Heliyon 10 (2024) e35178

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Revealing alcoholization-related volatile compounds and determining alcoholization indices in tobacco using GC-IMS coupled with chemometrics

Guangwei Xiao [a,1], Jianyu Ding [b,1], Shizhou Shao [a,**], Lin Wang [a], Lei Gao [a], Xiaohua Luo [a], Zhaozhao Wei [a], Xiaohong Tan [c], Jie Guo [c], Jiangjin Qian [b], Anhong Xiao [b], Jiahua Wang [b,*]

[a] China Tobacco Hubei Industrial Co., Ltd, Wuhan 430000, Hubei, China
[b] College of Food Science and Engineering, Wuhan Polytechnic University, Wuhan 430023, Hubei, China
[c] Hubei Tobacco Gold Leaf Compound Roasting CO., Ltd, Enshi Compound Roasting Plant, Enshi 445000, Hubei, China

ARTICLE INFO

Keywords:
Gas chromatography-ion mobility spectrometry (GC-IMS)
Chemometrics
Tobacco
Volatile compounds
Alcoholization index

ABSTRACT

Alcoholization is an integral part of tobacco processing and volatile compounds are key to assessing tobacco alcoholization. In this study, a total of 154 volatiles from nine categories were determined by gas chromatography-ion mobility spectrometry (GC-IMS) from four grades of tobacco, of which 114 were better identified. And then, the dynamic trends of volatile compounds with significant changes in tobacco alcoholization were analyzed. The relevant volatiles with the alcoholization indices (AIs) ($R > 0.8$) were screened as indicators of tobacco alcoholization. Cinnamyl isobutyrate, linolenic acid alcohol, propanoic acid-M and propanoic acid-D in all tobacco samples were highly correlated with the AIs and tended to increase during the alcoholization process. In addition, linear discriminant analysis (LDA), back-propagation neural network (BPNN) and random forest (RF) classifiers were constructed for discrimination of tobacco AIs. Three classifiers trained with a combination of 20 volatiles achieved satisfactory results with area under the curve (AUC) of 0.95 (LDA), 0.94 (BPNN) and 0.97 (RF), respectively. The RF classifier gained optimal accuracy of 100 % and 96.1 % for the training and test sets, respectively. The study confirmed that GC-IMS can be used to characterize the changes of volatile compounds in tobacco during alcoholization and combined with machine learning to achieve the determination of AIs. The results of the study may provide a new means for the tobacco industry to monitor the alcoholization process and determine the degree of alcoholization.

1. Introduction

Freshly harvested tobacco leaves are first roasted, leaf-punched, re-roasted, and flaked to produce flue-cured tobacco laminas (FCTL), which are commercial raw materials for the tobacco industry. Since the newly prepared FCTLs have the disadvantage of being

** Corresponding author.
* Corresponding author.
E-mail addresses: shaoshzh@hbtobacco.cn (S. Shao), w.jiahua@163.com (J. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e35178
Received 20 July 2024; Accepted 24 July 2024
Available online 25 July 2024

heavily green and irritating, they must be alcoholized to improve the availability and stability of tobacco quality and meet the raw material requirements for the processing chain [1,2]. Alcoholization of tobacco is an important part of raw material quality improvement in cigarette production [3]. The chemical composition of tobacco leaves undergoes slow changes under the combined effects of microorganisms, enzymes, and chemical actions, resulting in an alcoholization process that may take to 1–3 years [2]. Exploring the trends of quality changes during the alcoholization process of FCTL is currently a key and persistent topic in the tobacco industry.

FCTLs contain a large number of functional small molecules with biologically active, such as nicotine, cannabinol, and chlorogenic acid, as well as many macromolecules with nutritional value, such as proteins, cellulose, and carbohydrates with different molecular structures [4–7]. Importantly, smell, taste and sensation depend on the combination of compounds containing volatile flavors, which are directly related to the smoker's enjoyment and directly affect the consumer's purchase propensity [8–10]. During the alcoholization process, the macromolecules in FCTLs are involved in biochemical reactions, enzymes, and microbial activities, and are degraded into more flavorful substances to improve the flavor of tobacco. At present, the widely used natural alcoholization in the industry faces significant operational issues, such as the intractability of alcoholization cycle management, and the short duration of maintaining optimal alcoholization quality. The quality of FCTL shows a trend of increasing initially and then decreasing during the alcoholization process [1,2]. Therefore, it is of great significance to carry out research on the trend of volatile components of tobacco alcoholization in real warehouses, and to screen indicator compounds for determining the degree of alcoholization, which contributes to the determination of the optimal alcoholization period for FCTL.

Tobacco flavor is mainly determined by volatile and semi-volatile compounds known as neutral aroma components, including aldehydes, ketones, alcohols, esters (lactones), and olefins [11–13], which can be detected by sensors. However, accurate and rapid determination of the content of aroma components in FCTLs from different regions, varieties and parts is challenging. Conventional methods are very complex and are mainly based on liquid chromatography or gas chromatography combined with flame ionization detection for quantitative determination of small organic molecules [14–18]. These methods require complex sample pretreatment, long analysis time, high cost, and complicated data analysis, which hinder their practical application in the detection of aromatic compounds. Therefore, there is a need to develop more powerful techniques to fulfill the requirement of detecting all volatile compounds in FCTLs efficiently and accurately with a single detection device.

Currently, a technique that combines the advantages of gas chromatography and ion mobility spectrometry (GC-IMS) has attracted much attention because of its high sensitivity, short analysis time, high selectivity, and detection capability without sample pretreatment [19]. More importantly, GC-IMS can identify volatile monomers and dimers [16], which is more suitable for the analysis of tobacco samples with complex flavors. For example, GC-IMS was used to characterize the key odorants of green, pu-erh, oolong, and black teas [16,20–23], as well as the dynamic changes of aroma throughout the production process [20,24]. In tobacco applications, Budzyńsk et al. [25] used GC-IMS to analyze potentially hazardous substances in e-cigarettes and obtained good quantitative prediction performance. Wang et al. [26] used GC-IMS to comparatively study the volatile compounds of eight types of cigar filler tobaccos, and identified 84 nitrogen-containing and ketone compounds at high levels. Zhu et al. [27] found that the volatile compound contents of cigar tobacco from different regions of China varied considerably, revealing that the volatile compound contents may be affected by the geographic origin and were dominated by phenolic, pyrazine and aldehyde compounds. In addition, Qin et al. [6] used GC-IMS to compare the aroma differences between different tobacco samples, and through principal component analysis and similarity studies, the origin and grade of two types of tobacco could be distinguished. Similarly, the composition/content of aroma compounds of eight tobacco samples of different origins were analyzed by GC-IMS, 197 volatile aroma compounds were identified, and ion mobility spectra, differential ion mobility spectra, and fingerprinting spectra were also constructed for two-dimensional analyses [7]. Therefore, GC-IMS has a good discriminatory ability to identify small changes in volatile compounds in tobacco alcoholization, and in combination with machine learning methods, it can be used to evaluate tobacco alcoholization degree.

Searching for feature variables is one of the key aspects of building high-performance machine learning models. Machine learning models are relatively simple and usually use linear models, decision trees, and support vector machines, which require manual extraction of features from data. The Pearson's correlation coefficient [28] evaluates the strength of the relationship between two vectors based on the covariance matrix of the data, and is commonly used to explore a wide range of relationships between analyzed variables. To improve the overall accuracy of the model, Pearson's correlation analysis combined with heat map method is commonly used to optimize the combination of input variables [29,30]. Based on linear features, partial least squares discriminant analysis, linear discriminant analysis (LDA) are widely used to train classifiers for categorization and classification of food or agricultural products [30,31]. However, for complex dataset processing, some nonlinear machine learning methods are introduced to obtain better results, such as support vector machines (SVM), back-propagation neural network (BPNN) and random forest (RF). SVM maps the original feature space to a higher dimensional space through a mapping function [32], which turns the indivisible data in the original space into linearly divisible, and therefore, SVM have a unique advantage in high-dimensional pattern recognition. BPNN is the most representative neural network algorithm, which is a three-layer structure consisting of input, hidden and output layers, also known as a "black box" because it is difficult to explain how and why it produces a given output [33]. As an integrated learning algorithm, RF is a tree-predictor combination consisting of a large number of decision trees [34], and has demonstrated superior performance after combining multiple decision tree models working together as a base model [35]. Machine learning methods combined with analytical techniques (including sensor, chromatography, spectroscopy, etc.) have received increasing interest in recent years, and are widely used for grading and classification of food and agricultural products [36–39].

In this study, three varieties of FCTLs from different production areas were naturally alcoholized in different warehouses. The volatile components of FCTLs in the alcoholization cycle were characterized by GC-IMS technique. The specific objectives were to identify the characteristic volatile compounds in FCTLs, screen the alcoholization-related volatile compounds in all FCTLs as

alcoholization indicators, analyze the dynamic trends in volatile compounds during the alcoholization process, and train classifiers for discriminating of alcoholization degree. The results are expected to provide a basis for the tobacco industry to control the alcoholization process and determine the optimal alcoholization quality period.

