



OPEN ACCESS

EDITED BY

Tong Wang,
University of Connecticut, United States

REVIEWED BY

Yuan Gao,
Genentech Inc., United States
He Gao,
University of California, San Diego,
United States

*CORRESPONDENCE

Xuewei Zhuang
✉ zhuangxuewei@sdu.edu.cn

RECEIVED 24 August 2025

ACCEPTED 29 October 2025

PUBLISHED 18 November 2025

CITATION

Li X, Zhao X, Zhang R and Zhuang X (2025)
Harnessing serum VOCs and machine
learning for the early detection of MAFLD.
Front. Endocrinol. 16:1691853.
doi: 10.3389/fendo.2025.1691853

COPYRIGHT

© 2025 Li, Zhao, Zhang and Zhuang. 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.

Harnessing serum VOCs and machine learning for the early detection of MAFLD

Xin Li¹, Xiaoyue Zhao¹, Ruonan Zhang¹ and Xuewei Zhuang^{2*}

¹Shandong University of Traditional Chinese Medicine, Jinan, China, ²Shandong University Affiliated Shandong Provincial Third Hospital Department of Clinical Laboratory, Jinan, China

Introduction: Metabolic dysfunction-associated fatty liver disease (MAFLD) is a complex metabolic disorder and one of the leading causes of chronic liver disease worldwide. Current diagnostic tools, such as ultrasound, lack sufficient sensitivity for detecting early-stage disease, emphasizing the urgent need for novel and non-invasive diagnostic strategies. Metabolomics, particularly the profiling of volatile organic compounds (VOCs) in biofluids, has emerged as a promising approach for biomarker discovery in metabolic diseases.

Methods: In this preliminary single-center study, serum samples were collected from 199 participants, including 110 MAFLD patients and 89 healthy controls. Volatile organic compounds were analyzed using gas chromatography–ion mobility spectrometry (GC-IMS). Machine learning algorithms, including random forest, were applied to construct diagnostic models and identify key discriminatory metabolites. Clinical and biochemical parameters such as age, body mass index, liver function, and lipid profiles were also compared between groups.

Results: A total of 79 serum VOCs were detected, among which 54 showed significant differences between MAFLD patients and controls (29 identified and 25 unidentified). The random forest model exhibited the best diagnostic performance, achieving a test AUC of 0.941, with 86.7% sensitivity and 88.5% specificity. Seven key VOCs were identified as important contributors to the model, including two upregulated compounds (2-Butoxyethanol and Cyclopentanone-D) and five downregulated compounds ((E)-3-hexenoic acid, 2-Ethylbutanal, 2-Propyl acetate, Benzaldehyde-M, and Furaneol). Notably, 2-pentylfuran displayed significant variation across different pathological grades of MAFLD, suggesting potential as a stage-specific biomarker.

Discussion: This study demonstrates that serum VOC profiling using GC-IMS combined with machine learning can effectively distinguish MAFLD patients from healthy individuals. The identified VOC signatures, particularly 2-pentylfuran, may serve as non-invasive biomarkers for MAFLD diagnosis and staging. However, due to the limited sample size and single-center design, these findings require validation in larger, multi-center, and longitudinal studies to confirm their clinical applicability, especially for early disease detection.

KEYWORDS

metabolic dysfunction-associated fatty liver disease, volatile organic compounds, gas chromatography–ion mobility spectrometry, machine learning, biomarker, early diagnosis

1 Introduction

Metabolic dysfunction-associated fatty liver disease (MAFLD) is a complex metabolic disorder arising from an intricate interplay of genetic susceptibility, metabolic disturbances, and environmental factors. An international consensus group recommended in 2020 that MAFLD should replace the historically used term “non-alcoholic fatty liver disease” (NAFLD) (1–3). It is a leading cause of chronic liver disease worldwide (4, 5), affecting approximately 25% of the global population. It is a common reason for abnormal liver function tests and can account for up to 90% of cases of asymptomatic elevation in alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels once other causes of liver disease are excluded (6). Substantial research shows that MAFLD can progress to severe outcomes, including end-stage liver disease and hepatocellular carcinoma (HCC) (7). Liver biopsy remains the gold standard for diagnosis, but its invasiveness and associated clinical risks limit routine use. Consequently, diagnosis typically relies on a combination of medical history, laboratory tests, and imaging. Ultrasound can reliably detect moderate-to-severe steatosis and is currently the preferred imaging technique for screening in clinical and population-based settings (8). On ultrasound, the diffuse fatty infiltration of MAFLD often produces a hyperechoic texture, or “bright liver.” However, MAFLD is often asymptomatic in its early stages. When it comes to early detection, conventional methods have limited utility, and there is an urgent need for novel, non-invasive diagnostic methods.

Metabolomics, i.e., the systematic identification and quantification of small-molecule metabolites (<1500 Da) in a biological sample, offers a promising approach. The metabolome, or the complete set of metabolites in a biological system, provides a functional snapshot of the physiological state (9). Volatile organic compounds (VOCs), a subset of the metabolome, have emerged as powerful biomarkers. They can serve as non-invasive indicators of an individual’s metabolic status, and over the past decades, the VOC profiles of diverse biological fluids, including urine, breath, sweat, and feces, have been linked to various physiological and pathological states (10–13). For example, diagnostic models have been developed to identify cholangiocarcinoma using bile VOCs (14) and HCC using urinary VOCs (15). In addition, combining data from different sample types, such as blood and urine, has been shown to enhance the discriminative power and accuracy of diagnostic models (16).

Some researchers explored VOC signatures for NAFLD using breath and urine, each with distinct cohorts, platforms, and key findings. Shen et al. (17) applied 10 machine learning algorithms to 341 exhaled VOCs in a large cohort ($n = 1,501$). They showed that

adding breath VOCs to demographic and biochemical predictors significantly improved NAFLD classification, identified specific breath compounds (e.g., 2-propanol, acetone) as influential features, and discussed environmental and exposure confounders and the need for external validation. Cozzolino et al. (18) profiled urinary VOCs in cohorts with NAFLD, type 2 diabetes, and their coexistence. They reported a set of urinary volatiles that discriminated disease groups and highlighting links to altered amino acid and lipid metabolism and noted that urine VOCs may reflect systemic metabolic changes rather than liver-specific processes. Skarysz et al. (19) developed convolutional neural networks to detect VOCs directly from raw GC–MS breath data, bypassing expert-led deconvolution and demonstrating automated detection of more compounds with high specificity and far shorter processing time. The CNN pipeline, evaluated on 120 clinical samples targeting 30 VOCs, outperformed conventional expert analysis and produced robust, scalable VOC lists suitable for rapid breath-based biomarker discovery. While these methods have demonstrated impressive success, breathomics can be influenced by environmental exposures, and urinary analyses may reflect systemic rather than liver-specific metabolic changes. Indeed, Masoodi et al. (20) reviewed metabolomics and lipidomics in NAFLD and emphasized the need for multi-matrix validation. They warned that the variability in sampling, analytic workflows, and identification confidence tends to limit cross-study comparability, and argued for orthogonal confirmation of VOC candidates by MS-based techniques and standardization of collection/analysis procedures. Therefore, in this study, we turned to serum VOCs instead, which may more directly reflect hepatic metabolic processes.

