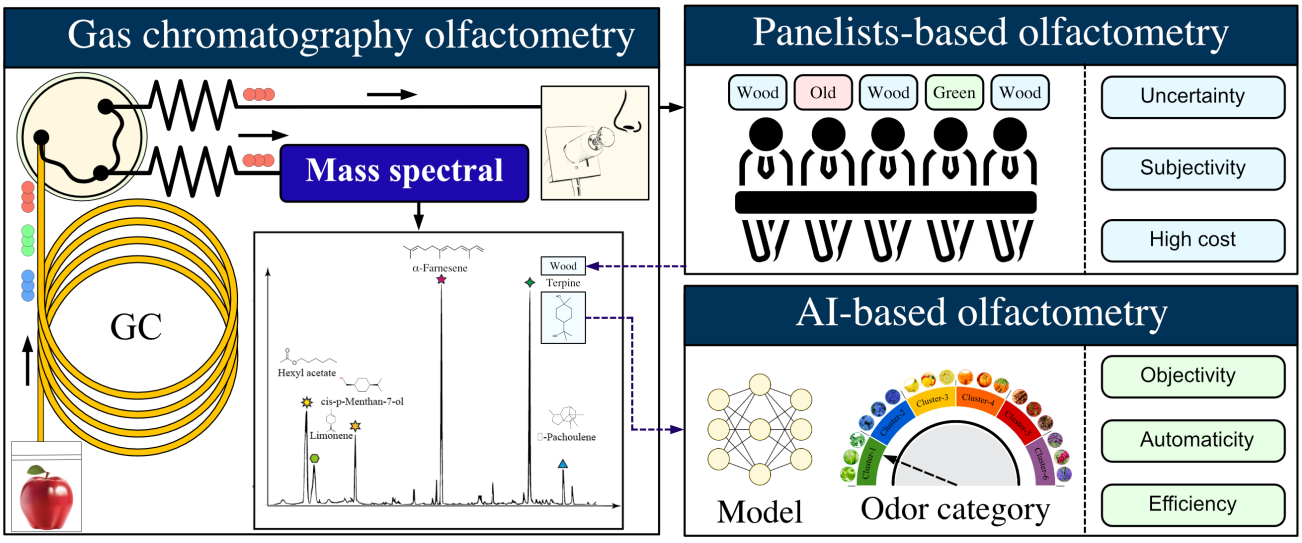
Graphical Abstract

Artificial intelligence-based gas chromatography-olfactometry for sensory evaluation of key compounds in food ingredients

Liang Shang,Chuanjun Liu,Fengzhen Tang,Bin Chen,Lianqing Liu,Kenshi Hayashi



Highlights

Artificial intelligence-based gas chromatography-olfactometry for sensory evaluation of key compounds in food ingredients

Liang Shang,Chuanjun Liu,Fengzhen Tang,Bin Chen,Lianqing Liu,Kenshi Hayashi

- Different deep neural networks were used to predict categorized odor descriptors.
- End-to-end-based representation learning was performed for molecular feature extraction.
- Molecular graphs with pre-training of convolution neural networks was most accurate.

Artificial intelligence-based gas chromatography-olfactometry for sensory evaluation of key compounds in food ingredients

Liang Shang[a,b,*], Chuanjun Liu[c,d], Fengzhen Tang[a,b,*], Bin Chen[e], Lianqing Liu[a,b] and Kenshi Hayashi[c]

[a]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China
[b]Institute for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, 110016, China
[c]Department of Electronics, Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, 819-0395, Japan
[d]Research Laboratory, U.S.E. Co., Ltd., Tokyo, 150-0013, Japan
[e]Chongqing Key Laboratory of Non-linear Circuit and Intelligent Information Processing, College of Electronic and Information Engineering, Southwest University, Chongqing, 400715, China

ARTICLE INFO

Keywords:
Key aroma compounds
Gas chromatography-olfactometry
Structure odor relationship
Odorant molecules structure features extraction
Machine learning

ABSTRACT

The use of an artificial intelligence (AI)-based prediction model of the structure-odor relationship (SOR) has shown great potential in the replacement of human panelists in gas chromatography-olfactometry (GCO). However, the Al-based GCO encounters issues such as poor accuracy, generalization, and practicality, owning to the insufficient feature extraction of odorant molecular structure. The purpose of this study is to the prediction of odor perception categories based on the odorant structure feature extraction by diverse deep neural networks, including molecular graphic convolution neural network (MG-CNN), molecular graph transformer neural network, and atom interaction neural networks. The results of the performance comparison of different feature extractors demonstrate that the MG-CNN model produces the highest accuracy and thus may be most suitable for the SOR prediction. It is hoped that the proposed method can be applied in practice as an auxiliary tool of GCO for the sensory evaluation of key compounds in food ingredients.

1. Introduction

Characterization of the key aroma-active compounds from food is an important topic in food science. To obtain the sensory information of odor, gas chromatography-olfactometry (GCO) has been widely applied as a powerful odor analytic strategy in various research areas, especially in food science (Maurya et al., 2021). Recently, GCO combined with headspace solid-phase microextraction (HS-SPME) has been employed to determine the key aroma compounds of tea (Ma et al., 2022), coffee (Caporaso et al., 2022), pork (Wu et al., 2022), Baijiu (Huang et al., 2022), spices (Ni et al., 2022), and celeries (Turner et al., 2021) et al.

Although GCO can be used to attain accurate sensory and chemical characteristics of odorants, this approach is expensive and time-consuming, which can limit its application. The main expense associated with GCO is the cost of hiring and training panelists to characterize the sensory qualities of odorants using their sense of smell (Acree, 1997). In addition, the sensory assessment of human panelists is personally dependent, which inevitably leads to subjectivity and inconsistency into the evaluation results. In order to overcome the above problems, some auxiliary business software modules, such as GC/MS Off-Flavor Analyzer (Shimadzu Co. Ltd.), AromaOffice (GERSTEL K.K.), and AroChemBase (Alpha MOS Co. Ltd.), have been widely applied as supplements to GCO analysis. The number of odorants with odor descriptors (around 2000 compounds for AroChemBase, 600 compounds for GC/MS Off-Flavor Analyzer) that are collected in their aroma compound databases. However, there are around 44,000 odorants in the world, and most of them are unknown to us. It indicated that existing systems may not meet the harsh requirements of GCO in food aroma analysis. Recently, the use of artificial intelligence (AI) to support analytical purposes has been a potential technology for overcoming the drawbacks of traditional analytical methods. Therefore, AI-based auxiliary for human assessors would reduce the above problems.

Research on the response patterns of neurons in the olfactory bulbs (OB) has illuminated the mechanisms underlying biological olfaction. However, many questions remain, such as why these molecules smell different from one another and why we link smell feelings with certain semantic descriptors, known as odor descriptors (ODs) (Barwich, 2020). These questions may be answered by examining structure odor relationships (SORs), the development of which presents a difficult and interesting challenge. In recent years, many types of research have been conducted to predict odor perception of odorants using various parameters, such as electronic or physicochemical characteristics (Chacko et al., 2020), mass spectrometry (MS) (Nozaki & Nakamoto, 2018), and social network interactions (Snitz et al., 2019). Additionally, novel methods, such as odor-based social networks (Kumar et al., 2015; Snitz et al., 2013; Liu et al., 2019), machine learning (ML) (Loetsch et al., 2019; Choi et al., 2022), deep neural network (DNN) models (Debnath & Nakamoto, 2020), and semantic-based

*Corresponding author
shangliang0225@gmail.com (L. Shang); tangfengzhen@sia.cn (F. Tang)
ORCID(s): 0000-0001-8369-3049 (L. Shang); 0000-0002-7232-2573 (C. Liu); 0000-0002-4654-9440 (F. Tang); 0000-0002-3722-4740 (B. Chen); 0000-0002-2271-5870 (L. Liu); 0000-0001-8679-4953 (K. Hayashi)

approaches (Gutiérrez et al., 2018) have been developed to calibrate models that express the relationships between odorants and ODs. These studies have demonstrated the possibility to use data-driven approaches to solve the SOR problems. In line with this trend, we have proposed a concept of ML-based GC/MS olfactometry in which the sensory evaluation of panelists is expected to be replaced by machine learning prediction models (Shang et al., 2017).