2. Materials and methods

2.1. Materials and reagents

Twenty-milliliter headspace vials covered with 18-mm magnetic PTFE/silicone caps were purchased from Agilent Technologies Inc. (Palo Alto, CA, USA). Ortho-ketone C4–C9 as an external reference was obtained from Sinopharm Chemical Reagent Beijing Co. (Beijing, China).

2.2. Sample preparation

In the present study, three commercial tobacco varieties (Yunyan 87, Cuibi 1, Longjiang 911) were harvested in 2021 from four production areas in China (Table S1). Cultivar Yunyin 87 from Xuanen, Hubei and Malong, Yunnan were labeled YY-HB and YY-YN, respectively. Cuibi 1 from Sanming, Fujian was labeled CB-FJ, and Longjiang 911 from Mudanjiang, Heilongjiang was labeled LJ-HLJ. According to the enterprise standard QB/HNZY. GY1-2020, the harvested fresh tobacco leaves were prepared into FCTLs and stored in commercial warehouses in Wuhan (WH), Enshi (ES) and Xiangyang (XY) for natural alcoholization. The environmental parameters (temperature and relative humidity) of the three warehouses during alcoholization and sampling times are shown in Table S2. The FCTL samples were initially stored in the warehouse in August 2022, indexed as the baseline (month 0) for the alcoholization process, and subsequent sampling time points were 2, 5, 9, 12, 14, and 17 months after alcoholization. Since FCTLs are naturally fermented in commercial warehouses where temperature and humidity vary seasonally, and temperature and humidity are key factors affecting the alcoholization process. The alcoholization process is fast at high temperature environments. In order to obtain variability and representative samples, non-equal time intervals sampling method was chosen, i.e., samples were taken at two-month intervals at high temperatures, four-month intervals at low temperatures, and three-month intervals at medium temperatures (Table S2).

A tobacco box at a fixed location in the center of the warehouse was chosen as the sampling point. About 1 kg of tobacco was randomly collected from selected boxes in different warehouses at the specified sampling time, sealed in polyethylene bags and immediately transferred to cold storage (−20 °C) for future analysis. At the end of sampling, all FCTLs were cut into filaments (approximately 1.5 mm wide and 5 mm long), mixed and placed in glass vials and kept at room temperature for 4 h before being analyzed by GC-IMS. Samples were weighed at room temperature and three independent FCTL samples were prepared for each sampling.

2.3. GC-IMS analysis

2.3.1. GC-IMS determination

The GC-IMS analytical method was slightly modified from Qin et al. [6] and Zhu et al. [27]. A GC-IMS instrument (Flavourspec®, G. A.S, Dortmund, Germany) equipped with an autosampler (CTC Analytics AG, Zwingen, Switzerland) was used for the detection of volatile compounds in FCTLs. The FCTL samples (0.5000 g) were transferred directly into 20 mL headspace vials and subsequently incubated at 80 °C for 20 min at a stirring rate of 500 rpm. Then, 0.5 mL of the headspace solution was automatically injected into the wells via a heated syringe at 85 °C. Chromatographic separation was performed on an MXT-5 capillary column (15 m × 0.53 mm × 1.0 μm, Restek, Beijing, China), and the chromatographic column temperature was 60 °C. High-purity nitrogen (99.99 %) was used as the carrier gas/drift gas. The carrier gas flow rate was programmed as follows: 0–2 min, the flow rate of 2 mL/min; 2–10 min, 10 mL/min; 10–20 min, 100 mL/min; and 20–30 min, 150 mL/min. The flow rate of the drift gas was 75 mL/min and the temperatures of drift tube was maintained at 45 °C. Each sample was analyzed in triplicate.

2.3.2. Qualitative and quantitative analysis of volatile compounds

The retention indices (RI) of the volatile compounds were determined using ortho-ketone C4–C9 as an external reference. By comparing the RI and drift time in the GC–IMS library provided by G.A.S (Dortmund, Germany), the volatile compounds in the CFTL samples were identified. Finally, the intensities of the volatile compounds were analyzed according to the peak volumes of the selected signal peaks. Visual and quantitative comparisons of volatility differences between samples were performed using the Gallery plot plug-in. Two-dimensional top views and three-dimensional fingerprints were constructed using the Reporter plug-in.

2.4. Chemometrics and statistical analysis

2.4.1. AIs-related volatile compounds selection

Alcoholization indices (AI) responds to the degree of natural alcoholization was defined in this study as a variable related to alcoholization time, i.e., AIs 0, 1, 2, 3, 4, 5, and 6 represent the cumulative alcoholization time of 0, 2, 5, 9, 12, 14, and 17 months (Table S2), respectively, in the warehouse. It is important to note that the AI reflects the availability and quality stability of tobacco and shows a pattern of increasing and decreasing with increasing alcoholization time. The determination of AI can provide a basis for companies to choose the best quality period for processing tobacco.

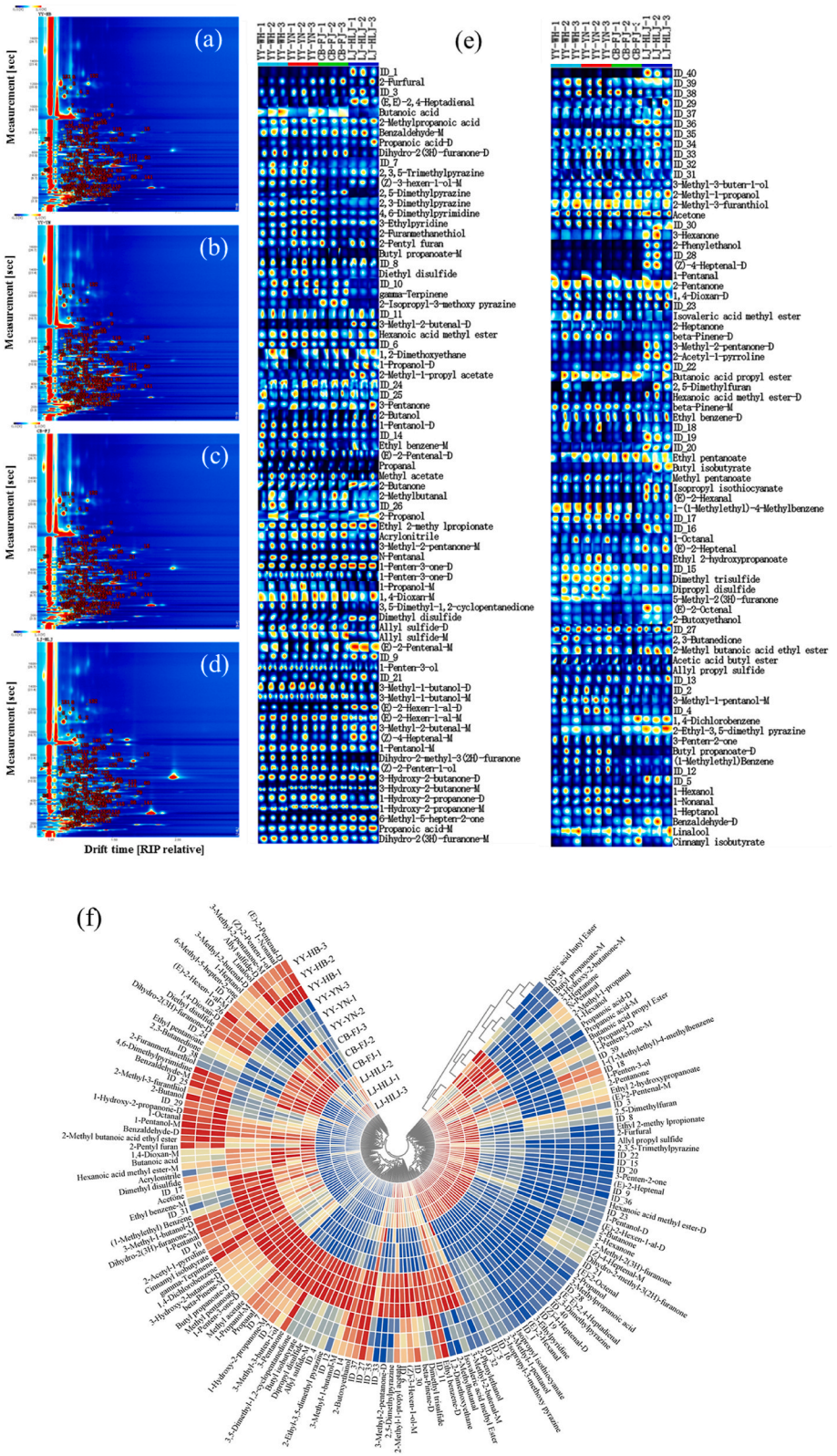

Fig. 1. 2D–topographic plots of volatiles detected by GC-IMS in YY-HB (a), YY-YN (b), CB-FJ (c) and LJ-HLJ (d) FCTLs; Fingerprints of volatile compounds identified by GC-IMS in four FCTLs (e); Heatmap of identified volatile compounds of FCTLs from four production regions (f). All peak volumes of volatile compounds were standardized on a scale of 0–1.