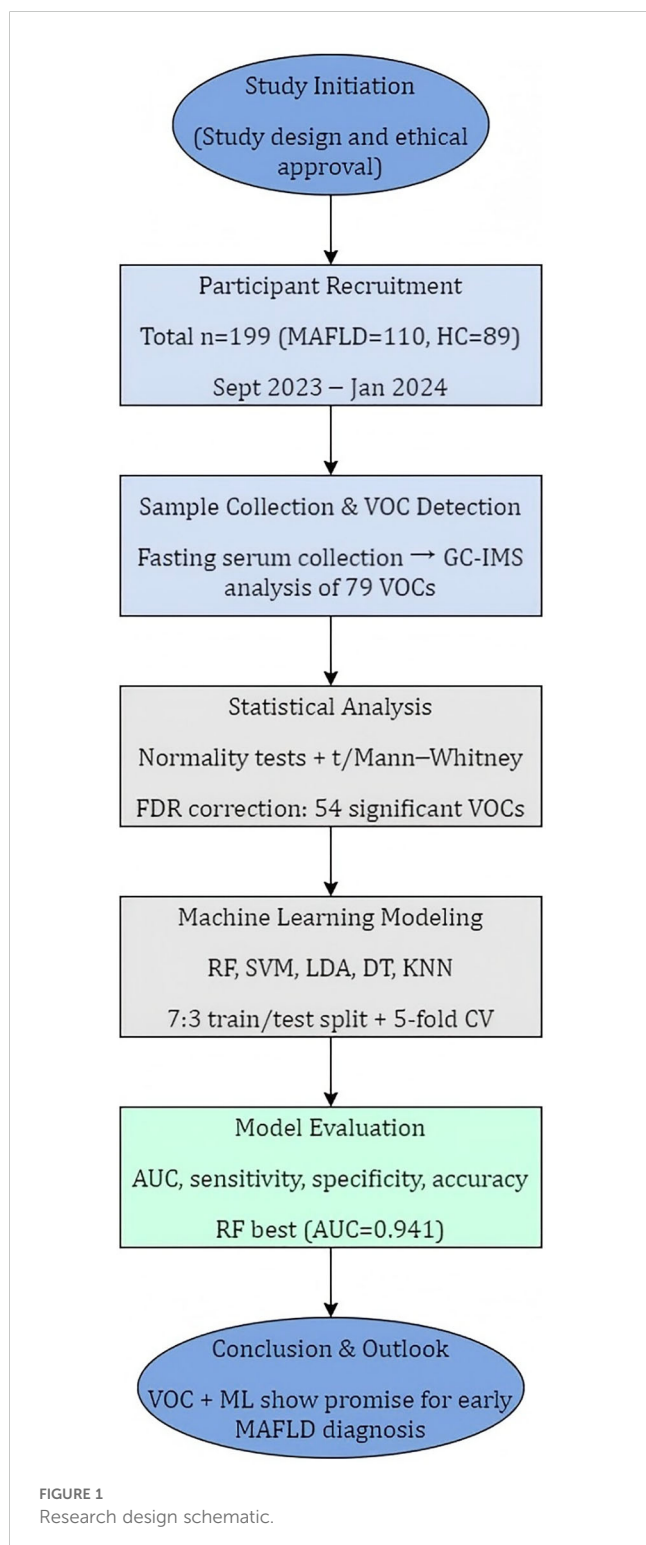
Headspace solid-phase microextraction (HS-SPME) coupled with gas chromatography–mass spectrometry (GC–MS) has proven to be a simple, rapid, and effective technique for analyzing VOCs in biofluids. This approach pairs chromatographic separation with mass spectral identification and extensive libraries for definitive compound assignment, but for large clinical screening, it is limited by long run times, intensive sample preparation, reliance on operator expertise, moderate throughput, high capital and per-sample costs, and occasional poor sensitivity or ambiguous identification for trace, highly polar, or coeluting breath VOCs. In contrast, gas chromatography–ion mobility spectrometry (GC–IMS) addresses these limitations by adding an orthogonal ion–mobility (drift–time) separation that improves resolution of isobaric/coeluting species, thus offering enhanced sensitivity for low-abundance metabolites, faster analyses with simpler headspace workflows, a smaller bench-top footprint, and lower operational burden. It can be readily integrated into clinical laboratory workflows and is particularly suitable for a large cohort study focusing on clinically deployable, high-throughput screening (21). Hence, in this work, we used GC–IMS to compare serum VOC profiles between MAFLD patients and healthy controls, to characterize disease-related metabolic changes and construct a novel, serum-based diagnostic model for MAFLD.

Abbreviations: MAFLD, Metabolic dysfunction-associated fatty liver disease; VOCs, Volatile organic compounds; GC–IMS, Gas chromatography–ion mobility spectrometry; NAFLD, Non-alcoholic fatty liver disease; FDR, False discovery rate; RF, Random forest; SVM, Support vector machine; LDA, Linear discriminant analysis; DT, Decision tree; KNN, K-nearest neighbors.

2 Materials and methods

2.1 Study design and participants

Figure 1 illustrates the overall study design. A cohort was recruited between September 2023 and January 2024 from the Department of Physical Examination at Shandong Provincial



Third Hospital. To minimize selection bias in this non-randomized study, we enrolled consecutive, eligible patients presenting to the hepatology clinic. Healthy controls were recruited from routine health check-up programs and were required to have normal liver imaging, normal liver enzyme levels, and no history of chronic liver or metabolic disease.

As our study's inclusion criteria aligned with the 2020 MAFLD definition, we use the term MAFLD throughout this manuscript to most accurately describe our patient cohort. MAFLD grade (mild, moderate, or severe) was determined using standard ultrasound criteria, including hepatic echogenicity, clarity of intrahepatic vasculature, diaphragm visualization, and posterior beam attenuation.

2.1.1 Inclusion criteria for the MAFLD group

1. Age between 18 and 75 years.
2. Evidence of hepatic steatosis on imaging studies or liver biopsy.
3. Presence of at least one of the following conditions:
 - Overweight or obesity (Body Mass Index [BMI] ≥ 23 kg/m² for Asians).
 - Diagnosed type 2 diabetes mellitus.
 - Evidence of metabolic dysfunction, defined as the presence of at least two of the following seven traits:
 - a. Waist circumference: ≥ 90 cm (male) or ≥ 80 cm (female) for Asians.
 - b. Blood pressure: $\geq 130/85$ mmHg, or use of antihypertensive medication.
 - c. Triglycerides: ≥ 1.70 mmol/L, or use of lipid-lowering medication.
 - d. High-density lipoprotein cholesterol (HDL-C): <1.0 mmol/L (male) or <1.3 mmol/L (female), or use of lipid-modifying medication.
 - e. Prediabetes: Fasting glucose 5.6–6.9 mmol/L, 2-hour postprandial blood glucose 7.8–11.0 mmol/L, or glycated hemoglobin (HbA1c) 5.7%–6.4%.
 - f. Homeostasis model assessment of insulin resistance (HOMA-IR) index: ≥ 2.5 .
 - g. High-sensitivity C-reactive protein (hs-CRP) level: ≥ 2 mg/L.

2.1.2 Inclusion criteria for the healthy control group

1. Age ≥ 18 years and provided voluntary written informed consent.
2. No history of major cardiac, hepatic, renal, neurological, or infectious disease.
3. BMI within the normal range (e.g., 18.5–22.9 kg/m²).
4. No heavy alcohol consumption, smoking, or any substance abuse.
5. No use of medications known to affect metabolic outcomes within the past month.
6. Not pregnant or breastfeeding.

2.2 Sample preparation

Peripheral blood samples were collected from all participants in the morning after an overnight fast. Samples were drawn into serum separator tubes containing a clot activator and centrifuged at 3000g for 10 min at 4°C within two hours of collection. The resulting serum was transferred to a new tube and centrifuged again at 12,000g for 10 min at 4°C to remove residual cellular debris. The final serum (supernatant) was aliquoted into cryotubes, flash-frozen, and stored at −80°C until analysis. Each sample was thawed no more than twice before analysis. To minimize exogenous contamination and avoid systematic bias, blood collection and processing were performed in a standardized environment using identical consumables for all samples.

2.3 GC-IMS analysis of serum VOCs

For each analysis, a 200 µL serum aliquot was placed in a 20 mL headspace vial, sealed, and incubated at 80°C for 10 min. Following incubation, 1 mL of the headspace gas was automatically injected into a GC-IMS system (FlavourSpec, G.A.S., Dortmund, Germany). The system used nitrogen as both the carrier and drift gas. The carrier gas flow was programmed as follows: 2 mL/min for 2 min, ramped linearly to 100 mL/min over 8 min, and then held at 150 mL/min for an additional 5 min. The IMS drift gas flow was constant at 150 mL/min. The drift tube, GC column, and inlet temperatures were maintained at 45, 60, and 60°C, respectively.

VOC signals were characterized by their retention index (RI), drift time (DT), and peak height (PH). Compounds were identified by comparing their spectral features against the National Institute of Standards and Technology (NIST) and integrated GC-IMS library databases, and external reference standards were used when appropriate. Peaks that could not be matched were numbered and reported as unidentified VOCs. Peak height was used for relative quantification across all samples.