In the ML-based GCO system, the molecular information obtained by MS analysis of the individual GC peaks is transfered into physicochemical parameters by molecular calculation software (DRAGON). By building models appropriately, the ODs of the odorants can be predicted with high accuracy from their physicochemical parameters. The ML-based GCO is a concept-of-proof study and there are many problems to be solved before its practical application. For example, in terms of the sample size, only limited ODs with a high frequency of occurrence are targeted in our models. In reality, it has been reported that over 500 ODs are used for odor evaluation (Dravnieks, 1982). Therefore, many ODs, especially for those rare ODs are not considered by the models. In order to solve this issue, we have recently proposed a categorization approach based on the semantic analysis of ODs, which make the model can cover the prediction of several hundreds of ODs (Shang et al., 2021). Another remaining problem is that the model prediction is based on physicochemical parameters of odorants. The recent development of computational chemistry and ML makes it possible to obtain various molecular structure information (Tsubaki & Mizoguchi, 2019). Unlike traditional methods, DNNs can directly learn latent presentations from molecular structures, such as atom types and their positions, via back-propagation. Moreover, an end-to-end strategy has been proposed as an effective nonlinear modeling method for learning molecular presentation in many fields, such as quantum chemical properties prediction, and compound-protein interaction identification (Sharma et al., 2021). However, since the detailed binding mechanism between the odorants and olfactory receptors is still not fully understood, which molecular features play the most important role in olfactory perception has not been cleared.

The purpose of this study is dedicated to molecular feature mining by diverse DNNs for the prediction of odor perception categories. We focus on using multi-type of feature extractors to understand the relationship between the odorant structure features and odor perception categories and thus construct a prediction model with much higher accuracy. This will make the ML-GCO system more generalizable and therefore more practical. A schematic of the data processing and modeling procedures is illustrated in Fig. 1. We used molecular structures, including 2D molecular images and atom spatial locations, to predict odor sensory categories. For minding useful information from molecular structures, some molecular structural feature extraction methods were applied and their results were compared and discussed. Specifically, we used a molecular graphic convolution neural network (MG-CNN) and molecular graph transformer neural network (MGTNN) to extract features from representations of molecular structure, such as molecular structure images and topology graphs. Moreover, we used atom interaction neural networks (AINNs) to generate features from the spatial structures of atoms. To overcome limitations related to insufficient samples, we also considered pre-trained models, such as pre-trained CNNs and MGTNNs. We compared the feasibility of SOR prediction via molecular fingerprints, molecular parameters, and molecular feature extraction based on MG-CNN, MGTNN, and AINN. We achieved the highest performance by applying molecular features extracted via a pre-trained MG-CNN combined with a multi-label learning model (area under the ROC curve: 0.877±0.028 and F1-score: 0.726±0.028). It indicated that molecularly image features would provide more sensory information than others. Thus, the proposed multi-label odor sensory category identification model is a feasible option for developing AI-based GCO. Although it would be still a challenge to replace human assessors with ML models perfectly, the proposed method could provide references for assessors to increase the efficiency of food aroma analysis.

2. Materials and methods

2.1. Data collection and preparation

To create an odor category prediction model, we collected an extensive dataset containing 2849 odorants. We used information from publicly available databases such as the Odor Map database (Johnson et al., 2009), Flavors and Fragrances database (Sigma-Aldrich) (Sigma-Aldrich, 2016), and the Good Scents database (Arn et al., 1998). As illustrated in Fig. 1, we collected and prepared 5 types of molecular descriptors, including molecular parameters, molecular fingerprints, molecular graphs, simplified molecular-input line-entry system (SMILES), and atom coordinates. Specifically, we used SMILES and RDKit Software (ver. 2021.03.1) to collect normalized molecular parameters (1826), molecular fingerprints (binary, 2048 bits), and 2D molecular graphs (RGB, 300×300 pixels). We used RDKit to obtain the 2D and 3D atom coordinates of the odorants, which we used to train the model regarding the atomic interactions between odorant structures and their odor sensory categories. In total, 256 ODs were clustered into 20 categories using a co-occurrence Bayesian embedding method. More detailed information regarding the cleaning and categorization of ODs can be found in Table S1 (Shang et al., 2021). Data were processed and analyzed using Python (ver. 3.9.0) and R (ver. 4.1.1). Because of the diverse molecular representations of odorants, we need to employ various feature extraction technologies to obtain embeddings for model calibration.

2.2. Model calibration

The calibration and validation process for the odor category model is shown in Fig. 1. First, all samples were divided via random splitting into training and test sets with a 4:1 ratio, and the reported results are averaged over 50 repetitions. Afterward, we employed molecular parameters,

molecular fingerprints, molecular graphic features, molecular graph transformers, and atomic interaction embedding to extract the molecular features of the odorants (Fig. 2). We used a DNN approach to develop a multi-label odor category learning model (the output size was 20) based on the aforementioned molecular information. The cost function $\mathcal{L}(\mathbf{\Theta})$ of the DNN multi-label classifier was calculated by summing the binary cross-entropy of each class, which was defined as follows:

$$\mathcal{L}(\mathbf{\Theta}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j} \times \log(\hat{y}_{i,j}) + (1 - y_{i,j}) \times \log(1 - \hat{y}_{i,j}) \tag{1}$$

where $\mathbf{\Theta}$ indicates the parameter set of the model, N indicates the sample number in the training set, and M indicates the number of odor categories, set as 20 in the present study. $y_{i,j}$ and $\hat{y}_{i,j}$ are the ground truth and prediction label for category j of sample i, respectively. By minimizing the cost function based on the stochastic steepest gradient descent algorithm, parameters from the DNN could be learned and updated. For each DNN method, the number of hidden layers and nodes was selected from 1 to 8 layers and {16, 32, 64, 128, 256, 512, 1024, 2048} nodes, respectively. Network parameters, including the dropout ratio, learning rate, and training epoch, were set as 0.1, 1×10^{-4}, and 200, respectively. The optimal models were determined according to their areas under the ROC curve (AUC) and F1-scores based on precision and recall, simultaneously. Both qualitative and quantitative data analyses were performed.

Because molecular graphs, molecular SMILES sequences, and atom coordinates are not tabular data, they cannot be used as classifier inputs directly. Therefore, diverse feature extraction methods were first employed to convert those unstructured data to tabular features. Specifically, pre-trained CNNs, sequence transformer, and atomic interaction embedding were utilized for molecular graphs, SMILES sequences, and atom coordinates feature extraction, respectively. Detailed information of the above-mentioned strategies was summarized as follows.

2.3. Molecular graphic feature extraction

CNNs are highly successful graphic feature extractors, and are commonly developed with high accuracy by large training datasets. Given the utility of convolution kernels and DNNs, CNNs have played a critical role in image and video processing. Recently, molecular graph embedding has been used to model the relationships between chemical compounds (Sharma et al., 2021). To investigate the feasibility of molecular graphic presentation for odor category prediction, we used 4 types of effective CNNs, including the VGG-16, Resnet, Densnet, and Alexnet, as feature extractors for generating embedding from molecular images in the present study (Fig. 2a). Detailed structures for these CNNs have been previously presented (Krizhevsky et al., 2017). In the present study, we used pre-trained CNN models for odor category prediction to overcome the limitation of sample size.