The relationship between the peak volume of volatile compounds and the AIs was determined by Pearson's correlation coefficient, which was then used in conjunction with a heat map to screen for AIs-related volatile compounds. Here, variables with high correlation are shown and those with low correlation are hidden, and subsequently the variables with high correlation are selected based on the correlation heat map. All these operations were performed in SPSS Statistics 26.0 software (SPSS Inc., Chicago, IL, USA) while the heatmaps were produced using TBtools-II software.

2.4.2. Machine learning methods

Machine learning methods such as LDA, BPNN and RF were proposed to train classifiers. The principles of LDA, BPNN and RF as well as the operational procedures were detailed in literatures [34,40,41]. These classifiers were trained in Matlab 2021a and the Statistics Toolbox (Mathworks Inc., Natick, MA, USA).

In this study, BPNN was a three-layer network structure, i.e., input layer (number of nodes is consistent with the selected feature volatiles), hidden layer (contains 6 neurons), and output layer (number of nodes is 6 consistent with the AI). The activation functions for the hidden and output layers of the BPNN were chosen as 'tansig' and 'softmax', respectively. The operational parameters of the BPNN were set to 1000 training rounds, a target error of 1×10^{-6} and a learning rate of 0.01. Hyperparameters were important in selecting an optimized RF algorithm, and the average prediction accuracy of 5-fold cross-validations of training sets was used as the selection standard, i. e., the maximum depth of each tree was 100, the minimum number of sample leaves was 1, and the number of decision trees was 50.

2.4.3. Classifier evaluation

In the present study, about 70 % of the samples were selected as the training set for training classifiers, while the remaining samples were used as the prediction set for model evaluation. The classifier outputs a confusion matrix presenting the number of true positive (TP), false negative (FN), true negative (TN), false positive (FP) samples. And then, the true positive rate (TPR) and the false positive rate (FPR) were calculated [30]. The performances of the classifiers were evaluated by obtaining the receiver operating characteristic (ROC) curves. A simple way of ROC curve evaluation is to estimate the area under the curve (AUC), with higher AUC values indicating better classification performance [42].

2.4.4. Statistical analysis

Analysis of significance between groups was assessed by one-way analysis of variance (ANOVA), implemented in SPSS Statistics 26.0 software (SPSS Inc., Chicago, IL, USA). Values of $p < 0.05$ were considered statistically significant. Dynamic change curves of volatile compounds during alcoholization process were plotted with Origin 2021 (OriginLab Corporation, Northampton, MA).

3. Results and discussion

3.1. Identification of volatile compounds in FCTLs raw materials

The volatile compounds of different FCTLs raw materials (unalcoholization) were determined using GC-IMS. The difference diagram of GC-IMS and representative three-dimensional fingerprints are shown in Fig. 1. The drift time of the reactant ions positive (RIP) peak was around 5.68 ms, and each point on either side of the RIP peak represented a volatile compound. In the GC-IMS spectra of FCTLs, the background was blue and the color of the points represented the intensity of the peak. The darker the color of the point, the higher the peak intensity.

As can be seen from Fig. 1(a–d), the volatile compounds of the three varieties of FCTLs from four production areas (of which YY87 was from two production areas) were well separated, with the retention time of most peaks being 144–1243 s and the drift time being 1.0–2.0 s. From the fingerprints of the FCTLs (Fig. 1(e)), it can be seen that the four FCTLs shared a considerable overlap in their volatile compounds. Table S2 lists the peak intensities of all identified volatile compounds that are present in all four FCTLs, and the significant difference analyses indicated that a larger number of volatile compounds also differed significantly ($p < 0.05$) between varieties or origins. A total of 154 volatile compounds were detected, including 20 alcohols, 20 aldehydes, 22 ketones, 17 esters, 7 ether, 4 acid, 5 benzene, 4 ethylene, 5 pyrazine, 10 others, and 40 are unknown substances (Table S2). The identified volatile compounds were similar to the results of another study, which detected 197 volatile aroma compounds in tobacco leaves, of which 75 were better identified [7]. From the heatmap of identified volatile compounds of four FCTLs (Fig. 1(f)), there was a significant difference between LJ-HLJ and YY-HB, YY-YN and CB-FJ, which could be caused by the fact that the annual average temperature in the LJ-HLJ planting region (Mudanjiang) was significantly lower than the other three regions, while the annual sunshine hours were significantly higher than the other regions (Table S1). The remaining three FCTLs exhibited closer similarities, with YY-HB occupying an intermediate position, some of whose components were more closely aligned with YY-YN, while others bore a stronger resemblance to CB-FJ (Fig. 1(f)).

Overall, both planting origins and tobacco varieties had a significant effect on the type and intensity of volatile compounds of FCTLs, and the effect of planting origins was relatively greater, mainly due to the average annual temperature and sunshine duration determination. These further confirmed the correlation between variety and origin and volatile compounds in tobacco [6,7,27]. Therefore, it is necessary to identify the indicative volatile compounds of FCTLs from different origins in natural alcoholization.

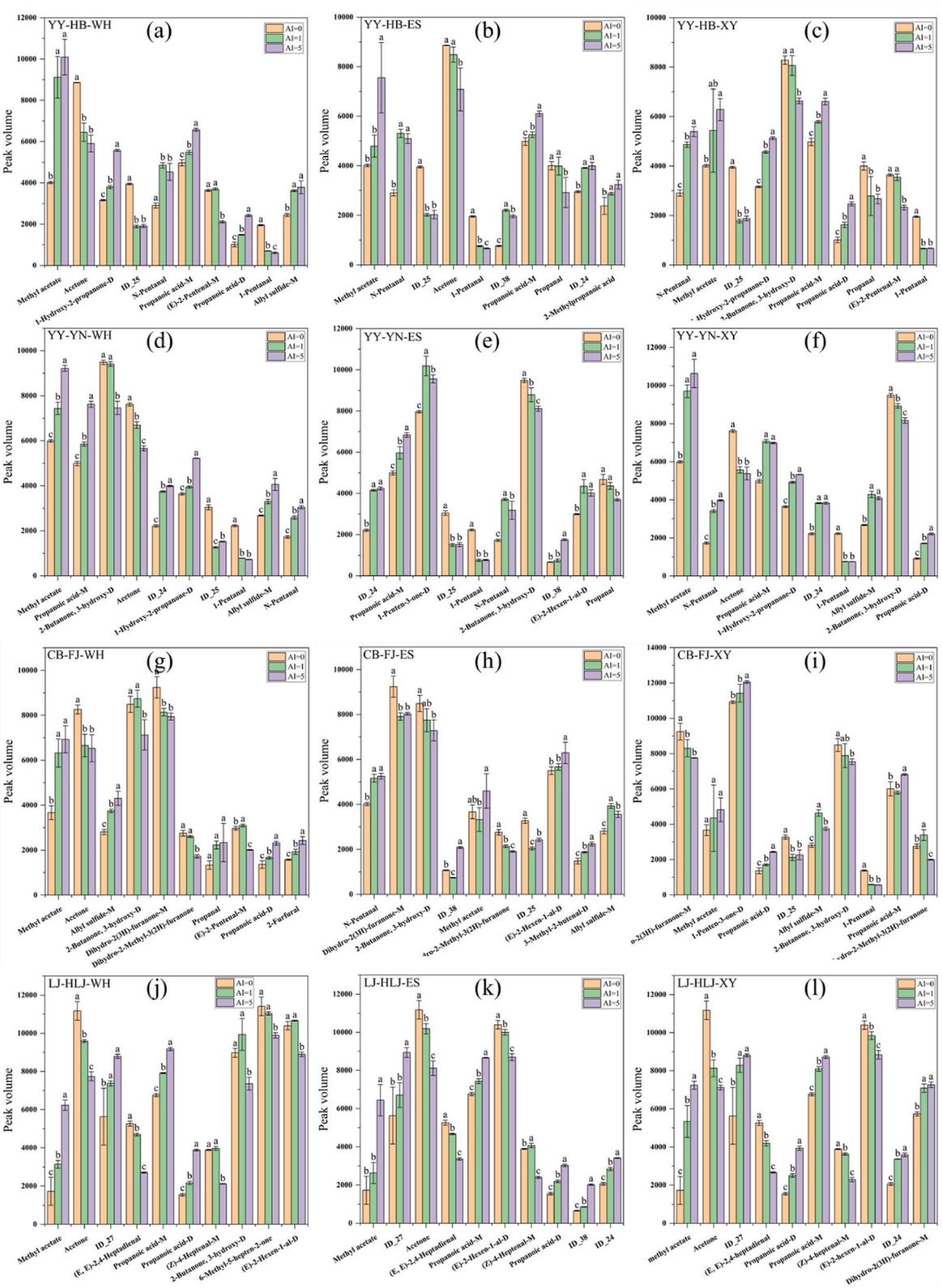

Fig. 2. Bar graphs of the top 10 volatile compounds with significant changes in peak volumes of (a–c) YY-HB, (d–f) YY-YN, (g–i) CB-FJ, and (j–l) LJ-HLJ stored in WH, ES, and XY warehouses, respectively.

