DOI: 10.23977/autml.2026.070119
ISSN 2516-5003 Vol. 7 Num. 1

Comparative Study on the Performance of Classification Models in Exhaled VOCs Sensing Applications

Ling Jiang[1], Chilin Zhang[2], Xianhua Zhong[1,2,3,*]

[1]*School of Life Health Information Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
[2]*Chongqing Postdoctoral Research Station of Medical Electronics and Information Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
[3]*Chongqing Engineering Laboratory of Digital Medical Equipment and Systems, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
[*]*Corresponding Author*

Keywords: Exhaled VOCs; Colorimetric Sensor Array; Classification Model; Performance Comparison

Abstract: The volatile organic compounds (VOCs) in exhaled breath are important disease biomarkers, and their rapid and accurate identification is of great significance for early disease screening. Regarding the issue of the applicability of classification models in the VOCs identification of colorimetric sensor arrays under small sample conditions, this study focused on the array response data of 20 typical exhaled VOCs, constructed and compared models such as linear discriminant analysis (LDA), hierarchical cluster analysis (HCA), support vector machine (SVM), and residual network (ResNet), and systematically evaluated their classification performance. The results showed that HCA and LDA are suitable for stability assessment and preliminary classification of the sensing array; SVM is suitable for small-scale classification tasks; ResNet34 achieved 100% recognition accuracy on the validation set and achieved a good balance between recognition performance and computational complexity, making it more suitable for high-precision and on-site detection scenarios. This study provides a basis for model selection for exhaled VOCs detection based on multiple sensors or multiple response modes, and has reference significance for promoting the application of colorimetric sensor arrays in clinical early diagnosis.

1. Introduction

Human breath contains hundreds of volatile organic compounds (VOCs) at varying concentrations, and their profiles are closely associated with an individual’s metabolic state.[1]In the presence of diseases such as lung cancer,[2]diabetes,[3]and gastrointestinal disorders,[4]relevant metabolic pathways may become dysregulated, leading to significant changes in the composition or concentrations of specific VOCs in exhaled breath. Therefore, breath VOCs are considered highly promising noninvasive biomarkers. Owing to their noninvasive nature, convenience, and potential for real-time monitoring, VOC-based diagnostic approaches have attracted increasing attention in recent years.[5]

The colorimetric sensor array technology, as a new type of sensing and analysis technology, works on the principle that different sensitive materials in the array undergo specific physical adsorption or chemical reactions with VOCs molecules, resulting in visual distinguishable color changes. By capturing these color change signals, it can achieve qualitative and quantitative analysis of VOCs [6]. This technology not only has significant advantages such as rapid response, high visualization degree, low manufacturing cost, and simple operation, but also can simultaneously detect multiple VOCs, perfectly meeting the application requirements of on-site rapid analysis of exhaled VOCs. It has shown good application prospects in scenarios such as lung cancer exhaled biomarker detection[7]and diabetes [8].

Nevertheless, colorimetric response patterns from different VOCs can exhibit cross-interference, and the resulting sensor-array datasets are typically high-dimensional yet small in sample size, posing challenges for feature extraction and reliable classification. How to select appropriate classification models under these constraints to enable effective VOC identification remains insufficiently explored. Therefore, under small-sample conditions, a systematic comparison of different classification algorithms and clarification of their applicability is important for guiding model selection in breath-analysis systems based on multi-sensor or multi-response patterns.

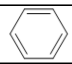
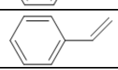
In this study, several representative classification models-HCA, LDA, SVM, and ResNet-were implemented to systematically evaluate their performance in colorimetric breath VOC recognition. Metrics including accuracy, sensitivity, and specificity were compared across recognition tasks involving 20 representative disease-related VOCs to examine model suitability for small-sample, high-dimensional data. In addition, classification confusions were analyzed in conjunction with VOC molecular structural characteristics, providing a reference for rational model selection in colorimetric sensor-array-based breath VOC analysis.