2.4 Statistical and bioinformatic analysis

Descriptive statistics were calculated for all baseline characteristics. Continuous variables were presented as mean ± standard deviation (SD) or median [P25, P75] based on their distribution, assessed using the Shapiro–Wilk test. Categorical variables were presented as counts and percentages. For comparisons of individual VOC concentrations between the MAFLD and control groups, we used the independent samples t-test for normally distributed data and the Mann–Whitney U test for non-normally distributed data. All tests were two-sided, and statistical significance was verified when $p < 0.05$.

To identify factors associated with MAFLD, we first performed univariate logistic regression. Variables with a p -value < 0.05 were then included in a multivariate binary logistic regression model to identify independent factors associated with MAFLD. For VOC feature selection, a multi-step filtering process was applied. First, we

corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) method, with a significance threshold of $q < 0.05$. To further ensure robustness, we only retained features that also exhibited an absolute fold change of at least 1.5 between the MAFLD and control groups and were detected in at least 20% of samples in either group. VOCs that met all these criteria were selected as the final feature set for machine learning.

We developed five machine learning models to classify MAFLD status based on the significant VOCs: random forest (RF), support vector machine (SVM), linear discriminant analysis (LDA), decision tree (DT), and k-nearest neighbors (KNN). The full dataset was randomly partitioned into a training set (70%) and a test set (30%) using stratified sampling to maintain the case-to-control ratio. Model hyperparameters were optimized using five-fold cross-validation within the training set. The final, optimized models were then evaluated on the independent test set. Diagnostic performance was evaluated using receiver operating characteristic (ROC) analysis, and performance was assessed using the area under the curve (AUC). To evaluate model robustness, we conducted a sensitivity analysis by excluding participants with ages in the lowest and highest 5% of the distribution, re-training the models, and comparing the performance metrics with those from the full dataset. All statistical analyses were performed using SPSS 22 (IBM, Armonk, NY, USA) and R (version 4.3.3), and the caret and pROC packages were utilized for machine learning tasks. Default parameters were applied unless specified otherwise.

3 Results

3.1 Baseline clinical characteristics of the study cohort

The study cohort included 199 participants: 110 patients with MAFLD and 89 healthy controls (HC). While the groups were balanced for sex, height, and several laboratory parameters (total bilirubin, creatinine, blood urea nitrogen), the MAFLD group was significantly older and had a higher weight and BMI ($p < 0.001$ for all). As detailed in [Table 1](#), the MAFLD patients also exhibited significant alterations in liver enzymes (ALT, AST, GGT) and lipid profiles (TC, LDL, TG) compared to the controls.

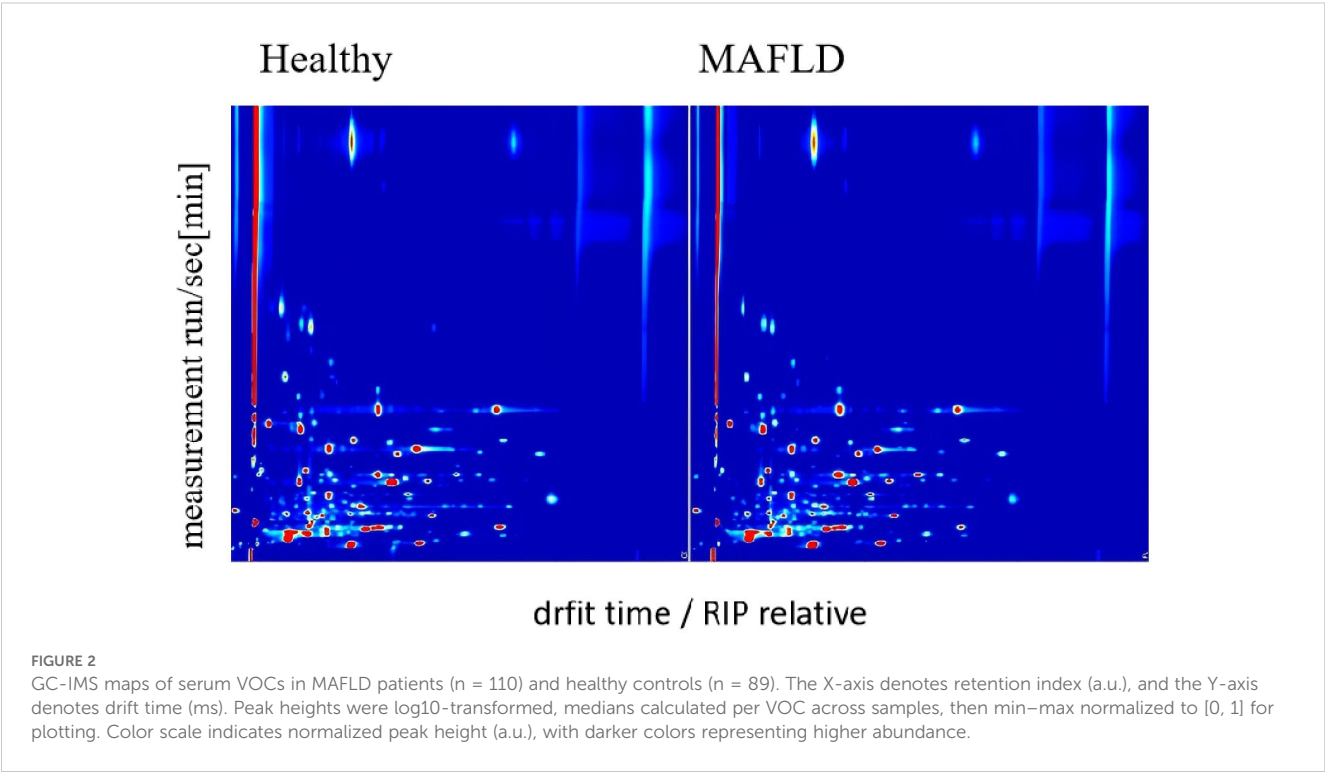
3.2 Serum VOC profiles distinguish MAFLD patients from healthy controls

The serum samples from all 199 participants were analyzed by GC-IMS to generate VOC profiles. Visual inspection of the resulting topographic plots revealed distinct differences in the VOC signatures between the MAFLD and HC groups ([Figure 2](#)). A total of 79 distinct VOC signals were quantified across all samples ([Supplementary Table S1](#)). Of these, 57 showed significant uncorrected differences between groups. After applying Benjamini–Hochberg (FDR) correction, 54 VOCs remained statistically significant ($q < 0.05$) and were selected as features for

TABLE 1 Patient characteristics.

Characteristics	Healthy Control (n = 89)	MAFLD (n = 110)	p-value
Sex			0.317
Male	36 (40.45%)	37 (36.64%)	
Female	53 (59.55%)	73 (66.36%)	
Age (years)	37.0 [28.5, 49.5]	49.0 [37.0, 60.0]	<0.001
Height (cm)	166.30 [162.40, 173.50]	165.45 [159.68, 172.90]	0.270
Weight (kg)	63.40 [56.60, 72.45]	73.95 [66.10, 81.85]	<0.001
BMI (kg/m ²)	22.90 [21.15, 24.85]	26.45 [24.40, 29.33]	<0.001
Lab test results			
ALT (U/L)	15.30 [11.75, 20.35]	20.75 [15.20, 31.28]	<0.001
AST (U/L)	18.70 [16.55, 21.85]	20.55 [17.50, 26.55]	0.002
GGT (U/L)	14.50 [11.30, 21.15]	22.10 [17.98, 31.90]	<0.001
TBIL (μmol/L)	12.60 [9.80, 17.55]	13.20 [10.08, 16.03]	0.681
TC (mmol/L)	4.54 [3.95, 4.88]	5.04 [4.44, 5.90]	<0.001
LDL (mmol/L)	2.73 [2.43, 3.09]	3.31 [2.83, 3.79]	<0.001
TG (mmol/L)	0.90 [0.68, 1.28]	1.67 [1.19, 2.41]	<0.001
GLU (mmol/L)	5.20 [4.90, 5.40]	5.45 [5.20, 6.00]	<0.001
Scr (μmol/L)	57.50 [52.20, 65.90]	57.00 [49.78, 66.30]	0.722
UA (μmol/L)	303.10 [274.10, 338.30]	364.8 [317.23, 431.65]	<0.001
BUN (mmol/L)	4.50 [3.85, 5.60]	4.75 [4.20, 5.50]	0.303

[§]Data are expressed as count (ratio) or median [P25, P75]. ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; TBIL, total bilirubin; TC, total cholesterol; LDL, low-density lipoprotein; TG, triglycerides; GLU, glucose; Scr, serum creatinine; UA, uric acid; BUN, blood urea nitrogen.



developing machine learning models (Supplementary Table S2). Some peaks were unidentified due to the current limitations of GC-IMS reference libraries, but they demonstrated consistent group differences and were thus retained for model construction.

