ratziu, et al., "A Multisociety Delphi Consensus Statement on New Fatty Liver Disease Nomenclature," *Journal of Hepatology* 79 (2023): 1542–1556.
22. N. Tanaka, T. Aoyama, S. Kimura, and F. J. Gonzalez, "Targeting Nuclear Receptors for the Treatment of Fatty Liver Disease," *Pharmacology & Therapeutics* 179 (2017): 142–157.
23. N. Tanaka, T. Kimura, N. Fujimori, T. Nagaya, M. Komatsu, and E. Tanaka, "Current Status, Problems, and Perspectives of Non-Alcoholic Fatty Liver Disease Research," *World Journal of Gastroenterology* 25 (2019): 163–177.
24. T. Kimura, A. Shinji, A. Horiuchi, et al., "Clinical Characteristics of Young-Onset Ischemic Colitis," *Digestive Diseases and Sciences* 57 (2012): 1652–1659.
25. T. Kimura, A. Shinji, N. Tanaka, et al., "Association Between Lower Air Pressure and the Onset of Ischemic Colitis: A Case-Control Study," *European Journal of Gastroenterology & Hepatology* 29 (2017): 1071–1078.
26. T. Kimura, A. Kobayashi, N. Tanaka, et al., "Clinicopathological Characteristics of Non-B Non-C Hepatocellular Carcinoma Without Past Hepatitis B Virus Infection," *Hepatology Research* 47 (2017): 405–418.
27. N. Fujimori, T. Kimura, N. Tanaka, et al., "2-Step PLT16-AST44 Method: Simplified Liver Fibrosis Detection System in Patients With Non-Alcoholic Fatty Liver Disease," *Hepatology Research* 52 (2022): 352–363.
28. D. E. Kleiner, E. M. Brunt, M. Van Natta, et al., "Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease," *Hepatology* 41 (2005): 1313–1321.
29. C. H. Ng, W. H. Lim, G. E. Hui Lim, et al., "Mortality Outcomes by Fibrosis Stage in Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis," *Clinical Gastroenterology and Hepatology* 21 (2023): 931–939.e5.
30. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.
31. X. Robin, N. Turck, A. Hainard, et al., "pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves," *BMC Bioinformatics* 12 (2011): 77.
32. M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support Vector Machines," *IEEE Intelligent Systems and Their Applications* 13 (1998): 18–28.
33. A.-L. Boulesteix, S. Janitz, J. Kruppa, and I. R. König, "Overview of Random Forest Methodology and Practical Guidance With Emphasis on Computational Biology and Bioinformatics," *WIREs Data Mining and Knowledge Discovery* 2 (2012): 493–507.
34. T. Chen, T. He, M. Benesty, et al., "Xgboost: Extreme Gradient Boosting," *R Package Version 04-2, 2015*, 1: 1–4.
35. G. Ke, Q. Meng, T. Finley, et al., "Lightgbm: A Highly Efficient Gradient Boosting Decision Tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc, 2017), 3146–3154.
36. D. M. P. Murti, U. Pujianto, A. P. Wibawa, and M. I. Akbar, "K-Nearest Neighbor (K-NN) Based Missing Data Imputation," in *5th International Conference on Science in Information Technology (ICSITech) 2019* (IEEE, 2019), 83–88, <https://doi.org/10.1109/ICSITech46713.2019.898753>.
37. S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation Over Hold-Out Validation on Colossal Datasets for Quality Classification," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)* (IEEE, 2016), 78–83.
38. WHO Expert Consultation, "Appropriate Body-Mass Index for Asian Populations and Its Implications for Policy and Intervention Strategies," *Lancet* 363 (2004): 157–163.
39. J. Boursier, M. Roux, C. Costentin, et al., "Practical Diagnosis of Cirrhosis in Non-Alcoholic Fatty Liver Disease Using Currently Available

Non-Invasive Fibrosis Tests,” *Nature Communications* 14, no. 1 (2023): 5219, <https://doi.org/10.1038/s41467-023-40328-4>.

40. T. Kimura, T. Iwadare, S. I. Wakabayashi, et al., “Thrombospondin 2 Is a Key Determinant of Fibrogenesis in Non-Alcoholic Fatty Liver Disease,” *Liver International* 44, no. 2 (2023): 483–496, <https://doi.org/10.1111/liv.15792>.