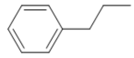
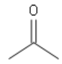
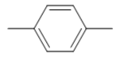
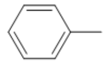
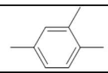
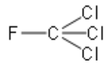
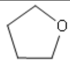
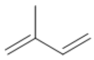
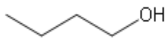
2. Materials and Methods

2.1 Experimental Data

In this study, a colorimetric sensor array was used to detect and analyze 20 VOCs, and the relevant information of the analytes is listed in Table 1. Before exposure to the analytes, images of the initial state of the sensor array were captured using a high-definition camera. After the target VOCs interacted with the sensor array for 4 min and reached response equilibrium, images of the array in its post-response state were acquired again. Subsequently, digital image subtraction was performed between the images before and after the reaction to obtain fingerprint maps that specifically represent the characteristic responses of the VOCs. Essentially, these fingerprint maps are visual representations of the response-difference vectors of the sensor array, which can effectively reflect the differences in sensor responses to different VOCs. The fingerprint maps of the 20 VOCs are shown in Figure 1.

Table 1: List of 20 volatile organic compounds (VOCs)

Analyte	Abbr.	Structure
1-Hexene	HexE	
Benzene	Ben	
Styrene	Sty	

Propylbenzene	ProB	
Acetone	Ace	
p-Xylene	PX	
Heptanal	HepA	
n-Decane	n-Dec	
Cyclohexane	CYH	
Hexanal	HexA	
Toluene	Tol	
Methanol	MeOH	
Trimethylbenzene	TMB	
Trichlorofluoromethane	TcFM	
Undecane	UND	
Tetrahydrofuran	THF	
Isoprene	ISO	
n-Butanol	n-BuOH	
n-Heptane	n-Hep	
n-Pentane	n-Pen	

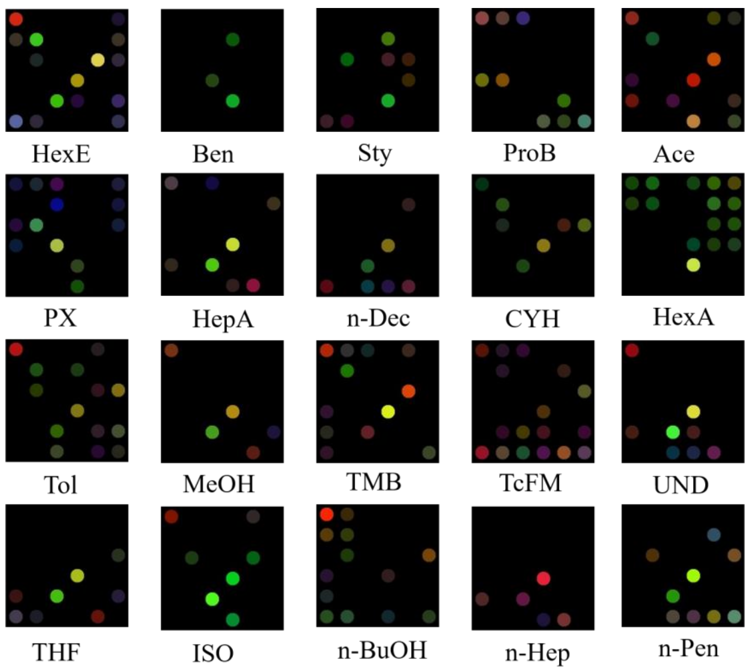


Figure 1: Fingerprint spectra of 20 kinds of VOCs

2.2 Methods

2.2.1 Hierarchical Cluster Analysis

HCA was carried out in the MATLAB environment. The colorimetric sensor array consists of 36 sensitive points. By extracting the differences in the RGB channels of the images before and after the reaction, a 108-dimensional feature vector was constructed to represent the VOCs response. For each type of VOC, 5 sets of data after sufficient reaction were selected for analysis, and their 108-dimensional difference vectors were used as sample features. HCA calculated the differences between samples using the Euclidean distance and performed clustering based on the Ward minimum variance method. Finally, the tree-like clustering diagram was used to display the similarity relationships between different VOCs, in order to evaluate the recognition and discrimination ability of the sensor array.