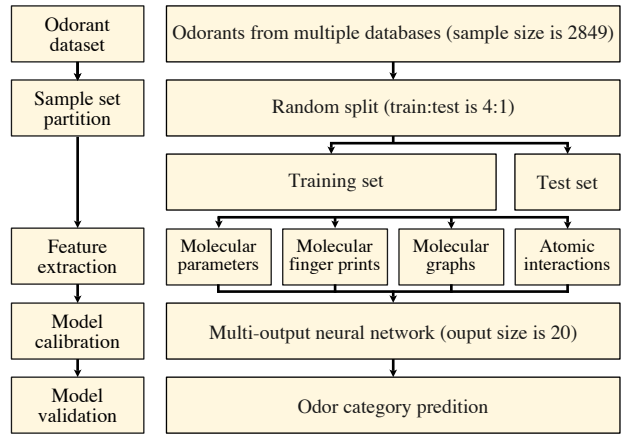


Figure 1: Data processing for calibration and validation of an odor category prediction model.

2.4. Molecular sequence feature extraction

As an end-to-end supervised learning algorithm, graph neural networks (GNNs) have been widely applied for sequence embedding in various fields. Given that odorants can be described as molecular topology graphs using SMILES, we considered GNNs to be appropriate for molecular presentation. Broadly speaking, graph transformers are considered to be a powerful tool for handling molecular presentation through encoding via SMILES, which has been used to predict compound protein interactions, virtual screening, and molecular parameters (Warikoo et al., 2021).

Although the molecular graph transformer neural network (MGTNN) has strong potential for molecular modeling, deep learning models always require a large amount of labeled data for training. To overcome the above problems, we used a self-supervised graph transformer (GROVER) to obtain presentation information from the odorants for odor category prediction. A brief description of the GROVER is given in Fig. 2b. The pre-training architecture was mainly composed of two parts: i) a transformer-based neural network, and ii) a GNN for molecular structure extraction. The input of the model was an odorant graph presentation $\mathcal{G} = (V, E)$, where V was the set of atoms and E was the set of bonds. Specifically, $v_i \in V$ and $e_{i,j} \in E$ were the i-th atom and bond between the i-th and j-th atom, respectively. The GNN was designed to embed extraction according to queries ($\mathbf{Q}$), keys ($\mathbf{K}$), and values ($\mathbf{V}$) from the atoms in molecular graphs ($\mathcal{G}$). The message transmission process of the GNN, as well as the neighborhood aggression between an atom (v) and its neighbors ($\mathcal{N}_v$) in an odorant ($\mathcal{G}$), were adopted to iteratively (L) update hidden states ($\mathbf{h}_v$) for atom v, which can be written as:

$$\begin{aligned} \mathbf{m}_v^{(l,k)} &= \text{Aggregate}^{(l)}(\{(\mathbf{h}_v^{(l,k-1)}, \mathbf{h}_u^{(l,k-1)}, \mathbf{e}_{v,u}) | u \in \mathcal{N}_v\}) \\ \mathbf{h}_v^{(l,k)} &= \sigma(\mathbf{W}^l \mathbf{m}_v^{(l,k)} + \mathbf{b}^{(l)}) \\ \mathbf{h}_{\mathcal{G}} &= \text{Readout}(\{\mathbf{h}_v^{(0,K_0)}, \ldots, \mathbf{h}_v^{(L,K_L)} | v \in V\}) \end{aligned} \tag{2}$$

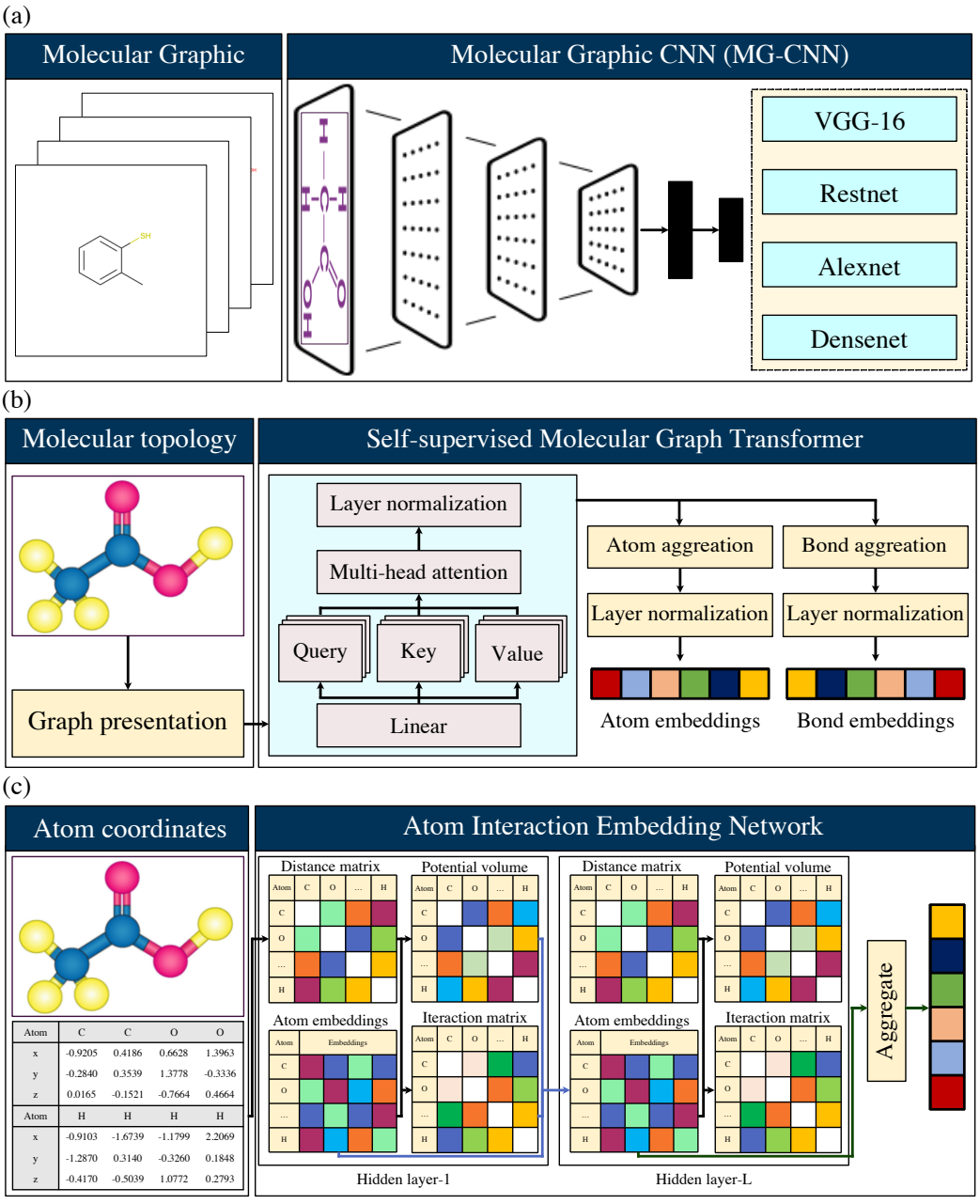


Figure 2: Overview of the design of molecular feature extractors, including (**a**) a molecular graphic convolution neural network (MG-CNN), (**b**) molecular graph transformer neural network (MGT-NN), and (**c**) atom interaction neural network (AINN).

where $\mathbf{m}_v^{(l,k)}$ indicates the passing message for atom v under the k-th step of the l-th iteration. Here, we suppose each iteration (l) contains K_l steps. $\text{Aggregate}^{(l)}(\cdot)$ is an aggregation function, which can be selected from the mean, max pooling, or graph attention mechanism. $\sigma(\cdot)$ is the activation function, and $\mathbf{h}_G$ is the graph-level representation generated by a Readout operation. The resulting matrices $(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ were fed to the transformer module, which was composed of graph multi-head attention blocks:

$$\begin{aligned}
\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_k)\mathbf{W}^O \\
\text{head}_i &= \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \\
\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{softmax}(\mathbf{Q}\mathbf{K}^T/\sqrt{d})\mathbf{V}
\end{aligned} \tag{3}$$

where $\mathbf{W}_i^{\mathbf{Q}}$, $\mathbf{W}_i^{\mathbf{K}}$, $\mathbf{W}_i^{\mathbf{V}}$ are the projection matrices of head_i. d indicates the dimension of $\mathbf{q}$ and $\mathbf{k}$.