3.2. Differences in main volatile compounds of FCTLs during natural alcoholization

The FCTLs with AIs of 0, 1 and 5, representing unalcoholized, after one summer (high temperature) and two summers of alcoholization, respectively, were used to explore the changes in major volatile compounds during natural alcoholization. Fig. S1 shows the GC-IMS difference mapping of the four FCTLs stored in the WH, ES and XY warehouses. In the difference spectra, volatile compounds with the same concentration had a white color bias. The red color indicated that the concentration of the volatiles was higher than the reference, while the blue color showed the opposite. The darker the color in the difference spectrum, the greater the difference in concentration [6,20]. There were more red dots in the differential spectrogram, indicating that more volatile substances were formed and accumulated to increase the flavor of FCTLs during the alcoholization process. Similarly, blue dots indicating that some substances were degraded by biochemical and microbial actions, promoting a more harmonious odor of tobacco [6,27]. For example, in Figs. S1 (a) and (c), the distribution of red and blue dots in the difference spectrogram with AI of 1 was similar to that with AI of 5, indicating that the volatile compounds of YY-HB and CB-FJ had a similar trend of change in alcoholization, with a faster rate of alcoholization in the early stage and a slower rate in the later stage. For the YY-YN and LJ-HLJ samples (Figs. S1(b) and (d)), there were relatively fewer dark dots in the difference spectrograms with an AI of 1 and more with an AI of 5, suggesting that these samples changed slowly in the early stage of alcoholization and faster in the later stage. Relatively more blue dots in the difference spectrograms of LJ-HLJ indicated the degradation of more volatile compounds in this variety (Fig. S1(d)). Similar phenomena were also found in the ES and XY libraries, and in comparison, the four RCTLs were more closely aligned with the alcoholization patterns in the WH and XY libraries (Figs. S1(e–l).

The volatile compounds corresponding to the darker dots (red and blue) were selected and the amount of change in peak volume was calculated for FCTLs with AIs of 0, 1 and 5, respectively. The top 10 volatile compounds were ranked according to the amount of change, and the top 10 compounds were analyzed for differences as shown in Fig. 2. For the same FCTL, there were significant differences among the top 10 compounds, confirming that the alcoholization of FCTL is a complex biochemical reaction process related to a variety of factors, such as the alcoholization environment, tobacco substrates and microorganisms [6,27]. Nonetheless, the same FCTL samples alcoholized in different warehouses had the same compounds with consistent trends. For example, methyl acetate, n-pentanal, and propanoic acid-M increeased while ID 25 and 1-pentanal decreasedin the YY-HB samples alcoholized at the WH, ES, and XY warehouses (Fig. 2(a–c)). For the alcoholized YY-YN sample, ID 24, propanoic acid-M, and n-pentanal accumulated, while 1-pentanal and 3-hydroxy-2-Butanone D decreased (Fig. 2(d–f)). For CB-FJ, methyl acetate increased while dihydro-2-methyl-3 (2H)-furanone decreased (Fig. 2(g–i)). During the alcoholization process, methyl acetate, ID 27, propanoic acid-M, and propanoic acid-D accumulated in sample LJ-HLJ, on the contrary acetone, (E, E)-2,4-heptadienal, (E)-2-hexen-1-al-D, and (Z)-4-heptenal-M were consumed, and the common pattern of changes in sample LJ-HLJ was most consistent among the three warehouses (Fig. 2(j–l)). Therefore, GC-IMS can be used as an effective tool for qualitative identification and quantitative characterization of volatile compounds in FCTL.

3.3. Screening of AIs-related volatile compounds in individual FCTL from different warehouses

The peak volumes of 154 determined volatile compounds were used to explore the relationship between compound concentration

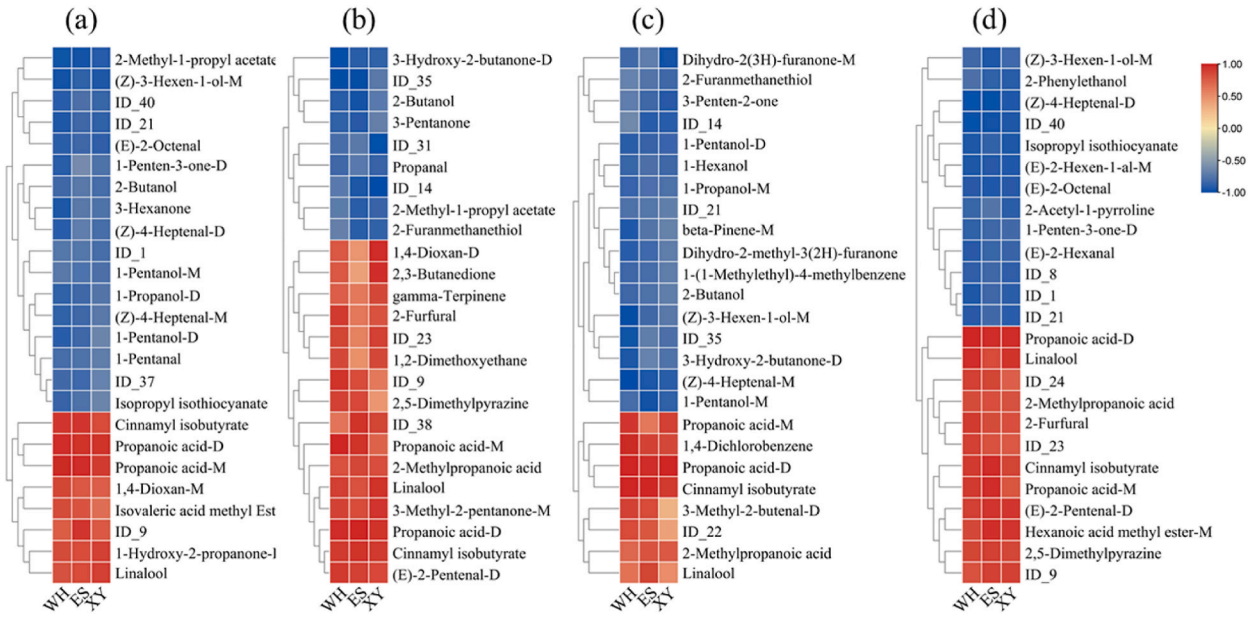

Fig. 3. Heat map of Pearson's correlation coefficients between peak volumes of volatile compounds and AIs of (a) YY-HB, (b) YY-YN, (c) CB-FJ, and (d) LJ-HLJ stored in WH, EN, and XY warehouses, respectively.