2.2.2 Linear Discriminant Analysis

The LDA model construction and visualization were completed in the Python 3.9 environment. Firstly, the difference vectors of 20 types of VOCs were organized into a sample matrix, with the gas category serving as the label. The model extracted discriminative features based on the Fisher discriminant criterion, improved the classification ability by maximizing the ratio of inter-class divergence to intra-class divergence, and constructed the discriminant boundary using Mahalanobis distance to enhance the classification stability. The sample distribution and category separation were presented through a two-dimensional scatter plot with 95% confidence ellipses.

To evaluate the ability of the model to identify the 20 VOCs, a cross-validation strategy was adopted, and the dataset was divided into a training set and a validation set at a ratio of 3:2. According to the prediction results on the validation set, the samples were categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). On this basis, the model performance was quantitatively assessed using three metrics: sensitivity (TP/(TP+FN)), specificity (TN/(TN+FP)), and accuracy ((TP+TN)/total number of samples).

2.2.3 Support Vector Machine

Based on the scikit-learn library, an SVM classification model was constructed to identify 20 types of VOCs spectra. The 600 fingerprint spectra were randomly divided into a training set and a test set in a 9:1 ratio. Firstly, a baseline SVM model was established using a linear kernel function, with the penalty parameter C set to 1 to evaluate the initial classification ability of the features. On this basis, grid search combined with five-fold cross-validation was used to optimize the kernel function type, penalty coefficient C, and kernel function parameter γ, and the optimal parameters were determined based on the cross-validation accuracy to establish the final SVM classification model.

2.2.4 ResNet Network

This study is based on Python 3.9 and the Pytorch deep learning framework, and builds ResNet18, ResNet34 and ResNet50 models respectively to classify and identify 20 types of VOCs difference spectra. 600 VOCs fingerprint spectra are randomly divided into training and validation sets at a ratio of 9:1. The relevant parameter configuration is shown in the Table 2.

Table 2: Settings for ResNet network parameters

Parameter	Configuration
CPU	Intel i5 12400F
GPU	NVIDIA RTX 1660 Ti
Running Memory	16G
Epochs	25
Batch size	4
Learning rate	0.0001
Optimizer	Adam
Loss function	Cross-Entropy Loss
Activation function	ReLU

3. Results and Analysis

3.1 Hierarchical Cluster Analysis

The HCA clustering results are shown in Figure 2. The dendrogram indicates that the vast majority of VOCs samples can be grouped into one category with a relatively small Euclidean distance, suggesting that this detection system has good repeatability and stability for the same gas. Only one sample of Tetrahydrofuran (THF) deviated from the main cluster and was mistakenly grouped into the Heptanal (HepA) sub-cluster, which might be related to experimental noise or the occasional overlap of their response characteristics. Apart from this outlier, the remaining 19 types of VOCs and most of the THF samples were correctly distinguished, indicating that HCA has a good ability to identify complex VOCs systems.