The self-supervised learning tasks in the present study were assigned as contextual property prediction, and graph-level motifs, as well as molecular components, were used

to predict links between both nodes and edges. In summary, we employed a pre-trained model, calibrated with 10 million molecules, as a molecular topology feature extractor in the present study. Instances of atom embedding (2048 dimensions) and bond embedding (2048 dimensions) generated by the above-mentioned procedure were used for odor category prediction. This simple strategy has been demonstrated to be a powerful method in terms of graph expression and structure information extraction. Details regarding graph transformers can be found elsewhere (Mao et al., 2021).

2.5. Atomic interaction embedding

Numerous studies have confirmed that atom interactions are crucial to odor perception (Fjaeldstad, 2018). Consequently, molecular features generated by atomic interactions may be feasible for odor category prediction. In the present study, we modeled interactions between atoms in an odorant molecule using a DNN-based model as a molecular feature extractor. A brief description of the process for the AINN is given in Fig. 2c. Formally, given an odorant $\mathcal{O} = \{(a_i, \mathbf{c}_i)\}_{i=1}^{\mathcal{N}_{atom}}$, where a_i is the i-th atom type, $\mathbf{c}_i \in \mathbb{R}^2 or \mathbb{R}^3$ is the coordinate vector of the i-th atom, and N_{atom} is the total number of atoms for the odorant. To obtain an atom embedding description for the odorant, $\mathcal{V}_{\mathcal{O}} = \{(\mathbf{v}_i, \mathbf{c}_i)\}_{i=1}^{\mathcal{N}_{atom}}$, where $\mathbf{v}_i \in \mathbb{R}^d$ is the embedded vector for the i-th atom. The embedding dimensionality d is a hyper-parameter that must be assigned before training, and these atom embeddings were initialized randomly and optimized via backpropagation. To select the update strategy for the above-mentioned embeddings, we referred to the previous use of DNNs with common graph-structured datasets (Chen et al., 2020). Accordingly, the atom embeddings were updated as follows:

$$\begin{aligned} \mathbf{v}_i^{(l+1)} &= f(\mathbf{v}_i^l) + \sum_{j \in \mathcal{O} \setminus i} g(\mathbf{v}_j^{(l)}, P_{i,j}^{(l)}, \alpha_{i,j}^{(l)}) \\ P_{i,j}^{(l)} &= f(\mathbf{v}_i^{(l)}, \mathbf{v}_j^{(l)}, dist_{i,j}) \\ \alpha_{i,j}^{(l)} &= f(\mathbf{v}_i^{(l)}, \mathbf{v}_j^{(l)}, dist_{i,j}) \\ dist_{i,j} &= ||\mathbf{c}_i - \mathbf{c}_j|| \end{aligned} \tag{4}$$

where $f(\cdot)$ and $g(\cdot)$ were the neural networks. $P_{i,j}^{(l)} \in \mathbb{R}$ and $\alpha_{i,j}^{(l)} \in \mathbb{R}$ indicate the potential volume and interaction factor between the i-th and j-th atoms at the l-th hidden layer, respectively. $dist_{i,j} \in \mathbb{R}$ was the Euclidean distance between the i-th and j-th atoms. Thus, the atom interaction embeddings for the odorant ($\mathbf{x}_{\mathcal{O}} \in \mathbb{R}^d$) could be calculated by:

$$\mathbf{x}_{\mathcal{O}} = \text{Aggregate}(\{\mathbf{v}_i\}_{i=1}^{\mathcal{N}_{atom}}) \tag{5}$$

where Aggregate$(\cdot)$ was the aggregate function, which was mean pooling in the present study. As an option, we added a residual part to prevent the vanishing gradient problem in the DNN (res-AINN), which could be defined as follows:

$$\mathbf{v}_i^{(l+1)} = Norm(f(\mathbf{v}_i^l) + \sum_{j \in \mathcal{O} \setminus i} g(\mathbf{v}_j^{(l)}, P_{i,j}^{(l)}, \alpha_{i,j}^{(l)}) + \mathbf{v}_i^l) \tag{6}$$

Finally, the odorant embeddings $\mathbf{X}_{\mathcal{O}} = \{\mathbf{x}_{\mathcal{O}_i}\}_{i=1}^{\mathcal{N}_{sample}}$ were selected as inputs for subsequent models. In this study, the hyper-parameters L, dimensions of embeddings d, learning rate η, and learning epochs were selected as 6, {32, 64, 128, 256, 512, 1024}, 0.001, and 200, respectively. We considered the feasibility of using molecular 2D and 3D coordinates, and discussed the embedding results. Detailed information regarding atom interaction embedding can be found in other publications (Tsubaki & Mizoguchi, 2019).

3. Results and discussion

3.1. Data analysis

We employed five different molecular structure representations, including odorant molecular parameters (MP), molecular fingerprints (FP), pre-trained molecular graphic embeddings, pre-trained molecular graph transformer embeddings, and atom interaction embeddings in the present study. First, we visualized the numeric vectors in low-dimensional space using Barnes-Hut t-distributed stochastic neighbor embedding (t-SNE) as an unsupervised statistical method. This method has been widely applied for high-dimensional data visualization. The t-SNE presentation of the odorants based on the above-mentioned vectors is illustrated in Fig. 3 and Fig. S1. As reported in previous studies, each odorant contained multi-odor category labels (Shang et al., 2017; Snitz et al., 2013; Kumar et al., 2015; Sánchez-Lengeling et al., 2019; Shang et al., 2021). Therefore, the distribution of odor categories was visualized using colors representing alpha values. Molecular graphic features extracted by Resnet (Fig. 3c) produced a better result than other molecular features because most odorants from the same odor category are clustered together. In contrast, odor cluster overlapping was observed more frequently in the t-SNE map generated from molecular fingerprints (Fig. 3a), molecular parameters (Fig. 3b), and molecular graph transformers (Fig. 3d). Molecular graphs generated from four combined types of pre-trained CNNs produced competitive results compared with other molecular descriptors (Fig. S1). This result demonstrates that odor categories are likely to be more strongly related to molecular graphs than other descriptors. Therefore, we inferred that an odor category identification model based on molecular graphic features would be superior.

3.2. Molecular graphic CNN-based feature analysis

Fig. 4 and Table S2 summarize the performance metrics of the odor category identification model based on molecular graphic feature extraction. The details of model calibration, including the optimal epochs, training loss, and elapsed time, are illustrated in Fig. S2. The pre-trained ResNet with DNNs (6 hidden layers) performed significantly better