3.3 Machine learning models achieve high diagnostic accuracy

Five machine learning models were trained using the 54 significant VOC features, which included 33 identified and 21 identified compounds, and the diagnostic performance of the models was evaluated on an independent test set. Table 2 shows that the RF model demonstrated the best performance, achieving an AUC of 0.941 on the test set, with an accuracy of 87.5%, sensitivity of 86.7%, and specificity of 88.5%. The SVM model also showed excellent performance (test set AUC = 0.927), whereas the LDA, DT, and GLM models were less effective. Figure 3 shows the ROC curves of all the models on the test set. Figure 4 shows the confusion matrix of the RF model on the test set.

To identify the most influential biomarkers among the 54 significant VOCs, we calculated, for each VOC, the Gini coefficient of its peak–height distribution across all 199 subjects and then ranked features by their importance in the top-performing RF model. The analysis revealed seven key VOCs that contributed most significantly to the model’s predictive power (Figure 5). Among them, two compounds were upregulated (2-butoxyethanol, cyclopentanone-D) and five were downregulated ((E)-3-hexenoic acid, 2-ethylbutanal, 2-propyl acetate, benzaldehyde-M, furaneol) in the MAFLD group. Figure 6 visualizes the distinct expression patterns of these top compounds. A sensitivity analysis that excluded age outliers confirmed the robustness of this VOC signature.

TABLE 2 Diagnostic performance of machine learning models.

Model [†]	RF	SVM	LDA	DT	GLM
CV_AUC	0.936	0.896	0.83	0.848	0.692
Test_AUC	0.941	0.927	0.876	0.878	0.764
Acc_0.5	0.875	0.875	0.857	0.839	0.804
Sens_0.5	0.867	0.867	0.867	0.833	0.867
Spec_0.5	0.885	0.885	0.846	0.846	0.731
Thr_Youden	0.566	0.598	0.347	0.850	0.496
Acc_Youden	0.875	0.875	0.875	0.857	0.804
Sens_Youden	0.833	0.833	0.900	0.800	–
Spec_Youden	0.923	0.923	0.846	0.923	–

[†]RF, random forest; SVM, support vector machine; LDA, linear discriminant analysis; DT, decision tree; KNN, k-nearest neighbors. Metrics include area under the receiver operating characteristic curve (AUC), accuracy (Acc), sensitivity (Sens), and specificity (Spec) at default (0.5) and Youden’s index thresholds (Thr). All results reflect the diagnostic performance of binary classification between MAFLD and healthy control groups. The model performance was evaluated by cross-validation (CV) and using independent test sets.

3.4 Association of VOCs with clinical parameters and disease severity

We next explored the association between the 33 identified significant VOCs and clinical parameters within the cohort. Numerous VOCs were significantly correlated with demographic factors such as sex and age (Figure 7). Notably, when analyzing the relationship with disease severity as determined by ultrasound, we found that the abundance of 2-pentyl furan varied significantly across mild, moderate, and severe stages of MAFLD ($p < 0.05$). This suggests that 2-pentyl furan may be a potential biomarker for monitoring MAFLD progression.

4 Discussion

VOC-based diagnostic model offers advantages such as convenient sampling, rapid detection, non-invasiveness, high patient acceptability, and the possibility of repeated sampling for dynamic monitoring (22, 23). This study demonstrates that a serum VOC-based signature, coupled with machine learning, can distinguish patients with MAFLD from healthy individuals with high accuracy. Current modalities of MAFLD diagnosis like ultrasound have limited sensitivity for early-stage disease, and our VOC-based model offers a promising, non-invasive alternative. The RF model, our top-performing algorithm, achieved an excellent AUC of 0.941, suggesting that serum VOCs reflect distinct metabolic perturbations in MAFLD. By identifying key VOCs linked to MAFLD, this work not only provides a potential diagnostic tool but also offers insights into the underlying pathophysiology of the disease.

A key strength of our study is the robustness of the VOC signal. Although the MAFLD and control groups had a significant age disparity, a sensitivity analysis excluding age extremes confirmed that the model’s performance was not driven by this demographic difference. Thus, the identified VOC signature must genuinely reflect disease-related metabolic changes. However, our findings should be interpreted with caution, since this work was a preliminary, single-center study with a limited sample size.

4.1 Comparison with existing literature and the novel contributions of this work

Our findings align with a growing body of research demonstrating the utility of VOCs as biomarkers in various diseases. For example, Hu et al. found a close relationship between urinary VOC metabolites and increased systemic inflammation, with smokers being more susceptible (24). This finding underscores the link between VOCs and inflammation—a key driver of MAFLD progression—and highlights the importance of accounting for lifestyle factors like smoking in metabolomic studies. Li et al. utilized GC-IMS to analyze VOC variations and explored the potential application of VOCs in the phenotypic detection of Carbapenem-Resistant *Klebsiella pneumoniae*

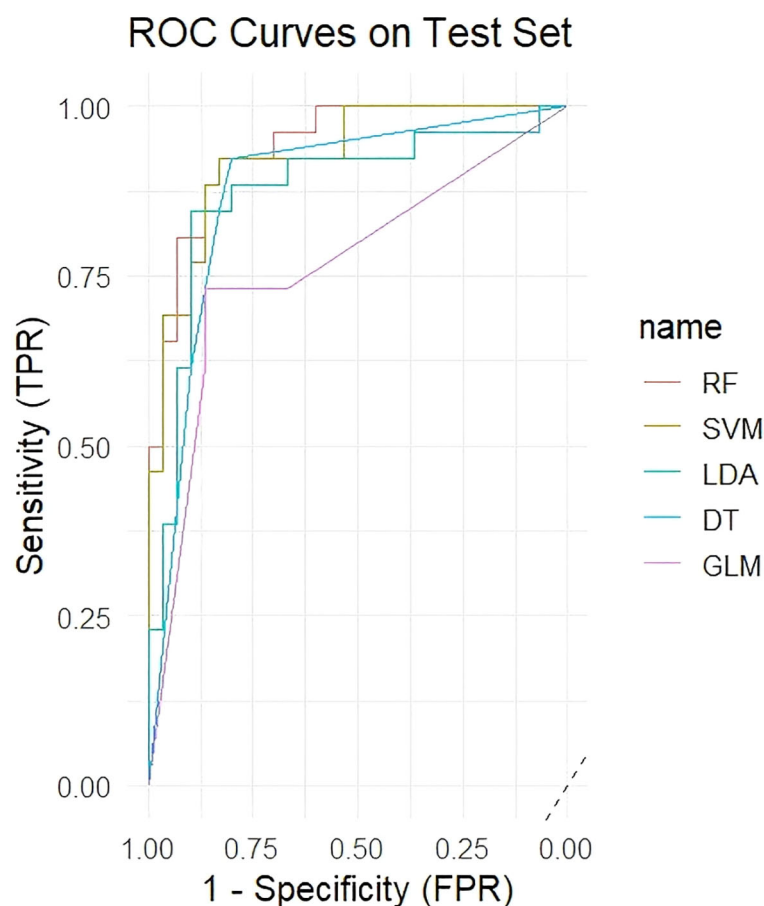


FIGURE 3
ROC curves of the five models on the test set.