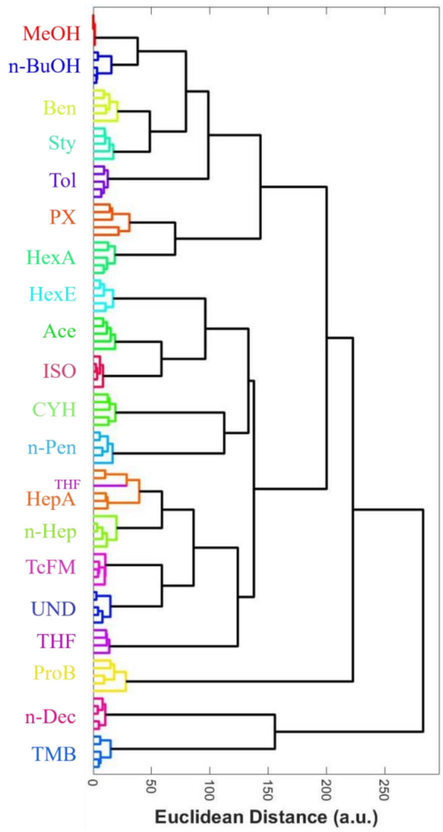


Figure 2: Dendrogram of 20 VOCs

3.2 Linear Discriminant Analysis

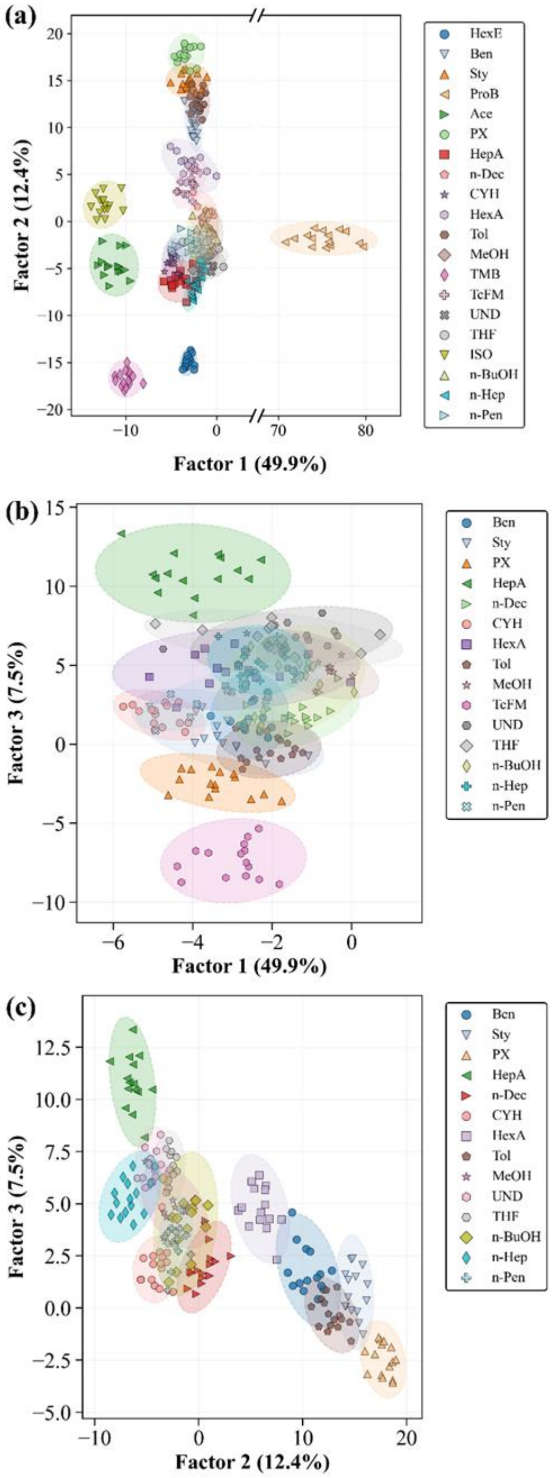


Figure 3: Discriminant analysis results for 20 VOCs based on (a) the first two discriminant factors, (b) the first and third discriminant factors, and (c) the second and third discriminant factors; the elliptical regions represent 95% confidence intervals

The LDA score plot based on the first three discriminant factors is shown in Figure 3. In Figure 3(a), only five VOCs, namely Propylbenzene (ProB), Isoprene (ISO), Trimethylbenzene (TMB), Acetone (Ace), and 1-Hexene (HexE), can be clearly distinguished, while the remaining 15 show significant overlap. Further analysis indicates that in Figure 3(b), only Trichlorofluoromethane (TcFM) is separated alone; in Figure 3(c), the undifferentiated VOCs mainly cluster into two groups: one group consists of aromatic hydrocarbons such as Benzene (Ben), Toluene (Tol), Styrene (Sty), and p-Xylene (PX), and the other group includes alkanes and their derivatives such as n-Pentane (n-Pen), n-Heptane (n-Hep), n-Decane (n-Dec), Undecane (UND), and Cyclohexane (CYH). Aromatic hydrocarbons are concentrated due to their similar benzene ring structures and π electron systems; alkanes and their derivatives, lacking characteristic functional groups, show strong overlap as their sensing responses vary continuously with carbon chain length. Some oxygen-containing compounds, such as n-Butanol (n-BuOH), Methanol (MeOH), HepA, and THF, also overlap with alkanes, indicating that their overall responses are still dominated by the non-polar hydrocarbon backbone. The partial overlap between Hexanal (HexA) and Ben may be related to the fact that HexA has both a long non-polar chain and a weakly polar carbonyl group.