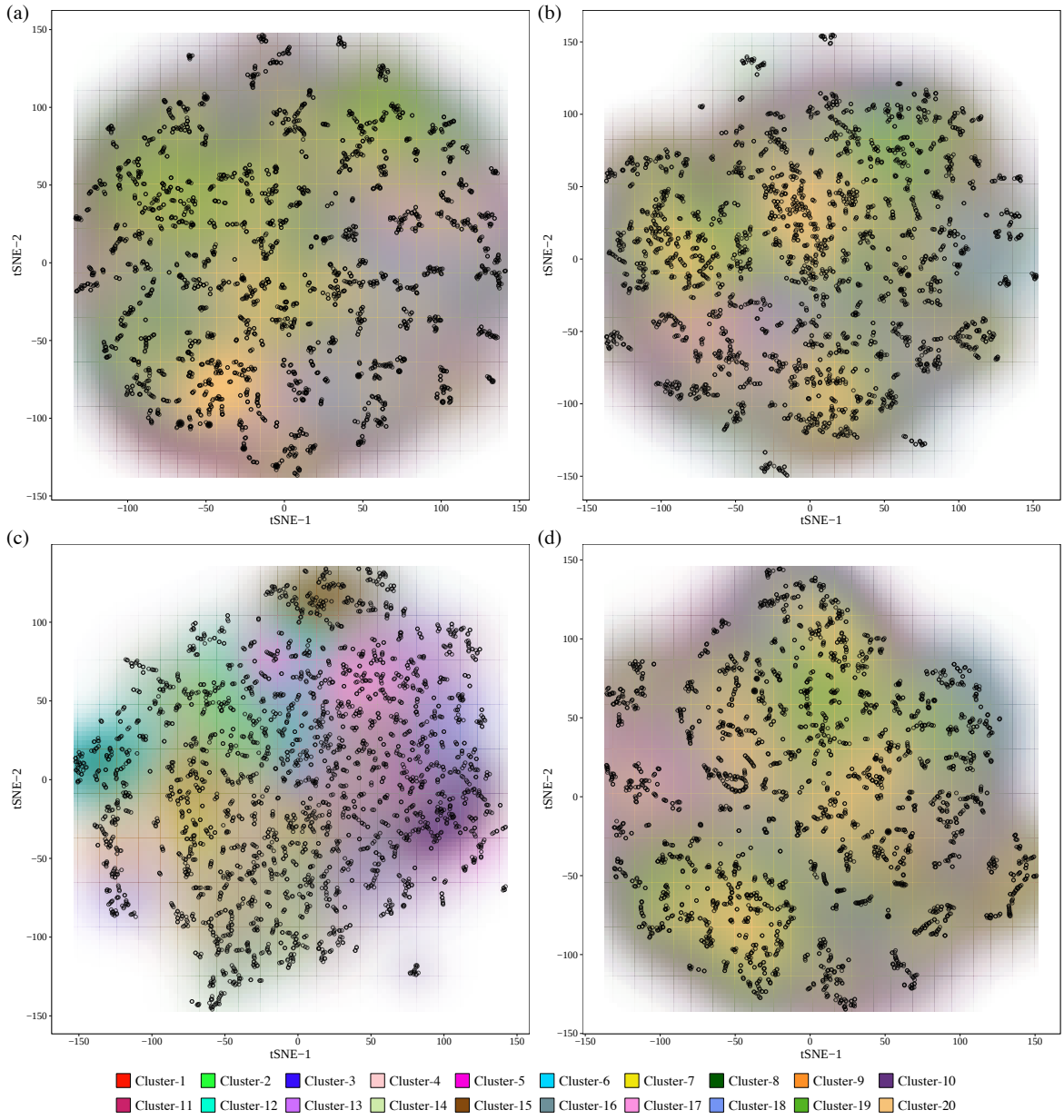


Figure 3: Odorant clustering using Barnes-Hut t-distributed stochastic neighbor embedding (t-SNE) based on (**a**) molecular fingerprints, (**b**) molecular parameters, and (**c**) molecular graphic features extracted via a pre-trained Resnet and (**d**) molecular graph transformer method based on the links between atoms and bonds. tSNE-1 and tSNE-2 were calculated using the t-SNE method. Each point indicates an odorant, colored according to its odor category labels, and the distributions of odor categories are given by the alpha values corresponding to the colors.

than the other models, with the highest AUC (0.877±0.028, $p<0.001$) and F1-score (0.725±0.0278, $p<0.001$) on the test sets. It was followed by the DenseNet (4 hidden layers, AUC 0.876±0.029, F1-score 0.716±0.035), VGG (6 hidden layers, AUC 0.875±0.028, F1-score 0.716±0.033), and AlexNet (5 hidden layers, AUC 0.873±0.029, F1-score 0.723±0.033). The deep residual framework of the most successful model may have overcome the degradation problem that affects deep networks. In addition, the number of hidden layers in the DNN did not play a necessary role in tuning the pre-trained CNN model. To verify the abilities of the models for transfer learning, we compared the prediction performances of the CNNs depending on whether they were pre-trained. These results are illustrated in Fig. 4. We found that the pre-trained models had significantly ($p<0.001$) higher accuracy compared with the models with un-trained CNNs. This indicates that CNNs could learn universal image feature extractors through training with a large dataset (ImageNet). This conclusion is supported by previous research (Liu et al., 2020).

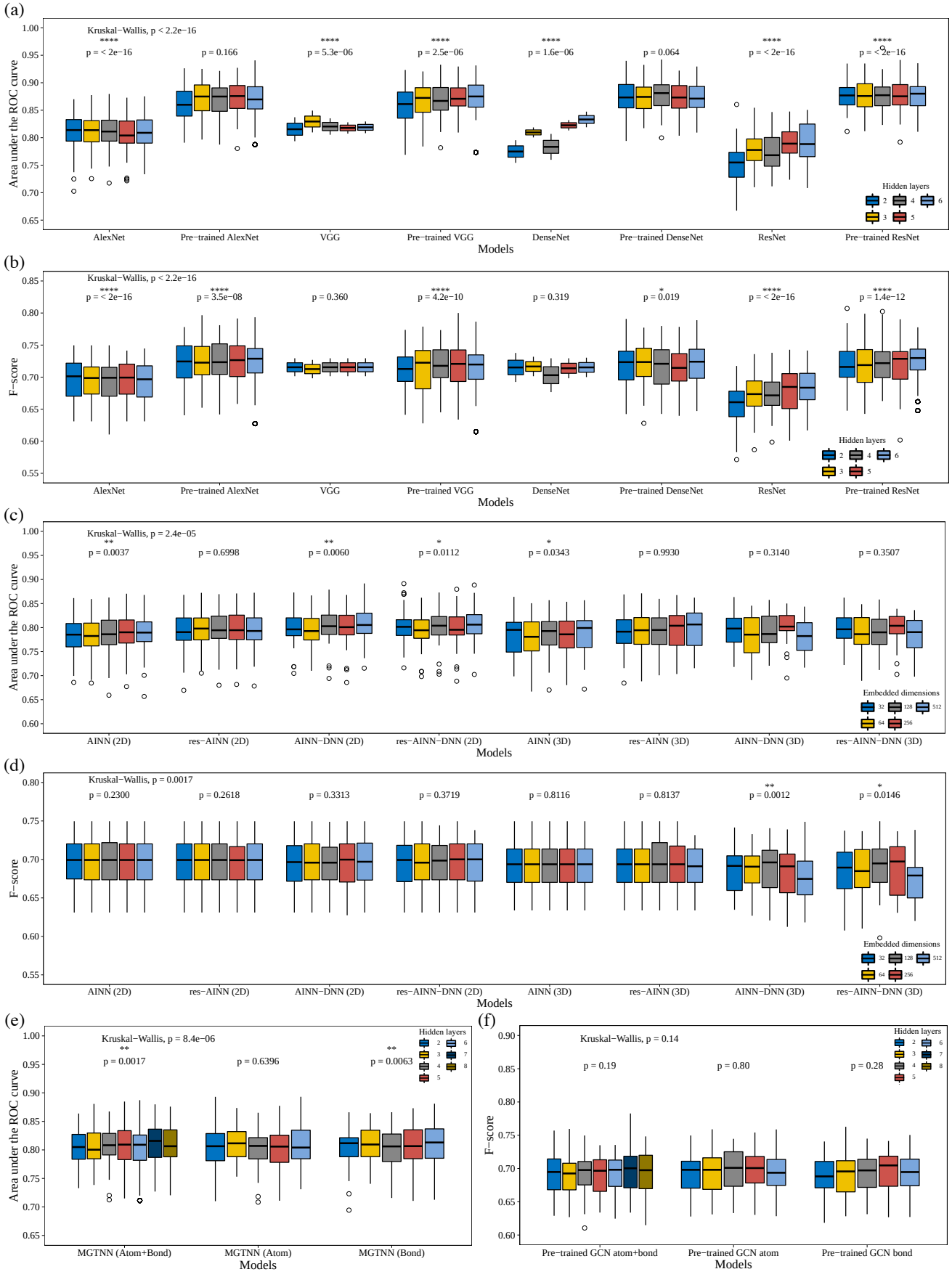


Figure 4: Identification performance of four molecular graphic convolution neural network (MG-CNN) models (**a**, **b**), molecular graph transformer neural network (MGTNN) models (**c**, **d**), and atom interaction neural network (AINN) models (**e**, **f**). Results were evaluated using the nonparametric Wilcoxon signed-rank test.