(CRKP) strains (25). Previous studies also examined the use of VOCs as diagnostic biomarkers for esophageal cancer (26), pancreatic cancer (27), prostate cancer (28), esophageal adenocarcinoma (29), HCC (15), and colorectal cancer (30). A number of studies have investigated the diagnostic value of VOCs in liver disease (31), particularly in breath-based analyses. Breathomics approaches have been explored for HCC, cirrhosis, and MAFLD (32), with some technologies even advancing toward commercialization (e.g., VOC-based breath analyzers) (33, 34).

The combination of VOCs and machine learning has been deployed successfully in various scenarios. Fu et al. developed a machine learning model using urinary VOC metabolites and demographic data to predict cardiovascular disease risk (35). Their Random Forest model achieved a high predictive accuracy (AUC = 0.8143), which reflects our findings of the RF model and shows the power of combining metabolomic data with machine learning for risk stratification in complex metabolic diseases. While their model was effective, our serum-based approach in a hepatic disease context yielded an even higher AUC (0.941 vs. 0.814), highlighting the strong, tissue-proximal signal captured from serum. Thomas et al. (36) reported that machine learning analysis of exhaled VOC profiles can non-invasively detect cirrhosis and portal hypertension, thus offering a promising biomarker strategy

for liver disease. Patnaik et al. (37) conducted a pilot study where they measured breath concentrations of isoprene, limonene, and dimethyl sulfide before and after exercise and used machine learning regression to predict liver-function scores. These studies all validate the overall approach.

Our work offers several unique contributions. First, unlike most prior MAFLD research that focused on exhaled breath or urine, we analyzed serum VOCs (17, 18). This sample matrix may more directly reflect endogenous hepatic metabolism and is less susceptible to confounding from environmental or dietary exposures (22). Second, our use of GC-IMS provided a rapid and highly sensitive platform suitable for clinical workflows (38). Third, by comparing multiple machine learning algorithms, we not only identified RF as the optimal model for this dataset but also demonstrated consistent discriminatory signals present in the VOC data, as even simpler models like GLM showed significant, albeit lower, performance (Test AUC = 0.764). This multi-model approach enhances the credibility of our findings compared to studies relying on a single algorithm. Finally, our model's high performance (AUC = 0.941) suggests it has the potential to complement or even surpass traditional non-invasive markers like liver enzymes and lipid profiles, which may overlap between healthy and diseased states.

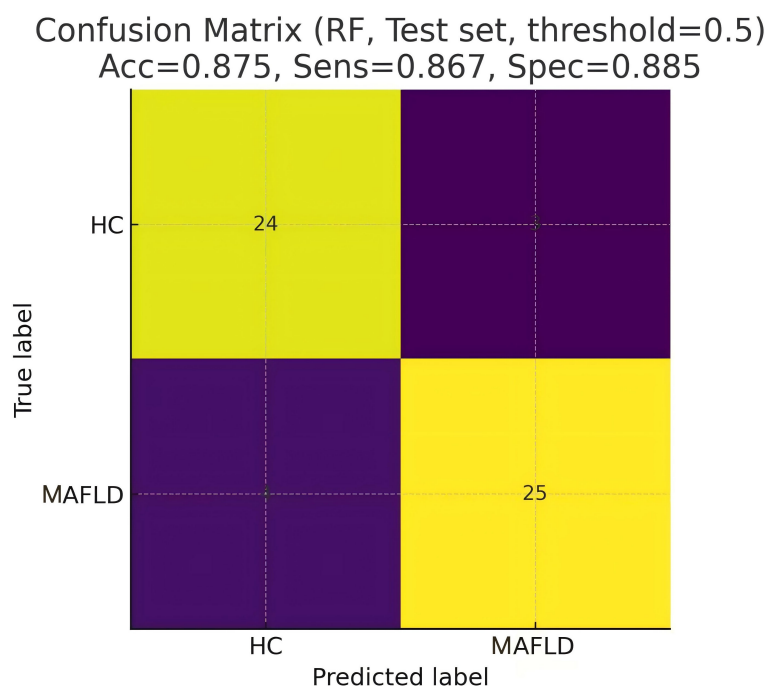


FIGURE 4
Confusion matrix of the RF model on the test set.

4.2 Biological and clinical interpretation of key VOCs

The VOC signature we identified reflects key metabolic hallmarks of MAFLD pathogenesis, including oxidative stress, disrupted lipid metabolism, and inflammation, as schematically illustrated in [Figure 8](#). Our feature importance analysis identified seven VOCs with the highest discriminatory power, and their putative biological origins align with these processes. Our feature importance analysis identified seven VOCs with the highest discriminatory power. The VOC signature we identified dovetails with the hallmarks of MAFLD—oxidative stress, disrupted lipid oxidation, and altered xenobiotic processing. Of these seven key VOCs, 2-butoxyethanol is a CYP-metabolized solvent derivative,

and its upregulation in MAFLD patients likely reflects enhanced hepatic detoxification activity in steatotic livers. The upregulation of cyclopentanone-D, a ketone byproduct of fatty-acid β -oxidation overload, reveals mitochondrial stress and lipid catabolism disruption. The downregulation of (*E*)-3-hexenoic acid and 2-ethylbutanal, both of which are α,β -unsaturated aldehydes formed during ω -6 lipid peroxidation, may indicate modified peroxidation flux or enhanced hepatic trapping/adduct formation. The downregulation of 2-propyl acetate and benzaldehyde-M, which are esters and aromatic aldehydes partly generated by the gut microbiome and amino acid metabolism, suggests dysbiosis and impaired hepatic clearance. Finally, the downregulation of furaneol, a furanone linked to Maillard/carbohydrate fermentation byproducts, may signal altered gut–liver crosstalk and glycation

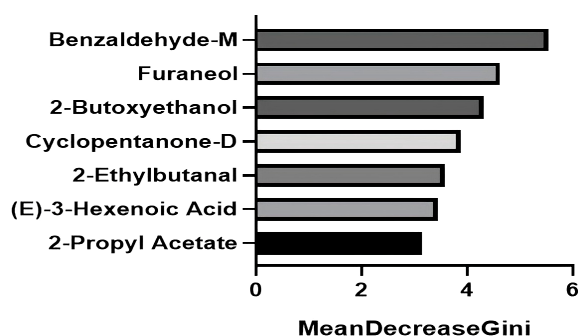


FIGURE 5
The top seven VOCs contributing most significantly to the predictive power of the random forest model. The bars represent the Gini coefficient calculated from the peak height of the respective VOC across all 199 patients. The seven selected VOCs all satisfy Gini coefficient ≥ 3 .