Table 3: Sensitivity, specificity, and accuracy of the array for 20 VOCs based on the cross-validation model

Analyte	Abbr.	Specificity/%	Sensitivity/%	Accuracy/%
1-Hexene	HexE	100	100	100
Benzene	Ben	83.3	100	99.2
Styrene	Sty	100	98.2	98.3
Propylbenzene	ProB	100	100	100
Acetone	Ace	100	100	100
p-Xylene	PX	83.3	100	99.2
Heptanal	HepA	100	100	100
n-Decane	n-Dec	66.7	100	98.3
Cyclohexane	CYH	83.3	100	99.2
Hexanal	HexA	100	99.1	99.2
Toluene	Tol	100	99.1	99.2
Methanol	MeOH	66.7	100	98.3
Trimethylbenzene	TMB	100	100	100
Trichlorofluoromethane	TcFM	100	99.1	99.2
Undecane	UND	100	100	100
Tetrahydrofuran	THF	100	100	100
Isoprene	ISO	100	100	100
n-Butanol	n-BuOH	100	98.2	98.3
n-Heptane	n-Hep	100	99.1	99.2
n-Pentane	n-Pen	83.3	100	99.2

Table 3 shows that the LDA model has a high overall recognition performance for the 20 VOCs, with accuracy rates for each category above 98.3%. Among them, HexE, ProB, Ace, HepA, TMB, UND, THF, and ISO have sensitivity, specificity, and accuracy all reaching 100%, demonstrating the best recognition effect. Sty, HexA, Tol, TcFM, n-BuOH, and n-Hep show high specificity but slightly lower sensitivity, indicating that they have few false positives but some false negatives; Ben, PX, CYH, n-Pen, n-Dec, and MeOH exhibit high sensitivity but relatively lower specificity, suggesting that although they can be completely recalled, there is still some cross-interference with similar substances, with n-Dec and MeOH having relatively higher misjudgment risks.

The LDA discrimination plots show some category overlap in the low-dimensional space, but the cross-validation results still indicate that the model has a high classification accuracy. This is mainly because the discrimination plots only reflect the projection distribution of samples in two or threedimensional space, and the dimension reduction process inevitably loses some discriminative information; in contrast, cross-validation is performed in the original high-dimensional feature space, which can retain more complete feature differences, thus the model can still establish effective classification boundaries and achieve good recognition results.

3.3 Support Vector Machine

In this study, the optimal SVM hyperparameters were determined using a grid-search strategy combined with five-fold cross-validation. The resulting configuration consisted of a penalty parameter C=10, a radial basis function (RBF) kernel, and a kernel parameter γ=0.01. The prediction results obtained under this optimal parameter setting, together with the corresponding confusion matrix, are presented in Table 4 and Figure 4.

Overall, the model demonstrated high accuracy and stability in the multi-classification task of VOCs. Except for a few categories, the specificity, sensitivity, and accuracy of most VOCs reached 100%. The diagonal elements dominated in the confusion matrix, while the off-diagonal distribution was sparse, indicating that the model had good overall classification performance and a low misjudgment rate. Specifically, the specificity, sensitivity, and accuracy of 15 types of VOCs, including HexE, Ben, ProB, Ace, PX, HepA, n-Dec, HexA, Tol, MeOH, TcFM, THF, ISO, n-Hep, and n-Pen, were all 100%, indicating that these categories had good separability and the model could accurately identify them. In contrast, a few categories showed some confusion. Sty and TMB had high sensitivity but relatively low specificity, suggesting that the model was sensitive to their identification but had a slightly weaker ability to exclude non-target samples. CYH and n-UND also had some misclassification, mainly due to the similarity of their features with those of structurally similar VOCs. Overall, the model had good recognition ability for most VOCs, and the few misclassifications mainly originated from the overlap of some categories in the feature space.