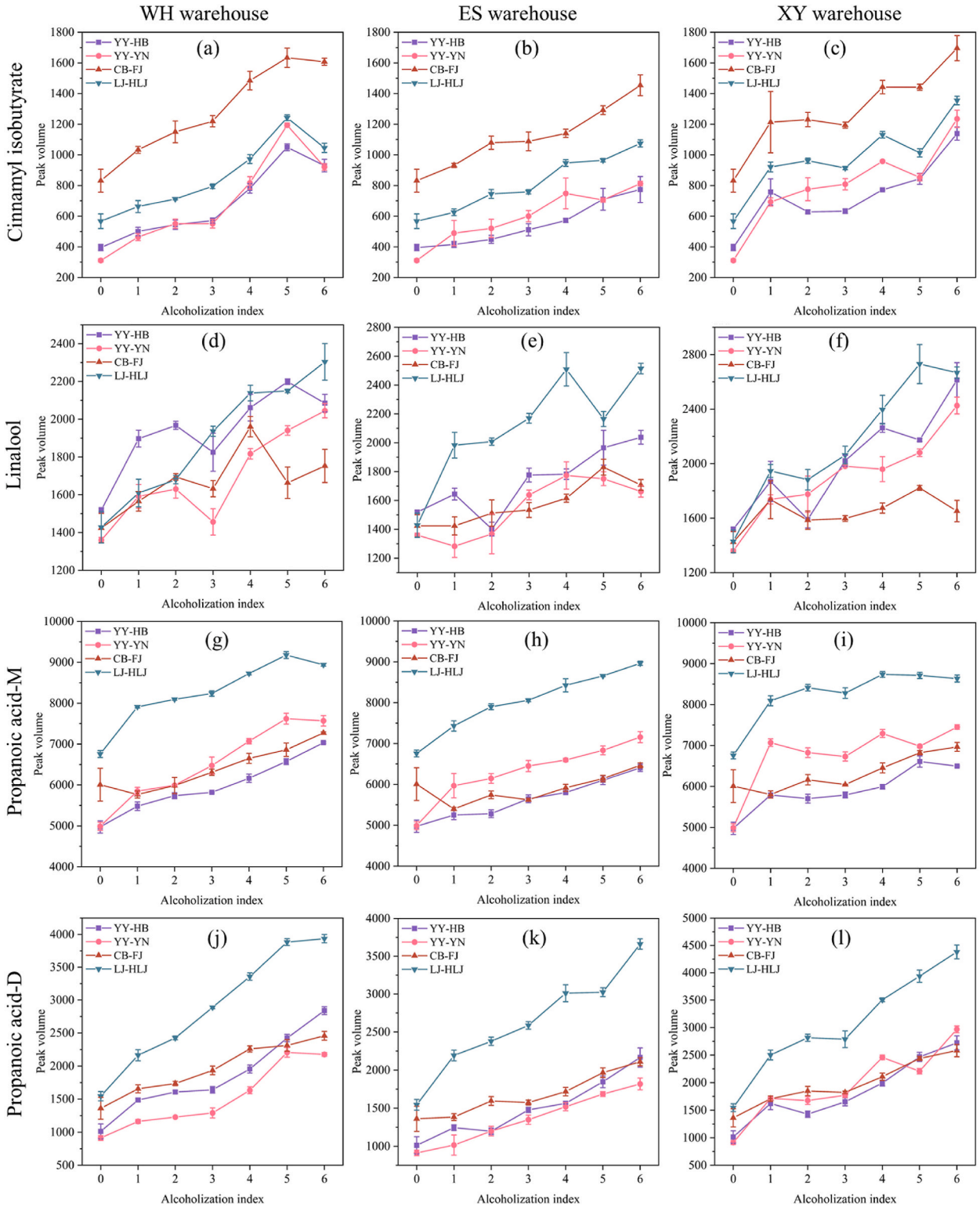

Fig. 4. Dynamic trends of peak volumes of cinnamyl isobutyrate (a–c), linalool (d–f), propanoic acid-M (g–i), and propanoic acid-D (j–l) in FCTLs accompanied by the AIs.

and AI. The Pearson's correlation coefficients between the peak volumes and AIs were calculated for FCTLs of YY-HB, YY-YN, CB-FJ and LJ-HLJ, respectively, and the clustered heat maps were obtained. In order to better understand the relationship between peak volume of volatile compounds and AI, the top 25 characteristic volatile compounds with high Pearson's correlation coefficients were selected for clustering heat map analysis, as shown in Fig. 3.

As can be seen from Fig. 3, for all FCTLs stored in different warehouses, the top 25 volatile components with high correlation were categorized into two groups. The red group indicated that volatile components were positively correlated with AIs, i.e., compounds accumulate during natural alcoholization. The blue group indicated that volatile compounds were negatively correlated with AIs, i.e., compounds break down during natural alcoholization. For individual FCTLs, the trends of change in volatile constituents in the WH, ES, and XY warehouses were consistent, despite some differences in temperature and humidity in the warehouses.

The volatile compounds with high Pearson's correlation coefficients ($R > 0.8$) in all three warehouses can be used as indicators of alcoholization for FCTLs. Thus, 10 volatile compounds were identified as indicators of alcoholization of YY-HB, i.e., propanoic acid-D, propanoic acid-M, 2-methyl-1-propyl acetate, (Z)-3-hexen-1-ol-M, cinnamyl isobutyrate, ID 21, (E)-2-octenal, 1-hydroxy-2-propanone-D, linalool, and ID 40 for YY-HB (Fig. 3(a)). Eight volatile compounds, propanoic acid-D, cinnamyl isobutyrate, linalool, propanoic acid-M, (E)-2-pentenal-D, 3-methyl-2-pentanone-M, 3-hydroxy-2-butanone-D, and 2-methylpropanoic acid, were used as indicators of YY-YN (Fig. 3(b)). Seven volatile compounds with correlation coefficients all higher than 0.8 were identified as CB-FJ indicators of alcoholization, such as propanoic acid-D, cinnamyl isobutyrate, 1,4-dichlorobenzene, (Z)-4-heptenal-M, 1-pentanol-M, 1-pentanol-D, and propanoic acid-M (Fig. 3(c)). In addition, propanoic acid-D, (Z)-4-heptenal-D, ID 40, (E)-2-pentenal-D, linalool, (E)-2-hexene-1-aldehyde-M, (E)-2-octenal, cinnamyl isobutyrate, methyl caproate-M, isopropyl isothiocyanate, and propanoic acid-M eleven compounds were screened as indicators of alcoholization for the LJ-HLJ (Fig. 3(d)). These volatile compounds were all screened from the same FCTLs and had similar trends in the alcoholization process across the three warehouses. Although there were obvious differences in the types of alcoholization indicators among the four FCTLs, four compounds including cinnamyl isobutyrate, linalool, propanoic acid-D, and propanoic acid-M, were present in all the FCTL groups and accumulated in alcoholization process (Fig. 3).

3.4. Dynamic trends of AIs-related volatile compounds during alcoholization

The trends of the four AIs-related volatile compounds in all FCTLs from different production areas during natural alcoholization are shown in Fig. 4. Four volatile compounds, cinnamyl isobutyrate, linalool, propanoic acid-D, and propanoic acid-M all increased in natural alcoholization, and were positively correlated with AIs (Fig. 3). Cinnamyl isobutyrate has a characteristic fruity and slight floral aroma, which has been detected in hookahs [11] and in cosmetics [43]. Linalool has a strong woody-green scent, resembling rosewood and green tea aroma, and is detected in almost all tobaccos [44,45], as well as in green tea leaves [24]. Propanoic acid-D, and propanoic acid-M are volatile organic acids found in tobacco smoke and are widely present in tobacco leaves and tobacco products [46].

Fig. 4(a–c) presents the dynamic trends of cinnamyl isobutyrate changes in the four FTCLs in WH, ES and XY warehouses, respectively. The highest content of cinnamyl isobutyrate was found in CB-FJ, followed by LJ-HLJ, then YY-YN and YY-HB, and the last two were similar, probably because they belonged to a single variety (YY 87) coming from different production areas, and the varietal factor played a dominant role. Among the three warehouses in the process of alcoholization, cinnamyl isobutyrate peaked earlier (corresponding to an AI of 5) in the WH warehouse (Fig. 4(a)), whereas the ES and XY warehouses peaked at an AI of 6 (Fig. 4(b) and (c)), and this may be due to the WH warehouse having a higher accumulation temperature. There was a significant increase in cinnamyl isobutyrate in all four FCTLs alcoholized in the three warehouses, with cumulative increases of 74.8%–103.9 % (CB-FJ), 84.3%–138.8 % (LJ-HLJ), 95.7%–188.1 % (YY-HB) and 161.9%–298.1 % (YY-YN), respectively.

The linalool content in all FTCLs raw materials was relatively consistent, while the largest amount of change was observed in LJ-HLJ during the alcoholization process, followed by YY-HB, YY-YN, and CB-FJ (Fig. 4(d–f)). The whole change trend was constantly fluctuating, with some variations in different warehouses. For example, out of LJ-HLJ, the other three FCTLs had a certain decrease in the AI of 3 (WH warehouse) and 2 (ES and XY), which was presumed to be related to the temperature change of the warehouse.

As can be seen from Fig. 4(g–l), the initial levels of propanoic acid-M and propanoic acid-D in LJ-HLJ were the highest, followed by CB-FJ, YY-YN and YY-HB, while the last two were similar. In different warehouses, the levels of propanoic acid-M and propanoic acid-D were significantly higher in LJ-HLJ than in the other three FCTLs throughout the alcoholization process. The average peak values of propanoic acid-M in FCTLs were 9176 (LJ-HLJ), 7602 (YY-YN), 7273 (CB-FJ), and 7034 (YY-HB) in the WH warehouse, 8964 (LJ-HLJ), 7157 (YY-YN), 6464 (CB-FJ), and 6404 (YY-HB) in the ES library, respectively, and 8736 (LJ-HLJ), 7293 (YY-YN), 6966 (CB-FJ), and 6605 (YY-HB) in the XY library, respectively (Fig. 4(g–i)). In comparison, the highest peak was observed in the WH warehouse, followed by the XY warehouse, and the lowest peak was observed in the ES warehouse, which may be related to the cumulative temperature of the warehouses.