3.3. Molecular graph transformer-based feature analysis

A summary of the identification accuracy of the MGTNN models is given in Fig. 4 and Table S3. The optimal training epoch, loss, and elapsed time for the MGTNN models are presented in Fig. S3. When the selected atom and bond embeddings were included with 7 hidden layers, the MGTNN model had the highest AUC (0.813±0.035) and F1-score (0.696±0.032) in the test set. In addition, the AUC values for the models independently trained via atom or bond embeddings were 0.812±0.031 and 0.810±0.030, respectively. However, the data were not sufficient to conclude that considering atoms and bonds together produced a significantly more accurate result than when they were included individually ($p>0.001$).

3.4. Atom interaction-based feature analysis

Fig. 4 and Table S4 compare odor sensory category identification according to the molecular features extracted by AINNs. The results indicated that the AINN-DNN model (2D, 512 embedded dimensions) had the highest identification performance in terms of the AUC and F1-score, which were 0.807±0.035 ($p<0.01$) and 0.696±0.023, respectively. However, we cannot claim that the molecular 2D coordinates were better for identification than the 3D coordinates because the analyses for both had a high p-value. Furthermore, the models with residual modules did not exhibit a significant increase, likely because the vanishing gradient is not the critical obstacle limiting AINN performance. In addition, the dimension of atom embedding vectors did not have a significant effect on the accuracy of odor category identification. The optimal training epoch, loss, and elapsed time for AINN models are listed in Fig. S4. We found that the modeling time for 2D coordinates was significantly smaller than that for 3D coordinates ($p<0.001$). This suggests that the presented AINN models do not need spatial embedding for odor sensory identification.

3.5. Performance comparison

To identify the model with the best comprehensive performance for odor category identification, we compared the five types of models in terms of performance metrics, as presented in Fig. 5 and Table 1. The predicted accuracies for each odor sensory category are summarized in Fig. S6-S10. The results confirmed that the model trained using molecular graphic features extracted via a pre-trained ResNet had significantly better performance than the other models (AUC 0.877±0.028, F1-score 0.726±0.028, $p<0.0001$), followed by the AINN-DNN (AUC 0.807±0.035, F1-score 0.696±0.030), MPs (AUC 0.806±0.033, F1-score 0.689±0.031), MGTNN (AUC 0.804±0.028, F1-score 0.692±0.029), and FPs (AUC 0.796±0.036, F1-score 0.688±0.033). This ranking could likely be explained by the high correlation between the olfactory sensory information and the molecular graphic features of the odorants compared with the other molecular descriptors. We found that the AINN-DNN model had the highest precision (0.861±0.038, $p<0.0001$). Fig. S5 illustrates the optimal training epoch, loss, and elapsed time for the above-mentioned models. We confirmed that although more epochs were needed to train the ResNet models, the training time was shorter than that for the AINN-DNN and MGTNN models. The fast convergence speed could contribute to the transfer-learning mechanism. Although the number of parameters was abundant for the ResNet model, we did not train these parameters, but instead used those from the pre-trained models. The pre-trained models could overcome the limitation of insufficient samples for training DNN models. A similar conclusion was found previously (Pezzotti et al., 2017). In summary, we suggest that an end-to-end DNN with molecular graphic features extracted via a pre-trained ResNet is an optimal model for predicting the sensory categories of odorants.

3.6. Discussion

The accurate and effective prediction of odor sensory categories is vital for developing machine-learning-based GCO. To develop an olfaction-based sensory system, we need not only bio-sensors to encode odorants (odor receptor imitation), but also a brain-like odor signal decoding algorithm. Although many studies have examined SORs, most have focused on predicting ODs (Snitz et al., 2013; Debnath & Nakamoto, 2020; Loetsch et al., 2019; Sharma et al., 2021; Kumar et al., 2015). Even though most ODs can be predicted, infrequent ODs were difficult to be identified. For example, Snitz proposed a mostly perfect result in predicting 64 smell percepts with 100 % precision and 102 smells with 90.35 %, but infrequent smells, such as almond, apricot, and chocolate, had been found to have poor prediction performance (Snitz et al., 2013). This could be explained by the extreme imbalance in the data distribution, as well as the insufficient number of training samples. Unlike the above-mentioned studies, we want to find useful odorant structure features for odor sensory categories identification.

Here, we focused on establishing relationships between molecular features and odor sensory categories via an end-to-end learning strategy, which is expected to play a decoding role in bio-olfaction. The MPs and FPs in the present study had poor performance, indicating that focusing solely on physiology-chemical parameters could result in the loss of some critical information related to olfaction. In contrast to relying on tabular features, molecular graph CNNs-based features would be more appropriate for learning useful odor sensory expression. We also considered a transfer learning strategy for dealing with the problem of insufficient training samples. Our results confirm that pre-trained CNNs, combined with a 'vanilla' DNN, can effectively establish relationships between molecular features and odor sensory categories. Furthermore, our data suggest that molecular graphic features are optimal for describing odorant protein interactions according to human olfaction. Existing GCO methods have focused on just 8 ODs in one olfaction sensory evaluation task, as limited by the odor memory of the assessors (Brattoli et al., 2013; Vene et al., 2013). Odor analysis precision is also limited by their odor memory. Therefore,

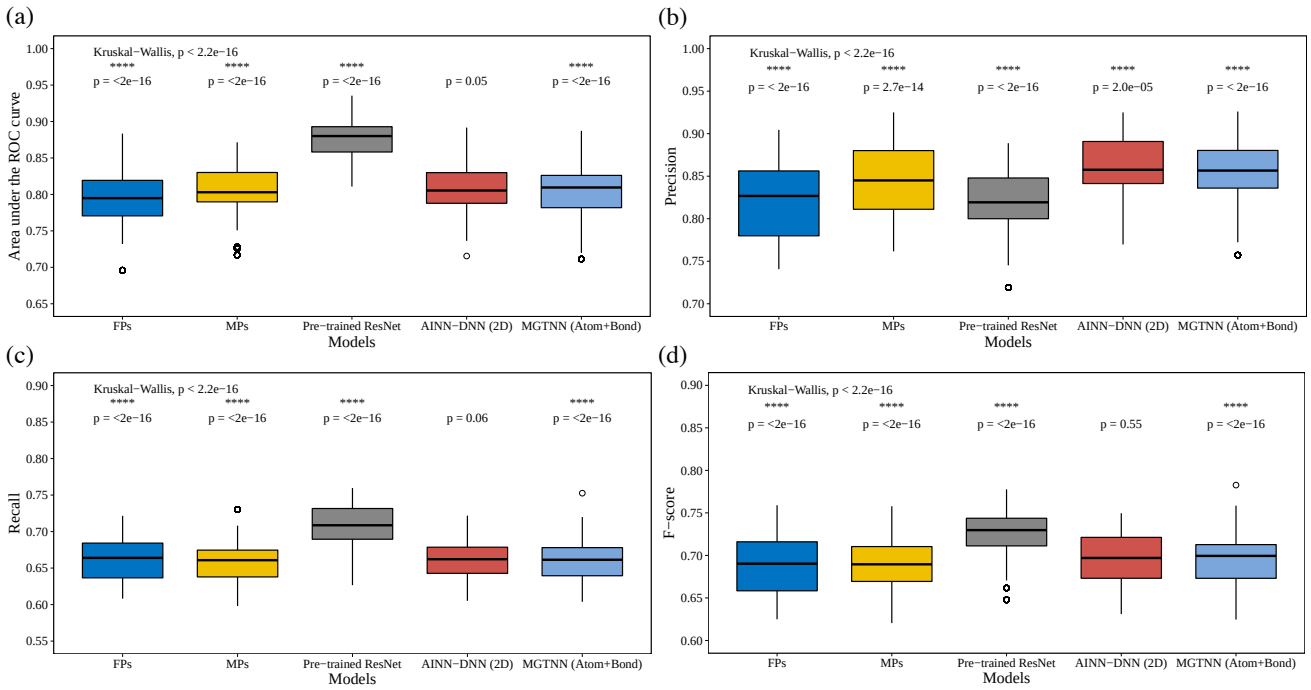


Figure 5: Identification accuracies of DNN models using molecular features extracted via FP-, MP-, MG-CNN-, MGTNN-, and AINN-based DNN models. The data were evaluated using the nonparametric Wilcoxon signed-rank test.