Table 4: Sensitivity, specificity, and accuracy of SVM for 20 VOCs

Analyte	Abbr.	Specificity/%	Sensitivity/%	Accuracy/%
1-Hexene	HexE	100	100	100
Benzene	Ben	100	100	100
Styrene	Sty	80	100	99
Propylbenzene	ProB	100	100	100
Acetone	Ace	100	100	100
p-Xylene	PX	100	100	100
Heptanal	HepA	100	100	100
n-Decane	n-Dec	100	100	100
Cyclohexane	CYH	50	97.3	95
Hexanal	HexA	100	100	100
Toluene	Tol	100	100	100
Methanol	MeOH	100	100	100
Trimethylbenzene	TMB	50	100	97.5
Trichlorofluoromethane	TcFM	100	100	100
Undecane	UND	80	94.7	94
Tetrahydrofuran	THF	100	100	100
Isoprene	ISO	100	100	100
n-Butanol	n-BuOH	50	98.9	96.5
n-Heptane	n-Hep	100	100	100
n-Pentane	n-Pen	100	100	100

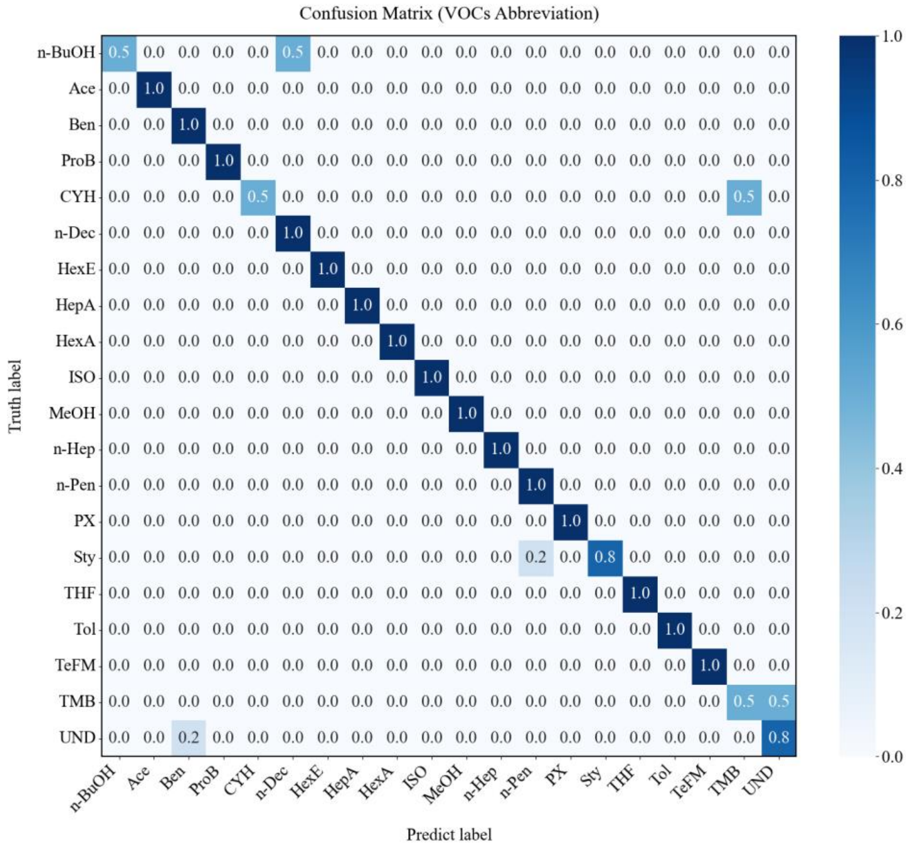


Figure 4: Classification confusion matrix of SVM for 20 VOCs

3.4 ResNet Network

The training results of each model are shown in Figure 5. All three ResNet models demonstrated excellent convergence during the training process. The training accuracy increased rapidly within a few epochs and stabilized, while the validation accuracy remained at a relatively high level overall, indicating that the residual network can effectively extract the discriminative features of the VOCs fingerprint patterns. Specifically, the training and validation processes of ResNet18 were relatively stable, and the final validation set accuracy was 95.7%; ResNet34 and ResNet50 exhibited stronger feature learning capabilities, with both validation accuracies reaching 100%. Considering the comprehensive recognition accuracy and model complexity, ResNet34 demonstrated superior overall performance.