During natural alcoholization, the propanoic acid-D content of FCTLs from all origins in the three warehouses showed an increasing trend, and the rate of propanoic acid-D accumulation in LJ-HLJ was higher than that in the other three origins. Propanoic acid-D increased by 154.8 % (WH warehouse), 137.1 % (ES warehouse), 183.7 % (XY warehouse) in LJ-HLJ; 180.3 % (Wuhan warehouse), 113.9 % (ES warehouse), 179.5 % (XY warehouse) in YY-HB; YY-YN increased by 141.6 % (WH warehouse), 99.1 % (ES warehouse), 224.8 % (XY warehouse); 80.5 % (WH warehouse), 54.9 % (ES warehouse), and 89.7 % (XY warehouse) in CB-FJ, respectively (Fig. 4(j–l)). Similar results were obtained with the changes of propanoic acid-M, and the increases of propanoic acid-M in WH and XY warehouses were significantly higher than those in ES warehouse.

3.5. Discrimination of AIs of FCTLs based on characteristic volatile compounds

A total of 252 FCTLs were used to calculate the correlation between AI and compound peak volume. The top 25 volatile compounds with the higher correlation were plotted on a correlation coefficient heat map (Fig. 5). The above 25 volatile compounds were ranked in descending order according to the relevance to AIs. The correlation coefficients of the compounds with AIs ranged from 0.44 to 0.68, with six compounds, 1-pentanal, propanoic acid-D, 1-pentanol-D, linalool, 3-penten-2-one, and cinnamyl isobutyrate, having correlation coefficients greater than 0.6. It can be seen that among these six compounds only propanoic acid-D and cinnamyl isobutyrate were selected in individual YY-HB, YY-YN, CB-FJ, LJ-HLJ with higher R above 0.9 (Fig. 5). Whereas 1-pentanal and 1-pentanol-D were present in some of FCTLs, 3-penten-2-one was not found in any of the individual FCTLs. This may be due to the fact that the correlation coefficients of these four compounds with AIs were small, whereas only the top 25 compounds are listed in Fig. 3. These further confirmed that there was a correlation between variety and origin and volatile compounds in tobacco [6,7,27].

The screened 25 volatile compounds were used as input variables for the model, and their peak volumes were used as variable values to construct classifiers of AIs. The importance of all variables was calculated first, and then some series of models were constructed after progressively removing the least contributing variables. The total discriminant accuracy was used to determine the best combination of variables for the model. After evaluation, a combination of 20 characteristic volatiles was determined to be used for training the classifiers and better results were obtained. The selected variables corresponding to the 20 volatile compounds, i.e., 1-pentanal, propanoic acid-D, 1-pentanol-D, linalool, 3-penten-2-one, cinnamyl isobutyrate, ID 35, 3-methyl-2-pentanone- M, ID 25, ID 21, ID 8, 2-butanol, propanoic acid-M, 2-furfural, (E)-2-pentenal-D, hexanoic acid, methyl ester-M, acetone, (E)-2-octenal, and 1,2-dimethoxyethane, can be used as characterization compounds for the identification of AIs. These 20 volatiles were mainly categorized as alcohols, ketones, aldehydes, and esters, which are important for the smokeability of tobacco. Alcohols act as moisturizers during smoking and also improve the aroma of tobacco [47]. Ketones have a strong influence on the taste, aroma and satisfaction of tobacco, harmonizing aroma, masking miscellaneous gas, and giving cigarettes different aroma profiles [48,49]. Aldehydes are naturally

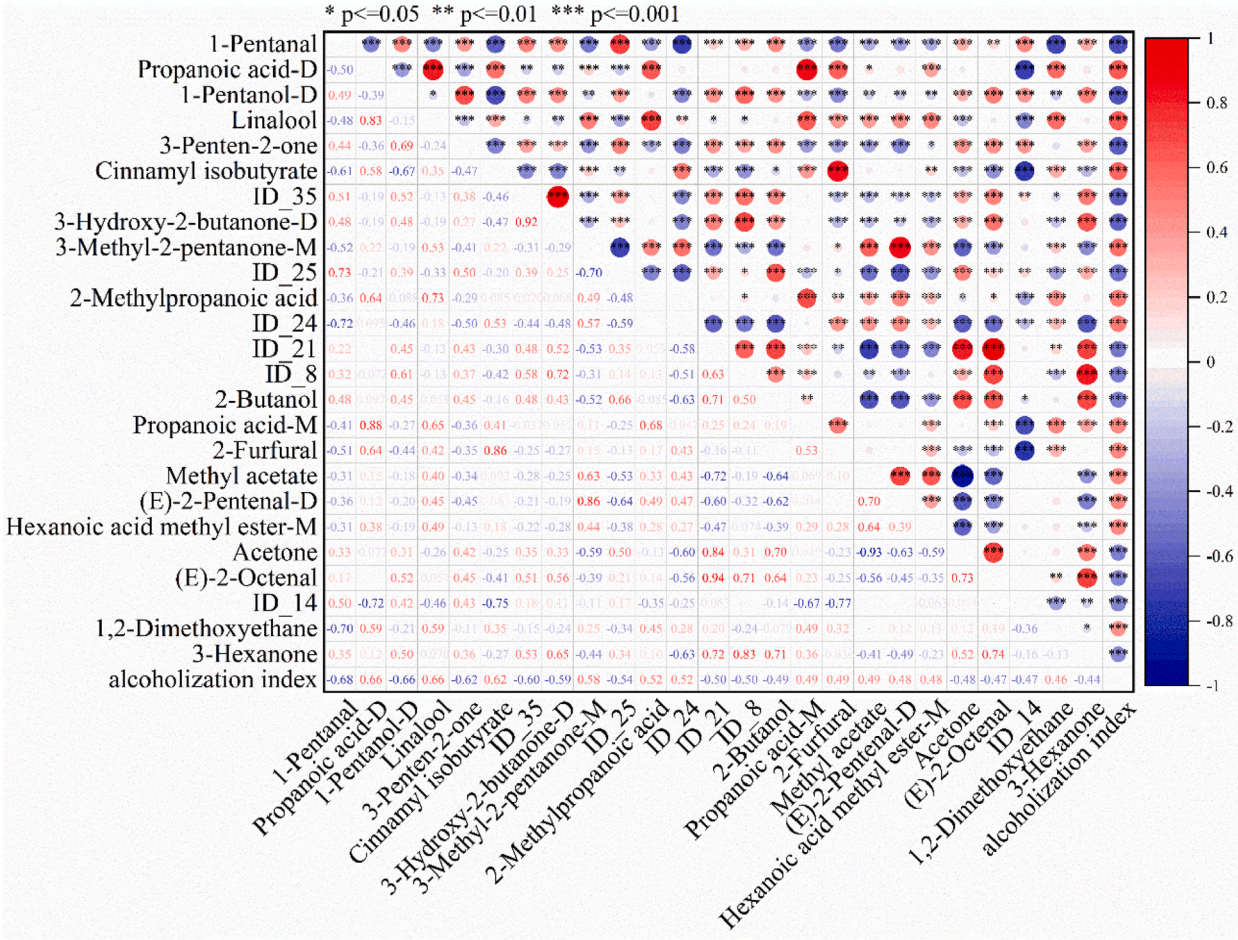

Fig. 5. Heat map of the correlations between volatile compounds and AIs. Red and blue colors indicate positive and negative correlations, respectively. Darker colors and larger circles indicate larger correlation coefficients, while the opposite indicates smaller correlation coefficients. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

present in plant essential oils and have rose and citrus-like aromas [50], whereas esters have pleasant fruity and toasty aromas, enhance the fullness of the tobacco aroma [51]. Therefore, the use of these volatiles to construct a classifier to identify the AIs of tobacco can guide the control of the alcoholization process and the appropriate time of release of tobacco.

The average ROC curves and AUC scores of the LDA, BPNN and RF are shown in Fig. 6. Three machine learning classifiers performed well in determining AIs with AUCs of 0.95 (LDA), 0.94 (BPNN) and 0.97 (RF). The RF algorithm introduces stochasticity relative to LDA and BPNN, which is less likely to be overfitted and has good noise immunity [34]. The confusion matrices of the three classifiers are shown in Fig. S2. The optimal RF model was obtained using the combination of 20 variables, resulted in the accuracy of 100 % and 96.1 % for the training and prediction sets, respectively; two samples with AI of 2 were misclassified as 3 and one sample with AI of 3 was misclassified as 4 in the prediction set (Fig. S2). According to the above results, the GC-IMS combined with RF can accurately classify FCTLs on the basis of the AI.

4. Conclusions

In this study, the changes of volatile compounds during natural alcoholization of FCTLs from different sources in different warehouses were investigated, and indicators to determine the Ais were selected. A total of 154 characteristic volatile compounds were detected by GC-IMS, among which 114 volatiles belonging to 9 categories were well identified. The concentration of volatile compounds varied significantly with the variety and origin of the FCTLs. By Pearson correlation analysis, 10, 8, 7 and 11 volatile compounds were selected as alcoholization indicators for YY-HB, YY-YN, CB-FJ and LJ-HLJ, respectively, and these compounds showed high correlation coefficients with AIs ($R > 0.8$). Four compounds, i.e., cinnamyl isobutyrate, linalool, propanoic acid-D and propanoic acid-M were present in FCTLs from all production areas and showed an increasing trend during natural alcoholization, which was positively correlated with AIs. In addition, the accumulation of the four volatile compounds was correlated with the cumulative temperature of the warehouse. The RF classifier trained using the combination of 20 volatile compounds achieved optimal results with recognition rates of 100 % and 96.1 % for the training and prediction sets, respectively. Since tobacco is a natural product, samples from different years and more production areas need to be collected for methodological validation, and to expand the dataset and improve the adaptability and accuracy of the model. These results provide new insights into the dynamics of volatile compounds during the natural alcoholization, thus providing new ideas for the control of alcoholization process and intelligent discrimination of AIs in the tobacco industry.