41. T. Kimura, N. Tanaka, N. Fujimori, et al., “Serum Thrombospondin 2 Is a Novel Predictor for the Severity in the Patients With NAFLD,” *Liver International: Official Journal of the International Association for the Study of the Liver* 41 (2021): 505–514.

42. M. Abe, T. Miyake, A. Kuno, et al., “Association Between *Wisteria floribunda* Agglutinin-Positive Mac-2 Binding Protein and the Fibrosis Stage of Non-Alcoholic Fatty Liver Disease,” *Journal of Gastroenterology* 50 (2015): 776–784.

43. N. Fujimori, N. Tanaka, S. Shibata, et al., “Controlled Attenuation Parameter Is Correlated With Actual Hepatic Fat Content in Patients With Non-Alcoholic Fatty Liver Disease With None-to-Mild Obesity and Liver Fibrosis,” *Hepatology Research* 46 (2016): 1019–1027.

44. T. Nagaya, N. Tanaka, T. Suzuki, et al., “Down-Regulation of SREBP-1c Is Associated With the Development of Burned-Out NASH,” *Journal of Hepatology* 53 (2010): 724–731.

45. A. Facciorusso, V. Del Prete, A. Turco, R. V. Buccino, M. C. Nacchiero, and N. Muscatiello, “Long-Term Liver Stiffness Assessment in Hepatitis C Virus Patients Undergoing Antiviral Therapy: Results From a 5-Year Cohort Study,” *Journal of Gastroenterology and Hepatology* 33 (2018): 942–949.

46. A. M. Dinani, K. V. Kowdley, and M. Nouredin, “Application of Artificial Intelligence for Diagnosis and Risk Stratification in NAFLD and NASH: The State of the Art,” *Hepatology* 74 (2021): 2233–2240.

47. G. L.-H. Wong, P.-C. Yuen, A. J. Ma, A. W.-H. Chan, H. H.-W. Leung, and V. W.-S. Wong, “Artificial Intelligence in Prediction of Non-Alcoholic Fatty Liver Disease and Fibrosis,” *Journal of Gastroenterology and Hepatology* 36 (2021): 543–550.

48. J. P. Sowa, D. Heider, L. P. Bechmann, G. Gerken, D. Hoffmann, and A. Canbay, “Novel Algorithm for Non-Invasive Assessment of Fibrosis in NAFLD,” *PLoS One* 8 (2013): e62439.

49. M. Docherty, S. A. Regnier, G. Capkun, et al., “Development of a Novel Machine Learning Model to Predict Presence of Nonalcoholic Steatohepatitis,” *Journal of the American Medical Informatics Association* 28 (2021): 1235–1241.

50. D. Chang, E. Truong, E. A. Mena, et al., “Machine Learning Models Are Superior to Noninvasive Tests in Identifying Clinically Significant Stages of NAFLD and NAFLD-Related Cirrhosis,” *Hepatology* 77 (2022): 546–557.

51. P. Aggarwal and N. Alkhouri, “Artificial Intelligence in Nonalcoholic Fatty Liver Disease: A New Frontier in Diagnosis and Treatment,” *Clinical Liver Disease* 17 (2021): 392–397.

52. A. Vellido, “The Importance of Interpretability and Visualization in Machine Learning for Applications in Medicine and Health Care,” *Neural Computing and Applications* 32 (2020): 18069–18083.

53. N. Tamaki, S. I. Wakabayashi, T. Kimura, et al., “Glycemic Control Target for Liver and Cardiovascular Events Risk in Metabolic Dysfunction-Associated Steatotic Liver Disease,” *Hepatology Research* 54 (2024): 753–762.

54. P. Loria, G. Marchesini, F. Nascimbeni, et al., “Cardiovascular Risk, Lipidemic Phenotype and Steatosis. A Comparative Analysis of Cirrhotic and Non-Cirrhotic Liver Disease due to Varying Etiology,” *Atherosclerosis* 232 (2014): 99–109.

55. R. D. Riley, J. Ensor, K. I. E. Snell, et al., “Calculating the Sample Size Required for Developing a Clinical Prediction Model,” *BMJ* 368 (2020): m441.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.