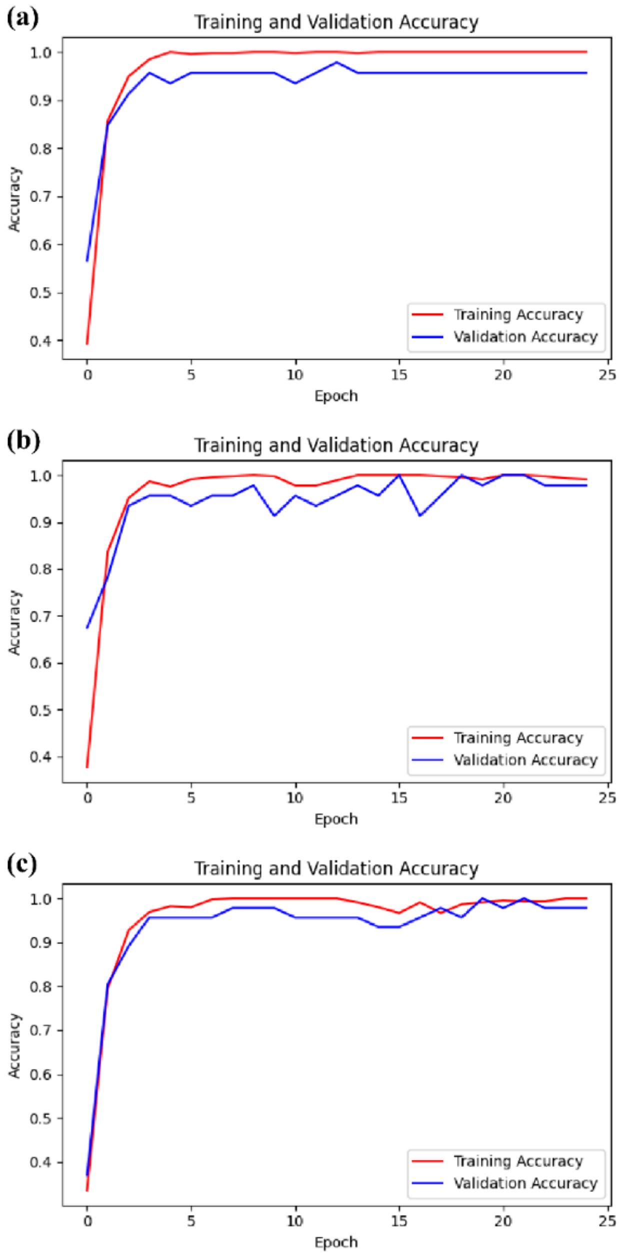


Figure 5: Training accuracy of ResNet models with different depths: (a) ResNet18 model, (b) ResNet34 model, and (c) ResNet50 model

3.5 Discussion

Table 5 compares the parameter quantities and overall accuracy rates of different models in the VOCs identification task. The results show that HCA, based on Euclidean distance and Ward's minimum variance method, achieved a 99% clustering accuracy rate, indicating that the sensor array has good repeatability and stability. However, its ability to distinguish outliers and structurally similar VOCs is limited. It is more suitable for system stability verification, sample distribution analysis, and preliminary classification assistance. LDA's overall accuracy rate is 93.3%, which has certain advantages in VOCs classification with significant feature differences. However, its discriminative ability for aromatic hydrocarbons and homologues is insufficient, so it is more suitable for feature visualization, category separability analysis, and preliminary screening scenarios. SVM achieved an overall accuracy rate of 90% under the optimal parameters, with a parameter quantity of only 0.012 M, demonstrating good lightweight advantages and small sample adaptability, and is suitable for portable or embedded VOCs detection tasks with limited computing resources. In contrast, the deep learning models show better comprehensive performance. Among them, ResNet18, ResNet34, and ResNet50 achieved 95.7%, 100%, and 100% respectively, indicating that residual networks can more effectively extract discriminative features in complex VOCs response signals. Considering the comprehensive recognition accuracy and model complexity, ResNet34 has better deployment applicability while maintaining high performance, and is more suitable for high-precision VOCs intelligent recognition applications in complex environments.