Table 1
Odor sensory category identification accuracy comparison of DNN models using multi-type of odorant structure features.

Model name	AUC-ROC	Precision	Recall	F1-score
FP-based	0.796±0.036	0.82±0.046	0.662±0.029	0.688±0.033
MP-based	0.806±0.033	0.843±0.039	0.658±0.028	0.689±0.031
MG-CNN-based	0.877±0.028	0.822±0.037	0.71±0.028	0.726±0.028
MGT-NN-based	0.804±0.036	0.855±0.036	0.659±0.025	0.692±0.029
AINN-based	0.807±0.035	0.861±0.038	0.662±0.027	0.696±0.03

the proposed model can apply reliable references for human panelists to reduce training costs.

This study has several limitations. First, more attention should be focused on atom interaction-based embeddings, although the AINN in the present study had poor performance. Biological studies have indicated that atom interactions play a critical role in mammal olfaction (Mantel et al., 2022). The poor accuracy of the AINN was likely caused by insufficient odorants and an inappropriate modeling approach. Moreover, we did not consider the electronic interactions between atoms, which may be suitable for olfaction sensory encoding. Self-supervised strategies combined with proper modeling techniques and trained with abundant molecules merit further investigation. Furthermore, synergism, odor neutralization, and the predictability of a fragrance mixture have still not been quantified (Hudon et al., 2000). In the present study, we focus on single odor molecule smell perception prediction, which is not appropriate for modeling odor synergism and neutralization. For fragrance mixture prediction, mass spectral would be feasible for model calibration. In the future, we plan to attempt to improve our framework for molecular structure feature extraction using other algorithms, and to try to explore the feasibility of metric modeling using Riemannian manifolds, such as the Grassmann or symmetric positive definite manifold (Tang et al., 2021). We expect that it will be difficult to find a reasonable algorithm when performing metric learning in Riemannian space. However, this is an interesting problem for future investigation. In addition, data fusion would also be an effective strategy for increasing the accuracy of odor category identification models.

4. Conclusions

In summary, we utilized a DNN-based multi-label classifier for odor sensory category identification using various molecular features. Specifically, we examined the possibility of predicting odor categories based on molecular parameters, fingerprints, and graphics, as well as graph attention network embedding and atom interactions. Our results indicated that molecular 2D graphic data were strongly related to sensory information about olfaction. Extensive

experiments confirmed that a 'vanilla' DNN with molecular graphic features, extracted via ResNet, was optimal for odor perception category identification. We anticipate that transfer learning is a viable and powerful technique for modeling the relationships between molecular structures and odor perception categories. Our proposed approach could be applied in the development of AI-based odor sensors. We believe that this study is among the first to examine the importance of molecular graphic features when establishing relational models between molecular structures and odor sensory categories. Our approach may not only serve as a realistic solution for introducing AI into olfactometry, but may also offer a novel auxiliary tool in the existing GCO by suggesting possible odor categories to the human assessors.

Acknowledgements

This research was supported by China Postdoctoral Science Foundation (2021M703399), the National Nature Science Foundation of China (No. 61803369 and 61801400), and JSPS KAKENHI Grant (No. 18H03782).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/.

CRediT authorship contribution statement

Liang Shang: Conceptualization, Methodology, Experiment, Writing. **Chuanjun Liu:** Conceptualization, Method, Supervision, Reviewing. **Fengzhen Tang:** Supervision, Reviewing. **Bin Chen:** Supervision, Reviewing. **Lianqing Liu:** Reviewing. **Kenshi Hayashi:** Reviewing.

References

Acree, T. E. (1997). GC/Olfactometry GC with a sense of smell. *Anal. Chem.*, *69*, 170A–175A. URL: https://doi.org/10.1021/ac971560s. doi:10.1021/ac971560s.

Arn, H., Acree, T. E., Contis, E. T., Ho, C.-T., Mussinan, C. J., Parliment, T. H., Shahidi, F., & Spanier, A. M. (1998). Flavornet: A database of aroma compounds based on odor potency in natural products. In *Developments in Food Science* (p. 27). Elsevier volume 40. URL: https://www.sciencedirect.com/science/article/pii/S0167450198800290. doi:10.1016/S0167-4501(98)80029-0.

Barwich, A.-S. (2020). What makes a discovery successful? the story of linda buck and the olfactory receptors. *Cell*, *181*, 749–753. URL: https://doi.org/10.1016/j.cell.2020.04.040. doi:10.1016/j.cell.2020.04.040.

Brattoli, M., Cisternino, E., Dambruoso, P. R., de Gennaro, G., Giungato, P., Mazzone, A., Palmisani, J., & Tutino, M. (2013). Gas chromatography analysis with olfactometric detection (GC-O) as a useful methodology for chemical characterization of odorous compounds. *Sens.*, *13*, 16759–16800. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3892869/. doi:10.3390/s131216759.

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2022). Prediction of coffee aroma from single roasted coffee beans by hyperspectral imaging. *Food Chem.*, *371*. doi:10.1016/j.foodchem.2021.131159.

Chacko, R., Jain, D., Patwardhan, M., Puri, A., Karande, S., & Rai, B. (2020). Data based predictive models for odor perception. *Sci. Rep.*, *10*, 17136. URL: https://doi.org/10.1038/s41598-020-73978-1. doi:10.1038/s41598-020-73978-1.

Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., Luo, X., Chen, K., Jiang, H., & Zheng, M. (2020). TransformerCPI: Improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, *36*, 4406–4414. doi:{10.1093/bioinformatics/btaa524}.

Choi, Y., Kim, K., Kim, S., & Kim, D. (2022). Identification of odor emission sources in urban areas using machine learning-based classification models. *Atmos*, *13*, 100156. URL: https://www.sciencedirect.com/science/article/pii/S2590162122000107. doi:https://doi.org/10.1016/j.aeaoa.2022.100156.

Debnath, T., & Nakamoto, T. (2020). Predicting human odor perception represented by continuous values from mass spectra of essential oils resembling chemical mixtures. *PLoS ONE*, *15*, e0234688. URL: https://doi.org/10.1371/journal.pone.0234688. doi:10.1371/journal.pone.0234688.

Dravnieks, A. (1982). Odor quality: Semantically generated multidimensional profiles are stable. *Science*, *218*, 799–801. URL: https://science.sciencemag.org/content/218/4574/799. doi:10.1126/science.7134974.

Fjaeldstad, A. (2018). Testing olfactory function and mapping the structural olfactory networks in the brain. *Dan. Med. J.*, *65*.

Gutiérrez, E. D., Dhurandhar, A., Keller, A., Meyer, P., & Cecchi, G. A. (2018). Predicting natural language descriptions of mono-molecular odorants. *Nat. Commun.*, *9*, 4979. URL: https://doi.org/10.1038/s41467-018-07439-9. doi:10.1038/s41467-018-07439-9.