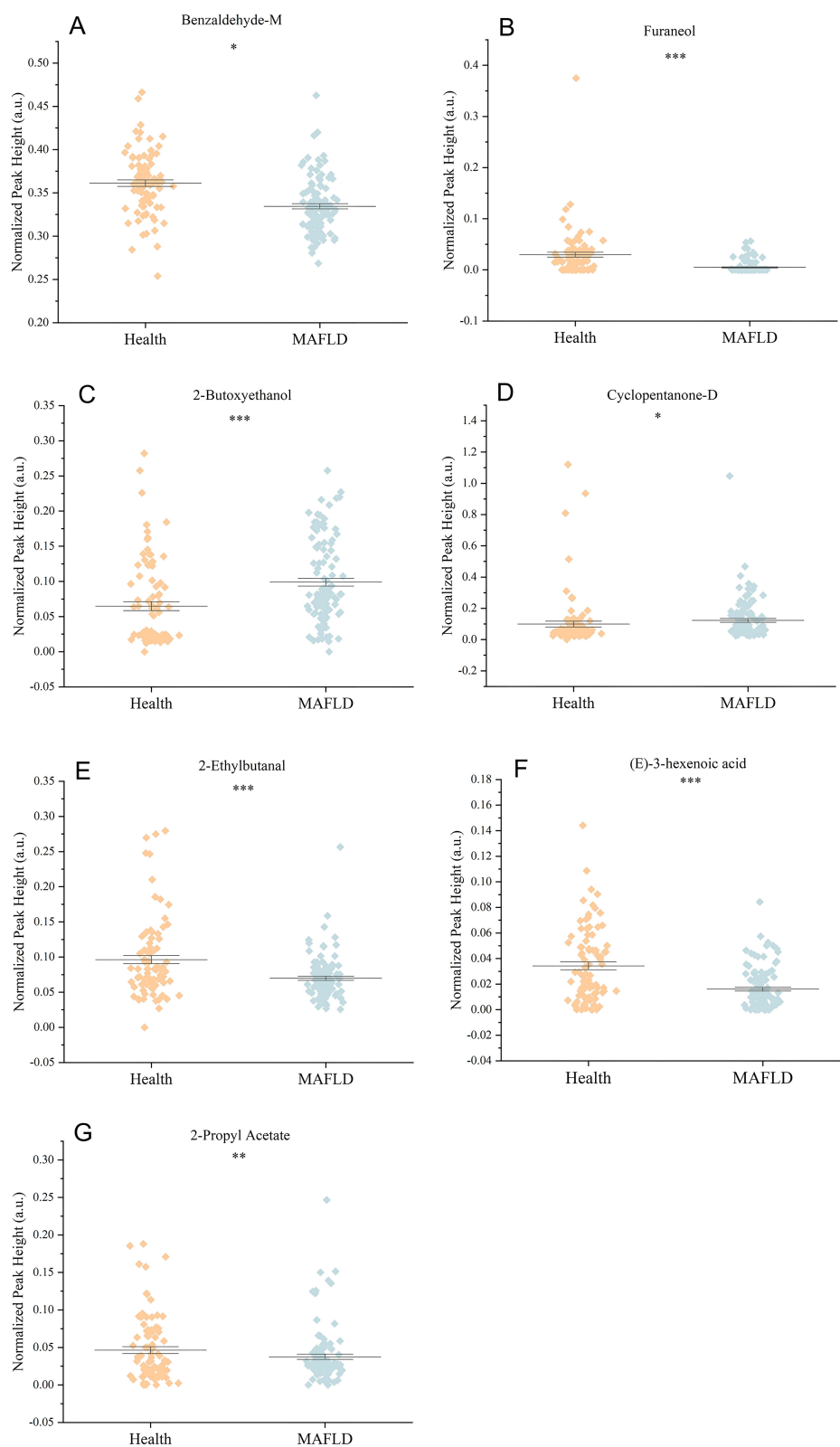


FIGURE 6

Comparison of the relative abundance (normalized peak height) of seven significantly altered VOCs between the Health and MAFLD groups.

(A) Benzaldehyde-M, (B) Furaneol, (C) 2-Butoxyethanol, (D) Cyclopentanone-D, (E) 2-Ethylbutanal, (F) (E)-3-Hexenoic acid, and (G) 2-Propyl acetate. Data are shown as mean $\pm$ SEM. *p < 0.05, **p < 0.01, ***p < 0.001.

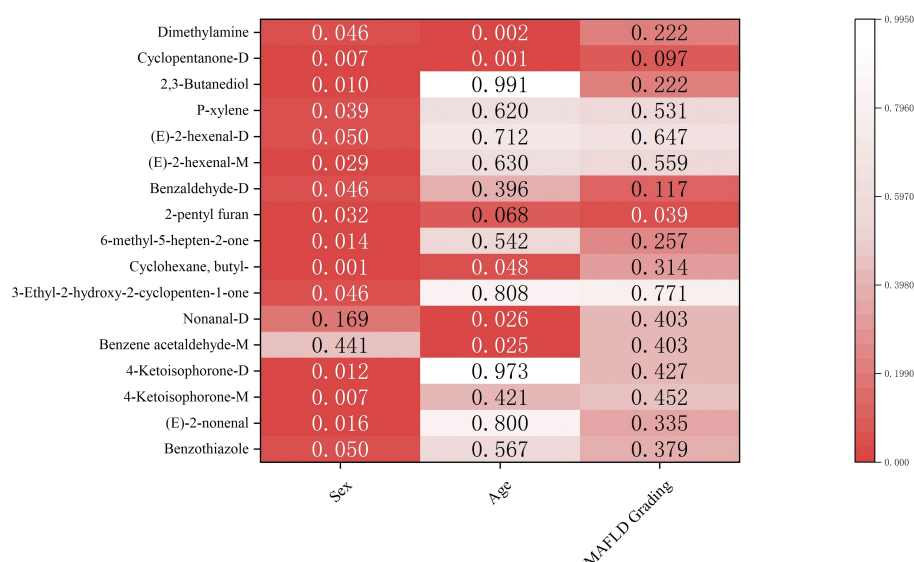


FIGURE 7

Associations between VOCs and pathological factors. Color intensity reflects $-\log_{10}(p\text{-value})$. Numbers colored in white denote statistically significant entries.

stress handling. Together, these patterns show that MAFLD involves not only lipid accumulation but also increased oxidative damage, impaired mitochondrial function, and changes in both host- and microbiome-derived volatile metabolites.

Interestingly, we observed significant associations between specific VOCs and demographic factors (Figure 6). Fifteen VOCs showed sex-specific differences, and six were correlated with age. For example, nonanal increased with age, which underscores cumulative lipid peroxidation and oxidative damage in MAFLD. Similarly, the sex-related difference in the ketone metabolite cyclopentanone-D may stem from hormonal regulation of mitochondrial fatty acid β -oxidation, thus suggesting that sex-specific metabolic pathways are involved in MAFLD. 2-Pentyl furan, highlighted in our analysis (Figure 8A) and known as an autooxidation product derived from linoleic acid (a polyunsaturated fatty acid), provides a direct link to lipid peroxidation and oxidative stress pathways implicated in MAFLD progression (Figure 8B). The most clinically relevant finding was its significant association with MAFLD severity graded by ultrasound. This correlation suggests that serum levels of 2-pentyl furan could potentially serve as a non-invasive biomarker for monitoring disease progression (Figure 8C) or stratifying patients based on severity, which would address a key unmet need in clinical practice. However, since we did not perform a parallel analysis of these demographic associations in the healthy control group, it remains unclear whether these effects are disease-specific or simply reflect baseline population variations.

4.3 Rationale for the machine learning approach

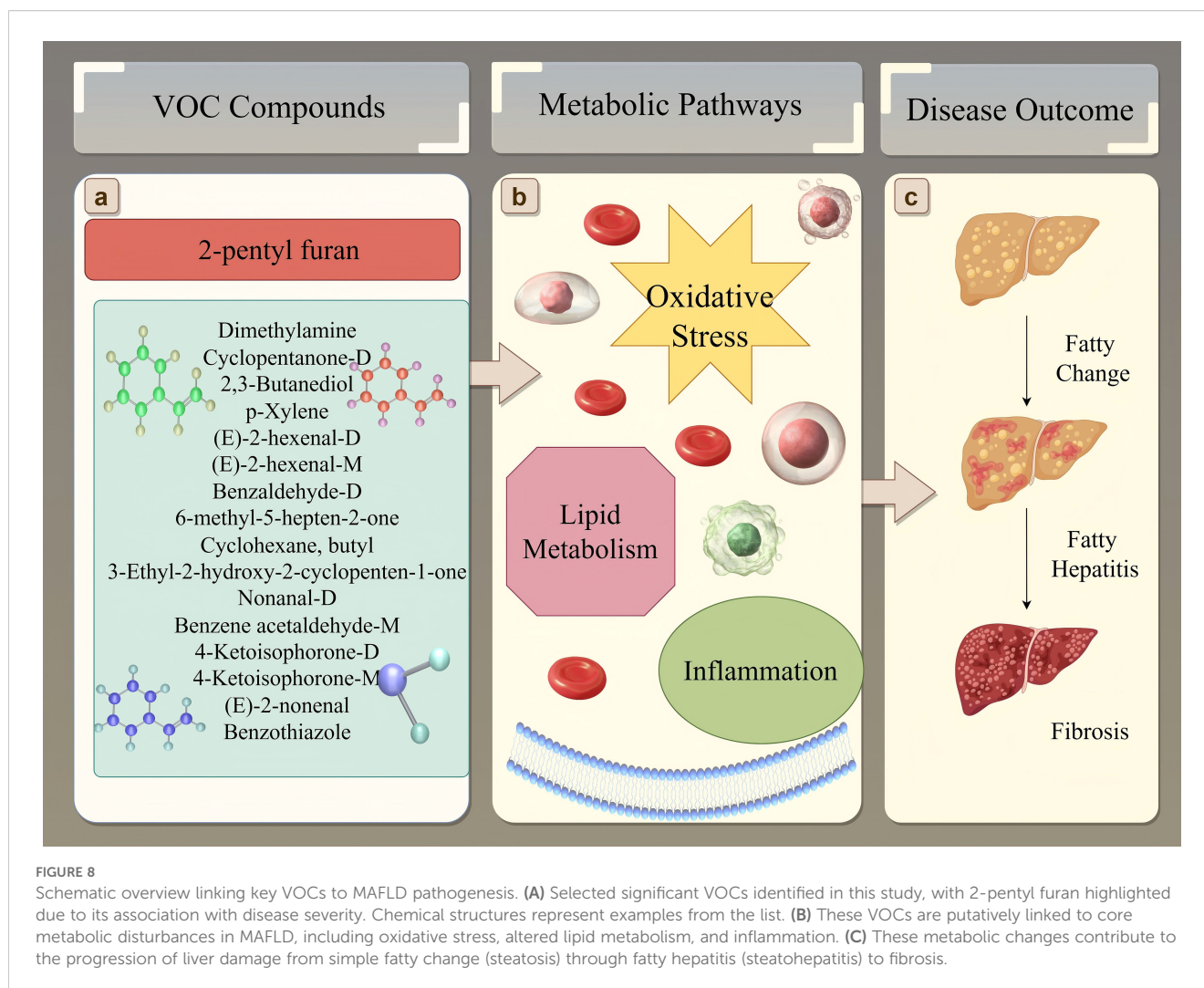
The superior performance of the RF model (Test AUC = 0.941) in our study can be attributed to its ability to handle high-