Table 5: Overall accuracy rate of different models

Model	Parameter (M)	Overall accuracy (%)
HCA	\	99
LDA	\	93.3
SVM	0.012	90
Resnet 18	11.19	95.7
Resnet 34	21.29	100
Resnet 50	23.55	100

4. Conclusions

This study systematically compared HCA, LDA, SVM, and ResNet-based models, thereby verifying the applicability of different methods in VOCs classification and identification tasks. HCA and LDA, characterized by simple structures, high computational efficiency, and strong interpretability, are well suited as fundamental tools for VOCs data analysis, particularly for sample distribution exploration, feature visualization, and sensor performance evaluation. SVM demonstrated relatively good classification performance and generalization ability under small-sample conditions, making it suitable for medium- and small-scale classification tasks under limited computational resources. In contrast, deep learning methods, especially the ResNet models, exhibited stronger feature extraction capability and higher classification accuracy in complex VOCs identification tasks.

Among the deep learning models investigated, ResNet18, ResNet34, and ResNet50 all showed good training stability and strong classification performance. Both ResNet34 and ResNet50 achieved a validation accuracy of 100%, while ResNet34 demonstrated the best overall performance by achieving a better balance between model complexity and recognition accuracy.

It should be noted that the training and validation of ResNet34 were conducted based on only 600 images, resulting in a relatively limited sample size. Although the model achieved a 100% classification accuracy under the current experimental conditions, this result mainly reflects its classification ability on the existing dataset. In the future, it is still necessary to expand the sample size, enrich the data diversity, and adopt more rigorous validation methods to further improve the stability and generalization performance of the model, laying the foundation for its application in actual VOCs sensing detection.

References

[1] Allard-Coutu A, Singh K, David D, et al. Volatile organic compounds: A promising new frontier for cancer screening[J]. Tumor Discovery, 2024, 3(2): 2061.

[2] Pereira J, Porto-Figueira P, Cavaco C, et al. Breath Analysis as a Potential and Non-Invasive Frontier in Disease

Diagnosis: An Overview[J]. Metabolites, 2015, 5(1): 3.
[3] Das S, Pal S, Mitra M. Significance of Exhaled Breath Test in Clinical Diagnosis: A Special Focus on the Detection of Diabetes Mellitus[J]. Journal of Medical & Biological Engineering, 2016, 36(5): 605-624.
[4] Dalis C, Mesfin F M, Manohar K, et al. Volatile Organic Compound Assessment as a Screening Tool for Early Detection of Gastrointestinal Diseases[J]. Microorganisms, 2023, 11(7): 25.
[5] Moura PC, Raposo M, Vassilenko V. Breath volatile organic compounds (VOCs) as biomarkers for the diagnosis of pathological conditions: A review[J]. Biomed J. 2023, 46(4): 100623.
[6] Schoolaert E, Hoogenboom R, De Clerck K. Colorimetric Nanofibers as Optical Sensors[J]. Advanced Functional Materials, 2017: 1702646.
[7] Queralto N, Berliner AN, Goldsmith B, et al. Lim SH. Detecting cancer by breath volatile organic compound analysis: a review of array-based sensors[J]. J Breath Res. 2014, 8(2): 027112.
[8] Faura G, Boix-Lemonche G, Holmeide AK, et al. Colorimetric and Electrochemical Screening for Early Detection of Diabetes Mellitus and Diabetic Retinopathy-Application of Sensor Arrays and Machine Learning[J]. Sensors (Basel). 2022, 22(3): 718.