Funding statement

The work was supported by the China Tobacco Hubei Industrial Co., Ltd (2022JCZL3WL2B048).

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Guangwei Xiao: Conceptualization, Methodology, Writing – original draft. **Jianyu Ding:** Data curation, Investigation, Methodology, Writing – original draft. **Shizhou Shao:** Funding acquisition, Resources, Supervision, Writing – review & editing. **Lin Wang:** Formal analysis, Methodology. **Lei Gao:** Investigation, Validation. **Xiaohua Luo:** Investigation, Validation. **Zhaozhao Wei:** Data

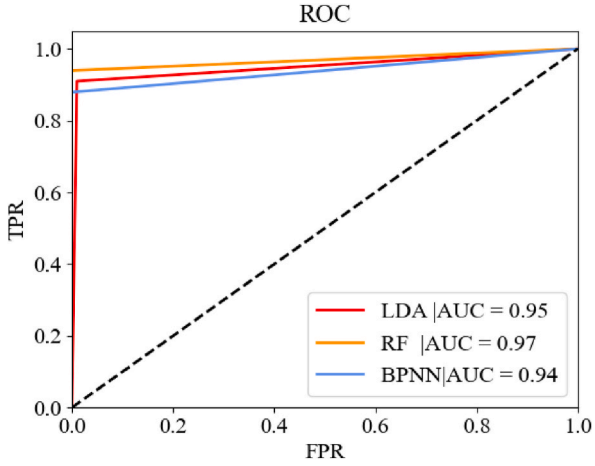

Fig. 6. ROC curves and AUC scores for machine learning classifiers based on the combination of 20 characteristic volatile compounds.

curation, Investigation. **Xiaohong Tan:** Investigation, Resources. **Jie Guo:** Investigation, Resources. **Jiangjin Qian:** Software. **Anhong Xiao:** Project administration, Visualization. **Jiahua Wang:** Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35178.

References

[1] Y. Ning, L.Y. Zhang, J. Mai, J.E. Su, J.Y. Cai, Y. Chen, B.B. Hu, Tobacco microbial screening and application in improving the quality of tobacco in different physical states, Bioresources and Bioprocessing 10 (1) (2023) 32, https://doi.org/10.1186/s40643-023-00651-6.
[2] C. Wen, Q. Zhang, P. Zhu, W. Hu, Y. Jia, S. Yang, Y. Huang, Z. Yang, Z. Chai, T. Zhai, others, High throughput screening of key functional strains based on improving tobacco quality and mixed fermentation, Front. Bioeng. Biotechnol. 11 (2023) 1108766, https://doi.org/10.3389/fbioe.2023.1108766.
[3] L. Yao, D. Li, C. Huang, Y. Mao, Z. Wang, J. Yu, X. Chen, Screening of cellulase-producing bacteria and their effect on the chemical composition and aroma quality improvement of cigar wrapper leaves, Bioresources 17 (1) (2022) 1566–1590, https://doi.org/10.15376/biores.17.1.1566-1590.
[4] T. Fujimori, R. Kasuga, H. Kaneko, M. Noguchi, Neutral volatile components of burley tobacco, Contributions to Tobacco & Nicotine Research 9 (5) (1978) 317–325, https://doi.org/10.2478/cttr-2013-0950.
[5] R.S. Hu, J. Wang, H. Li, H. Ni, Y.F. Chen, Y.W. Zhang, S.P. Xiang, H.H. Li, Simultaneous extraction of nicotine and solanesol from waste tobacco materials by the column chromatographic extraction method and their separation and purification, Separ. Purif. Technol. 146 (2015) 1–7, https://doi.org/10.1016/j.seppur.2015.03.016.
[6] G. Qin, G. Zhao, C. Ouyang, J. Liu, Aroma components of tobacco powder from different producing areas based on gas chromatography ion mobility spectrometry, Open Chem. 19 (1) (2021) 442–450, https://doi.org/10.1515/chem-2020-0116.
[7] T. Yan, P. Zhou, F. Long, J. Liu, F. Wu, M. Zhang, J. Liu, Unraveling the difference in the composition/content of the aroma compounds in different tobacco leaves: for better use, J. Chem. (2022) 1–10, https://doi.org/10.1155/2022/3293899, 2022.
[8] S. Dagnon, R. Tasheva, A. Stoilova, D. Christeva, A. Edreva, Evaluation of aroma in oriental tobaccos as based on valeric acid gas chromatography, Contributions to Tobacco & Nicotine Research 23 (2) (2008) 115–120, https://doi.org/10.2478/cttr-2013-0854.
[9] V. Popova, T. Ivanova, T. Prokopov, M. Nikolova, A. Stoyanova, V.D. Zheljazkov, Carotenoid-related volatile compounds of tobacco (Nicotiana tabacum L.) essential oils, Molecules 24 (19) (2019) 3446, https://doi.org/10.3390/molecules24193446.
[10] F. Liao, Y. Li, W. He, J. Tie, X. Hao, Y. Tian, S. Li, L. Zhang, L. Tang, J. Wu, others, Evaluation of aroma styles in flue-cured tobacco by near infrared spectroscopy combined with chemometric algorithms, J. Near Infrared Spectrosc. 28 (2) (2020) 93–102, https://doi.org/10.1177/0967033519898892.
[11] M.A. Farag, M.M. Elmassry, S.H. El-Ahmady, The characterization of flavored hookahs aroma profile and in response to heating as analyzed via headspace solid-phase microextraction (SPME) and chemometrics, Sci. Rep. 8 (1) (2018) 17028, https://doi.org/10.1038/s41598-018-35368-6.
[12] D. Nedeltcheva-Antonova, D. Ivanova, L. Antonov, I. Abe, Insight into the aroma profile of Bulgarian tobacco absolute oil, Ind. Crop. Prod. 94 (2016) 226–232, https://doi.org/10.1016/j.indcrop.2016.08.047.
[13] X. Li, J. Bin, X. Yan, M. Ding, M. Yang, Application of chromatographic technology to determine aromatic substances in tobacco during natural fermentation: a review, Separations 9 (8) (2022) 187, https://doi.org/10.3390/separations9080187.
[14] Y. Ding, L. Zhu, S. Liu, H. Yu, Y. Dai, Analytical method of free and conjugated neutral aroma components in tobacco by solvent extraction coupled with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, J. Chromatogr. A 1280 (2013) 122–127, https://doi.org/10.1016/j.chroma.2013.01.028.
[15] Q. He, Y. Zhang, S. Zhou, S. She, G. Chen, K. Chen, Z. Yan, D. Guo, Estimating the aroma glycosides in flue-cured tobacco by solid-phase extraction and gas chromatography–mass spectrometry: changes in the bound aroma profile during leaf maturity, Flavour Fragrance J. 30 (3) (2015) 230–237, https://doi.org/10.1002/ffj.3235.
[16] L. Liu, X. Wang, S. Wang, S. Liu, Y. Jia, Y. Qin, H. Cui, L. Pan, H. Liu, Simultaneous quantification of ten Amadori compounds in tobacco using liquid chromatography with tandem mass spectrometry, J. Separ. Sci. 40 (4) (2017) 849–857, https://doi.org/10.1002/jssc.201601168.
[17] K. Mitsui, F. David, E. Dumont, N. Ochiai, H. Tamura, P. Sandra, LC fractionation followed by pyrolysis GC–MS for the in-depth study of aroma compounds formed during tobacco combustion, J. Anal. Appl. Pyrol. 116 (2015) 68–74, https://doi.org/10.1016/j.jaap.2015.10.004.
[18] L. Wu, Q. Li, C. Li, J. Cao, Y. Lai, K. Qiu, S. Min, Determination of aroma components in Chinese southwest tobacco by directly suspended droplet microextraction combined with GC–MS, J. Chromatogr. Sci. 52 (10) (2014) 1317–1325, https://doi.org/10.1093/chromsci/bmt170.
[19] S. Gu, J. Zhang, J. Wang, X. Wang, D. Du, Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products, TrAC, Trends Anal. Chem. 144 (2021) 116435, https://doi.org/10.1016/j.trac.2021.116435.
[20] J. Xie, L. Wang, Y. Deng, H. Yuan, J. Zhu, Y. Jiang, Y. Yang, Characterization of the key odorants in floral aroma green tea based on GC-E-Nose, GC-IMS, GC-MS and aroma recombination and investigation of the dynamic changes and aroma formation during processing, Food Chem. 427 (2023) 136641, https://doi.org/10.1016/j.foodchem.2023.136641.
[21] X. Guo, W. Schwab, C.-T. Ho, C. Song, X. Wan, Characterization of the aroma profiles of oolong tea made from three tea cultivars by both GC–MS and GC-IMS, Food Chem. 376 (2022) 131933, https://doi.org/10.1016/j.foodchem.2021.131933.
[22] Y. Rong, J. Xie, H. Yuan, L. Wang, F. Liu, Y. Deng, Y. Jiang, Y. Yang, Characterization of volatile metabolites in Pu-erh teas with different storage years by combining GC-E-Nose, GC–MS, and GC-IMS, Food Chem. X 18 (2023) 100693, https://doi.org/10.1016/j.fochx.2023.100693.
[23] H. Liu, Y. Xu, J. Wu, J. Wen, Y. Yu, K. An, B. Zou, GC-IMS and olfactometry analysis on the tea aroma of Yingde black teas harvested in different seasons, Food Res. Int. 150 (2021) 110784, https://doi.org/10.1016/j.foodres.2021.110784.
[24] Y. Yang, M.C. Qian, Y. Deng, H. Yuan, Y. Jiang, Insight into aroma dynamic changes during the whole manufacturing process of chestnut-like aroma green tea by combining GC-E-Nose, GC-IMS, and GC\times GC-TOFMS, Food Chem. 387 (2022) 132813, https://doi.org/10.1016/j.foodchem.2022.132813.
[25] E. Budzyńska, S. Sielemann, J. Puton, A.L.R.M. Surminski, Analysis of e-liquids for electronic cigarettes using GC-IMS/MS with headspace sampling, Talanta 209 (2020) 120594, https://doi.org/10.1016/j.talanta.2019.120594.
[26] J. Wang, Y. Pan, L. Liu, C. Wu, Y. Shi, X. Yuan, Identification of key volatile flavorcompounds in cigar filler tobacco leaves via GC-IMS, Asian Journal of Agriculture and Biology 3 (2023) 2023013, https://doi.org/10.35495/ajab.2023.013.