dimensional data and model complex, non-linear relationships between variables without overfitting (39). By building an ensemble of decision trees, RF effectively reduces noise and automatically selects the most important predictive features (40). In contrast, DT is prone to overfitting and had a lower performance (Test AUC = 0.878), although it is simple to interpret (41). The SVM model also performed well (Test AUC = 0.927), as it is well-suited for high-dimensional data with small sample sizes, although it can be computationally intensive (42). The strong performance of these advanced algorithms, compared to the simpler GLM (Test AUC = 0.764), highlights the importance of capturing non-linear interactions among VOCs for accurate MAFLD diagnosis.

4.4 Limitations and future directions

This study has several limitations. First, the single-center design and relatively small sample size ($n = 199$) limit the generalizability of our findings. Second, the cross-sectional design prevents us from establishing a causal relationship between VOC changes and MAFLD progression. Third, our diagnosis and staging of MAFLD were based on ultrasound rather than liver biopsy, the gold standard. This prevents us from distinguishing between simple steatosis and the more aggressive non-alcoholic steatohepatitis (NASH). Fourth, several significant VOCs could not be identified due to the limited library coverage of GC-IMS, which hinders a complete biological interpretation. Finally, the demographic analysis was restricted to the MAFLD group, and we did not fully disentangle disease-specific effects from normal population variance.

Building on this work, future research should focus on several key areas.



4.4.1 Clinical validation

Large-scale, multi-center, prospective studies are the highest priority. These studies should include diverse populations and use histopathology to confirm diagnoses, allowing for the validation of VOCs as biomarkers for both early detection and staging (i.e., steatosis vs. NASH vs. fibrosis). Longitudinal studies are also needed to track changes in serum VOCs over time in target populations.

4.4.2 Mechanistic insight

Mechanistic studies are needed to understand the biological origins of the key VOCs identified. Combining metabolomics with other omics data (e.g., transcriptomics, proteomics) can help elucidate specific pathways (e.g., lipid metabolism, inflammation, gut–liver axis signaling) that are disrupted in MAFLD.

4.4.3 Technological advancement

Continued optimization of VOC detection technologies and exploration of advanced machine learning algorithms, like deep learning, will further enhance model performance. Future studies can also benefit from integrating GC-IMS with GC-MS for more comprehensive compound annotation.

4.4.4 Multimodal diagnostics

Research should explore how to integrate VOC profiling with existing diagnostic tools (ultrasound, blood tests, FibroScan) to create a more powerful, multi-modal strategy for MAFLD management, as such an approach should improve diagnostic accuracy and support personalized treatment decisions.

5 Conclusions

In summary, this study demonstrates that a serum VOC signature, when analyzed by machine learning, can accurately distinguish patients with MAFLD from healthy controls. Our model demonstrated high diagnostic performance and identified several promising biomarkers potentially linked to disease severity. However, the findings are constrained by the study's single-center design and lack of histopathological validation. Further validation in large, prospective cohorts is required to confirm the clinical utility of this approach for the early detection and management of MAFLD.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Third Hospital of Shandong Province KYLL-2024005. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XL: Writing – original draft. XYZ: Visualization, Writing – original draft. RZ: Conceptualization, Writing – original draft. XWZ: Writing – review & editing, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Shandong Provincial Natural Science Foundation via the General Program under the project “Study of lncRNA OGFRP1 on the mechanism and early diagnosis of non-small cell lung cancer” (No. ZR220MH321) and by Biobase China under the commissioned project “Research on Key Technologies of Molecular POCT Systems in the Field of Infectious Diseases” (No. SLSYKYB2022010).

References

1. Eslam M, Sanyal AJ, George J. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. *Gastroenterology*. (2020) 158:1999–2014.e1. doi: 10.1053/j.gastro.2019.11.312
2. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. *J Hepatol*. (2020) 73:202–9. doi: 10.1016/j.jhep.2020.03.039
3. Gofton C, Upendran Y, Zheng MH, George J. MAFLD: How is it different from NAFLD? *Clin Mol Hepatol*. (2023) 29:S17–s31.
4. Guo X, Yin X, Liu Z, Wang J. Non-Alcoholic Fatty Liver Disease (NAFLD) Pathogenesis and Natural Products for Prevention and Treatment. *Int J Mol Sci*. (2022) 23. doi: 10.3390/ijms232415489
5. Rong L, Zou J, Ran W, Qi X, Chen Y, Cui H, et al. Advancements in the treatment of non-alcoholic fatty liver disease (NAFLD). *Front Endocrinol (Lausanne)*. (2022) 13:1087260. doi: 10.3389/fendo.2022.1087260
6. Pouwels S, Sakran N, Graham Y, Leal A, Pintar T, Yang W, et al. Non-alcoholic fatty liver disease (NAFLD): a review of pathophysiology, clinical management and effects of weight loss. *BMC Endocr Disord*. (2022) 22:63. doi: 10.1186/s12902-022-00980-1
7. Ghazanfar H, Javed N, Qasim A, Zacharia GS, Ghazanfar A, Jyala A, et al. Metabolic Dysfunction-Associated Steatohepatitis and Progression to Hepatocellular Carcinoma: A Literature Review. *Cancers (Basel)*. (2024) 16. doi: 10.3390/cancers16061214
8. Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. *Hepatology*. (2011) 54:1082–90. doi: 10.1002/hep.24452
9. Bujak R, Struck-Lewicka W, Markuszewski MJ, Kaliszan R. Metabolomics for laboratory diagnostics. *J Pharm Biomed Analysis*. (2015) 113:108–20. doi: 10.1016/j.jpba.2014.12.017

Acknowledgments

We are grateful to the participants for generously providing data. We also thank the staff members for their diligent work in data collection.