Huang, Z. J., Zeng, Y. H., Sun, Q. Y., Zhang, W. H., Wang, S. T., Shen, C. H., & Shi, B. (2022). Insights into the mechanism of flavor compound changes in strong flavor baijiu during storage by using the density functional theory and molecular dynamics simulation. *Food Chem.*, *373*. doi:10.1016/j.foodchem.2021.131522.

Hudon, G., Guy, C., & Hermia, J. (2000). Measurement of odor intensity by an electronic nose. *J. Air Waste Manag Assoc.*, *50*, 1750–1578.

Johnson, B. A., Xu, Z., Ali, S. S., & Leon, M. (2009). Spatial representations of odorants in olfactory bulbs of rats and mice: SimilaritiesÍ and differences in chemotopic organization. *J. Comp. Neurol.*, *514*, 658–673. doi:10.1002/cne.22046.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM*, *60*, 84–90. URL: https://doi.org/10.1145/3065386. doi:10.1145/3065386.

Kumar, R., Kaur, R., Auffarth, B., & Bhondekar, A. P. (2015). Understanding the odour spaces: A step towards solving olfactory stimulus-percept problem. *PLoS ONE*, *10*, e0141263. doi:10.1371/journal.pone.0141263.

Liu, C., Liu, P., Zhao, W., & Tang, X. (2020). Visual tracking by structurally optimizing pre-trained CNN. *IEEE Trans. Circuits. Syst. Video Technol.*, *30*, 153–3166. doi:{10.1109/TCSVT.2019.2938038}.

Liu, C., Shang, L., & Hayashi, K. (2019). Co-occurrence-based clustering of odor descriptors for predicting structure-odor relationship. In *IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)* (pp. 1–4). doi:10.1109/ISOEN.2019.8823446.

Loetsch, J., Kringel, D., & Hummel, T. (2019). Machine learning in human olfactory research. *Chem. Senses*, *44*, 11–22. doi:{10.1093/chemse/bjy067}.

Ma, L. J., Gao, M. M., Zhang, L. Q., Qiao, Y., Li, J. X., Du, L. P., Zhang, H. L., & Wang, H. (2022). Characterization of the key aroma-active compounds in high-grade Dianhong tea using GC-MS and GC-O combined with sensory-directed flavor analysis. *Food Chem.*, *378*. doi:10.1016/j.foodchem.2022.132058.

Mantel, M., Fournel, A., Staedlé, I., Oelschlägel, A., Carro, J., Dubreuil, R., Herrier, C., Livache, T., Haehner, A., Hummel, T., Roy, J.-M., & Bensafi, M. (2022). Using a bio-inspired surface resonance plasmon electronic nose for fundamental research on human olfaction. *Sens. Actuators B Chem.*, *350*, 130846. URL: https://www.sciencedirect.com/science/article/pii/S0925400521014143. doi:10.1016/j.snb.2021.130846.

Mao, K., Xiao, X., Xu, T., Rong, Y., Huang, J., & Zhao, P. (2021). Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing*, *457*, 193–202. doi:{10.1016/j.neucom.2021.06.037}.

Maurya, A., Prasad, J., Das, S., & Dwivedy, A. K. (2021). Essential oils and their application in food safety. *Front. Sustain. Food Syst.*, *5*. doi:{10.3389/fsufs.2021.653420}.

Ni, R. J., Yan, H. Y., Tian, H. L., Zhan, P., & Zhang, Y. Y. (2022). Characterization of key odorants in fried red and green huajiao (Zanthoxylum bungeanum maxim. and Zanthoxylum schinifolium sieb. et Zucc.) oils. *Food Chem.*, *377*. doi:10.1016/j.foodchem.2021.131984.

Nozaki, Y., & Nakamoto, T. (2018). Predictive modeling for odor character of a chemical using machine learning combined with natural language processing. *PLoS ONE*, *13*, e0208962. doi:{10.1371/journal.pone.0198475}.

Pezzotti, N., Lelieveldt, B. P. F., van der Maaten, L., Hollt, T., Eisemann, E., & Vilanova, A. (2017). Approximated and user steerable tSNE for progressive visual analytics. *IEEE Trans. Vis. Comput. Graph*, *23*, 1739–1752. doi:{10.1109/TVCG.2016.2570755}.

Sánchez-Lengeling, B., Wei, J. N., Lee, B. K., Gerkin, R. C., Aspuru-Guzik, A., & Wiltschko, A. B. (2019). Machine learning for scent: Learning generalizable perceptual representations of small molecules. *ArXiv*, *abs/1910.10685*.

Shang, L., Liu, C., Chen, B., Tang, F., Liu, L., & Hayashi, K. (2021). Machine-learning-based olfactometry: Odor descriptor clustering based on bayesian embedding model and its prediction from molecular graphic features. *Submitted*, .

Shang, L., Liu, C. J., Tomiura, Y., & Hayashi, K. (2017). Machine-learning-based olfactometer: Prediction of odor perception from physicochemical features of odorant molecules. *Anal. Chem.*, *89*, 11999–12005. doi:10.1021/acs.analchem.7b02389.

Sharma, A., Kumar, R., Ranjta, S., & Varadwaj, P. K. (2021). SMILES to smell: Decoding the structure-odor relationship of chemical compounds using the deep neural network approach. *J. Chem. Inf. Model.*, *61*, 676–688. doi:10.1021/acs.jcim.0c01288.

Sigma-Aldrich (2016). Flavors and fragrances products catalog. *Merck KGaA: Darmstadt*, . URL: https://search.library.wisc.edu/catalog/999610513502121.

Snitz, K., Perl, O., Honigstein, D., Secundo, L., Ravia, A., Yablonka, A., Endevelt-Shapira, Y., & Sobel, N. (2019). Smellspace: An odor-based social network as a platform for collecting olfactory perceptual data. *Chem. Senses*, *44*, 267–278. URL: https://doi.org/10.1093/chemse/bjz014. doi:10.1093/chemse/bjz014.

Snitz, K., Yablonka, A., Weiss, T., Frumin, I., Khan, R. M., & Sobel, N. (2013). Predicting odor perceptual similarity from odor structure. *PLoS Comput. Biol.*, *9*, e1003184. doi:10.1371/journal.pcbi.1003184.

Tang, F., Feng, H., Tino, P., Si, B., & Ji, D. (2021). Probabilistic learning vector quantization on manifold of symmetric positive definite matrices. *Neural Netw.*, *142*, 105–118. doi:{10.1016/j.neunet.2021.04.024}.

Tsubaki, M., & Mizoguchi, T. (2019). Fast and accurate molecular property prediction: learning atomic interactions and potentials with neural networks. *J. Phys. Chem.*, *10*, 2066–2067. doi:{10.1021/acs.jpclett.9b00301}.

Turner, L., Dawda, D., Wagstaff, C., Gawthrop, F., & Lignou, S. (2021). Influence of harvest maturity on the aroma quality of two celery (Apium graveolens) genotypes. *Food Chem.*, *365*. doi:10.1016/j.foodchem.2021.130515.

Vene, K., Seisonen, S., Koppel, K., Leitner, E., & Paalme, T. (2013). A method for GC-Olfactometry panel training. *Chemosensory Perception*, *6*, 179–189. URL: https://doi.org/10.1007/s12078-013-9156-x. doi:10.1007/s12078-013-9156-x.

Warikoo, N., Chang, Y.-C., & Hsu, W.-L. (2021). LBERT: Lexically aware transformer-based bidirectional encoder representation model for learning universal bio-entity relations. *Bioinformatics*, *37*, 404–412. doi:{10.1093/bioinformatics/btaa721}.

Wu, W. D., Zhan, J. L., Tang, X. Y., Li, T., & Duan, S. N. (2022). Characterization and identification of pork flavor compounds and their precursors in Chinese indigenous pig breeds by volatile profiling and multivariate analysis. *Food Chem.*, *385*. doi:10.1016/j.foodchem.2022.132543.