[27] B. Zhu, H. An, L. Li, H. Zhang, J. Lv, W. Hu, F. Xue, L. Liu, S. He, D. Li, Characterization of flavor profiles of cigar tobacco leaves grown in China via headspace–gas chromatography–ion mobility spectrometry coupled with multivariate analysis and sensory evaluation, ACS Omega 9 (14) (2024) 15996–16005, https://doi.org/10.1021/acsomega.3c09499.
[28] Y. Mu, X. Liu, L. Wang, A Pearson's correlation coefficient based decision tree and its parallel implementation, Inf. Sci. 435 (2018) 40–58, https://doi.org/10.1016/j.ins.2017.12.059.
[29] J. Wang, J. Lv, T. Mei, M. Xu, C. Jia, C. Duan, F. Pi, Spectroscopic studies on thermal degradation and quantitative prediction on acid value of edible oil during frying by Raman spectroscopy, Spectrochim. Acta Mol. Biomol. Spectrosc. 293 (2023) 122477, https://doi.org/10.1016/j.saa.2023.122477.
[30] J. Wang, Y. Lin, Q. Li, Z. Lu, J. Qian, H. Dai, Y. He, Non-destructive detection and grading of chilling injury-induced lignification of kiwifruit using X-ray computer tomography and machine learning, Comput. Electron. Agric. 218 (2024) 108658, https://doi.org/10.1016/j.compag.2024.108658.
[31] Y. Zhang, Y. Lin, H. Tian, S. Tian, H. Xu, Non-destructive evaluation of the edible rate for pomelo using X-ray imaging method, Food Control 144 (2023) 109358, https://doi.org/10.1016/j.foodcont.2022.109358.
[32] J. Kremer, K.S. Pedersen, C. Igel, Active learning with support vector machines, Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery 4 (4) (2014) 313–326, https://doi.org/10.1002/widm.1132.
[33] M. Paliwal, U.A. Kumar, Neural networks and statistical techniques: a review of applications, Expert Syst. Appl. 36 (1) (2009) 2–17, https://doi.org/10.1016/j.eswa.2007.10.005.
[34] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[35] B. Talekar, S. Agrawal, A detailed review on decision tree and random forest, Bioscience Biotechnology Research Communications 13 (14SI) (2020) 245–248, https://doi.org/10.21786/bbrc/13.14/57.
[36] X. Deng, S. Cao, A.L. Horn, Emerging applications of machine learning in food safety, Annu. Rev. Food Sci. Technol. 12 (1) (2021) 513–538, https://doi.org/10.1146/annurev-food-071720-024112.
[37] X. Wang, Y. Bouzembrak, A.O. Lansink, H.J. Van Der Fels-Klerx, Application of machine learning to the monitoring and prediction of food safety: a review, Compr. Rev. Food Sci. Food Saf. 21 (1) (2022) 416–434, https://doi.org/10.1111/1541-4337.12868.
[38] L. Oliveira Chaves, A.L. Gomes Domingos, D. Louzada Fernandes, F. Ribeiro Cerqueira, R. Siqueira-Batista, J. Bressan, Applicability of machine learning techniques in food intake assessment: a systematic review, Crit. Rev. Food Sci. Nutr. 63 (7) (2023) 902–919, https://doi.org/10.1080/10408398.2021.1956425.
[39] X. Zeng, R. Cao, Y. Xi, X. Li, M. Yu, J. Zhao, J. Li, Food flavor analysis 4.0: a cross-domain application of machine learning, Trends Food Sci. Technol. 138 (2023) 116–125, https://doi.org/10.1016/j.tifs.2023.06.011.
[40] P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, in: Robust Data Mining, Springer, New York, 2013, https://doi.org/10.1007/978-1-4419-9878-1.
[41] H. Dai, C. MacBeth, Effects of learning parameters on learning procedure and performance of a BPNN, Neural Network. 10 (8) (1997) 1505–1521, https://doi.org/10.1016/S0893-6080(97)00014-2.
[42] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[43] S. Bhatia, G. Wellington, J. Cocchiara, J. Lalko, C. Letizia, A. Api, Fragrance material review on cinnamyl butyrate, Food Chem. Toxicol. 45 (1) (2007) S62–S65, https://doi.org/10.1016/j.fct.2007.09.028.
[44] J.H. Loughrin, T.R. Hamilton-Kemp, R.A. Andersen, D.F. Hildebrand, Headspace compounds from flowers of Nicotiana tabacum and related species, J. Agric. Food Chem. 38 (2) (1990) 455–460, https://doi.org/10.1021/jf00092a027.
[45] L. Reger, J. Moß, H. Hahn, J. Hahn, Analysis of menthol, menthol-like, and other tobacco flavoring compounds in cigarettes and in electrically heated tobacco products, Contributions to Tobacco & Nicotine Research 28 (2) (2018) 93–102, https://doi.org/10.2478/cttr-2018-0010.
[46] B. Wang, S. Yang, G. Chen, Y. Wu, Y. Hou, G. Xu, Simultaneous determination of non-volatile, semi-volatile, and volatile organic acids in tobacco by SIM–Scan mode GC-MS, J. Separ. Sci. 31 (4) (2008) 721–726, https://doi.org/10.1002/jssc.200700318.
[47] C. Li, S. Tian, J. You, J. Liu, C. Wang, Q. Wang, R. Tian, Qualitative determination of volatile substances in different flavored cigarette paper by using headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS) combined with chemometrics, Heliyon 9 (1) (2023) e12146, https://doi.org/10.1016/j.heliyon.2022.e12146.
[48] X. Fan, W. Zi, J. Ao, B. Li, J. Qiao, Y. Wang, Y. Nong, Analysis and application evaluation of the flavour-precursor and volatile-aroma-component differences between waste tobacco stems, Heliyon 8 (9) (2022) e10658, https://doi.org/10.1016/j.heliyon.2022.e10658.
[49] J.C. Morgan, M.J. Byron, S.A. Baig, I. Stepanov, N.T. Brewer, How people think about the chemicals in cigarette smoke: a systematic review, J. Behav. Med. 40 (2017) 553–564, https://doi.org/10.1007/s10865-017-9823-5.
[50] J. Lynch, L. Jin, A. Richardson, D.J. Conklin, Tobacco smoke and endothelial dysfunction: role of aldehydes? Curr. Hypertens. Rep. 22 (2020) 1–9, https://doi.org/10.1007/s11906-020-01085-7.
[51] M. Banožić, S. Jokić, Đ. Ačkar, M. Blažić, D. Šubarić, Carbohydrates—key players in tobacco aroma formation and quality determination, Molecules 25 (7) (2020) 1734, https://doi.org/10.3390/molecules25071734.