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 author(s) 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/fendo.2025.1691853/full#supplementary-material>

10. da Costa BRB, De Martinis BS. Analysis of urinary VOCs using mass spectrometric methods to diagnose cancer: A review. *Clin Mass Spectrom.* (2020) 18:27–37. doi: 10.1016/j.clinms.2020.10.004
11. Moura PC, Raposo M, Vassilenko V. Breath volatile organic compounds (VOCs) as biomarkers for the diagnosis of pathological conditions: A review. *BioMed J.* (2023) 46:100623. doi: 10.1016/j.bj.2023.100623
12. Zhou M, Wang Q, Lu X, Zhang P, Yang R, Chen Y, et al. Exhaled breath and urinary volatile organic compounds (VOCs) for cancer diagnoses, and microbial-related VOC metabolic pathway analysis: a systematic review and meta-analysis. *Int J Surg.* (2024) 110:1755–69. doi: 10.1097/JIS9.0000000000000999
13. Moura PC, Raposo M, Vassilenko V. Breath biomarkers in Non-Carcinogenic diseases. *Clin Chim Acta.* (2024) 552:117692. doi: 10.1016/j.cca.2023.117692
14. Gui X, Zhang X, Xin Y, Liu Q, Wang Y, Zhang Y, et al. Identification and validation of volatile organic compounds in bile for differential diagnosis of perihilar cholangiocarcinoma. *Clinica Chimica Acta.* (2023) 541. doi: 10.1016/j.cca.2023.117235
15. Bannaga AS, Tyagi H, Daulton E, Covington JA, Arasradnam RP. Exploratory Study Using Urinary Volatile Organic Compounds for the Detection of Hepatocellular Carcinoma. *Molecules.* (2021) 26. doi: 10.3390/molecules26092447
16. Einoch Amor R, Levy J, Broza YY, Vangravs R, Rapoport S, Zhang M, et al. Liquid Biopsy-Based Volatile Organic Compounds from Blood and Urine and Their Combined Data Sets for Highly Accurate Detection of Cancer. *ACS Sens.* (2023) 8:1450–61. doi: 10.1021/acssensors.2c02422
17. Shen CH, Huang RH, Li YK, Chu TW, Pei D. Using machine learning methods to investigate the role of volatile organic compounds in non-alcoholic fatty liver disease. *Front Mol Biosci.* (2025) 12:1631265. doi: 10.3389/fmolb.2025.1631265
18. Cozzolino R, De Giulio B, Martignetti A, Forte G, Dallio M, Romeo M, et al. Urinary volatile Organic compounds in non-alcoholic fatty liver disease (NAFLD), type two diabetes mellitus (T2DM) and NAFLD-T2DM coexistence. *Metabolomics.* (2022) 18:98. doi: 10.1007/s11306-022-01960-1
19. Skarysz A, Salman D, Eddleston M, Sykora M, Hunsicker E, Nailon WH, et al. Fast and automated biomarker detection in breath samples with machine learning. *PloS One.* (2022) 17:e0265399. doi: 10.1371/journal.pone.0265399
20. Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. *Nat Rev Gastroenterol Hepatol.* (2021) 18:835–56. doi: 10.1038/s41575-021-00502-9
21. Liu Q, Li S, Mao M, Gui X, Zhang Y, Zhao Y, et al. Serum-volatile organic compounds in the diagnostics of esophageal cancer. *Sci Rep.* (2024) 14:17722. doi: 10.1038/s41598-024-67818-9
22. Miao G, Wang Y, Wang B, Yu H, Liu J, Pan R, et al. Multi-omics analysis reveals hepatic lipid metabolism profiles and serum lipid biomarkers upon indoor relevant VOC exposure. *Environ Int.* (2023) 180:108221. doi: 10.1016/j.envint.2023.108221
23. Chen WY, Fu YP, Tu H, Zhong W, Zhou L. The association between exposure to volatile organic compounds and serum lipids in the US adult population. *Lipids Health Dis.* (2023) 22:129. doi: 10.1186/s12944-023-01895-z
24. Hu Y, Niu Z, Cao C, Gao J, Pan M, Cai Y, et al. Volatile organic compounds (VOC) metabolites in urine are associated with increased systemic inflammation levels, and smokers are identified as a vulnerable population. *Ecotoxicol Environ Saf.* (2024) 288:117398. doi: 10.1016/j.ecoenv.2024.117398
25. Li F, Zheng Y, Zhao C, Zhu J, Hang Y, Fang Y, et al. GC-IMS facilitates identification of carbapenem-resistant *Klebsiella pneumoniae* in simulated blood cultures. *AMB Express.* (2024) 14:40. doi: 10.1186/s13568-024-01708-1
26. Liu Q, Li S, Li Y, Yu L, Zhao Y, Wu Z, et al. Identification of urinary volatile organic compounds as a potential non-invasive biomarker for esophageal cancer. *Sci Rep.* (2023) 13:18587. doi: 10.1038/s41598-023-45989-1
27. Daulton E, Wicaksono AN, Tiele A, Kocher HM, Debernardi S, Crnogorac-Jurcevic T, et al. Volatile organic compounds (VOCs) for the non-invasive detection of pancreatic cancer from urine. *Talanta.* (2021) 221:121604. doi: 10.1016/j.talanta.2020.121604
28. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer. *Clin Genitourin Cancer.* (2019) 17:183–90. doi: 10.1016/j.clgc.2019.02.003
29. Bhatt A, Parsi MA, Stevens T, Gabbard S, Kumaravel A, Jang S, et al. Volatile organic compounds in plasma for the diagnosis of esophageal adenocarcinoma: a pilot study. *Gastrointest Endosc.* (2016) 84:597–603. doi: 10.1016/j.gie.2015.11.031
30. Mozdiak E, Wicaksono AN, Covington JA, Arasradnam RP. Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP). *Tech Coloproctol.* (2019) 23:343–51. doi: 10.1007/s10151-019-01963-6
31. P H, Rangarajan M, Pandya HJ. Breath VOC analysis and machine learning approaches for disease screening: a review. *J Breath Res.* (2023) 17. doi: 10.1088/1752-7163/acb283
32. Khoubnasabjafari M, Mogaddam MRA, Rahimpour E, Soleymani J, Saei AA, Jouyban A. Breathomics: Review of Sample Collection and Analysis, Data Modeling and Clinical Applications. *Crit Rev Anal Chem.* (2022) 52:1461–87. doi: 10.1080/10408347.2021.1889961
33. Liu W, Cao S, Shi D, Yu L, Qiu W, Chen W, et al. Single-chemical and mixture effects of multiple volatile organic compounds exposure on liver injury and risk of non-alcoholic fatty liver disease in a representative general adult population. *Chemosphere.* (2023) 339:139753. doi: 10.1016/j.chemosphere.2023.139753
34. De Vincentis A, Vespasiani-Gentilucci U, Sabatini A, Antonelli-Incalzi R, Picardi A. Exhaled breath analysis in hepatology: State-of-the-art and perspectives. *World J Gastroenterol.* (2019) 25:4043–50. doi: 10.3748/wjg.v25.i30.4043
35. Fu Q, Wu Y, Zhu M, Xia Y, Yu Q, Liu Z, et al. Identifying cardiovascular disease risk in the U.S. population using environmental volatile organic compounds exposure: A machine learning predictive model based on the SHAP methodology. *Ecotoxicol Environ Saf.* (2024) 286:117210. doi: 10.1016/j.ecoenv.2024.117210
36. Thomas JN, Roopkumar J, Patel T. Machine learning analysis of volatolomic profiles in breath can identify non-invasive biomarkers of liver disease: A pilot study. *PloS One.* (2021) 16:e0260098. doi: 10.1371/journal.pone.0260098
37. Patnaik RK, Lin YC, Agarwal A, Ho MC, Yeh JA. A pilot study for the prediction of liver function related scores using breath biomarkers and machine learning. *Sci Rep.* (2022) 12:2032. doi: 10.1038/s41598-022-05808-5
38. Maimaitiyiming R, Zhang H, Wang J, Wang L, Zhao L, Liu B, et al. A Novel Strategy for Mixed Jam Evaluation: Apparent Indicator, Sensory, Metabolomic, and GC-IMS Analysis. *Foods.* (2024) 13. doi: 10.3390/foods13071104
39. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* (2023) 24. doi: 10.1093/bib/bbad002
40. Zheng J, Lv X, Jiang L, Liu H, Zhao X. Development of a Pancreatic Fistula Prediction Model After Pancreaticoduodenectomy Based on a Decision Tree and Random Forest Algorithm. *Am Surg.* (2023). doi: 10.1177/00031348231158692
41. Huang JC, Guo Q, Li XH, Shi TQ. A comprehensive review on the application of neural network model in microbial fermentation. *Bioresour Technol.* (2025) 416:131801. doi: 10.1016/j.biortech.2024.131801
42. Hosseini MP, Hosseini A, Ahi K. A Review on Machine Learning for EEG Signal Processing in Bioengineering. *IEEE Rev BioMed Eng.* (2021) 14:204–18. doi: 10.1109/RBME.2020.2969915