Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI4Tech: AI Enabling Critical Technologies

# Multi-Hierarchical Fine-Grained Feature Mapping Driven by Feature Contribution for Molecular Odor Prediction

**Hongxin Xie**[1], **Jiande Sun**[1*], **Fanfu Xue**[2], **Zifei Han**[3], **Shanshan Feng**[1] and **Qi Chen**[4]
[1]Shandong Normal University, Jinan, China
[2]Shandong University, Jinan, China
[3]ZJU-UIUC Institute, Haining, China
[4]Huangshan University, Huangshan, China
139536763397@163.com, jiandesun@hotmail.com, fanfuxue@outlook.com, zifeih2@illinois.edu, fengshan2017@sdnu.edu.cn, chenqijiyu@126.com.

## Abstract

Molecular odor prediction is the process of using a molecule's structure to predict its smell. While accurate prediction remains challenging, AI models can suggest potential odors. Existing methods, however, often rely on basic descriptors or handcrafted fingerprints, which lack expressive power and hinder effective learning. Furthermore, these methods suffer from severe class imbalance, limiting the training effectiveness of AI models. To address these challenges, we propose a Feature Contribution-driven Hierarchical Multi-Feature Mapping Network (HMFNet). Specifically, we introduce a fine-grained, Local Multi-Hierarchy Feature Extraction module (LMFE) that performs deep feature extraction at the atomic level, capturing detailed features crucial for odor prediction. To enhance the extraction of discriminative atomic features, we integrate a Harmonic Modulated Feature Mapping (HMFM). This module dynamically learns feature importance and frequency modulation, improving the model's capability to capture relevant patterns. Additionally, a Global Multi-Hierarchy Feature Extraction module (GMFE) is designed to learn global features from the molecular graph topology, enabling the model to fully leverage global information and enhance its discriminative power for odor prediction. To further mitigate the issue of class imbalance, we propose a Chemically-Informed Loss (CIL). Experimental results demonstrate that our approach significantly improves performance across various deep learning models, highlighting its potential to advance molecular structure representation and accelerate the development of AI-driven technologies.

## 1 Introduction

Odor, a key sensory characteristic, significantly influences consumer experience and product perception [Keller *et al.*, 2017]. By understanding the molecular structure-odor relationship, AI models can predict how molecules interact with the human olfactory system [Sharma *et al.*, 2021; Liu *et al.*, 2022a]. In personalized medicine, AI-based odor prediction helps create custom scents for individual health and wellness needs [Zhang *et al.*, 2018]. These innovations highlight AI's transformative potential to advance critical technologies in biotechnology, health, and environmental sciences.

Early approaches primarily relied on chemistry-based statistical machine learning techniques [Ji *et al.*, 2023]. However, with advancements in artificial intelligence, modern methods now predominantly utilize deep learning technologies, such as graph neural networks [Wu *et al.*, 2020; Lee *et al.*, 2023], to model the intricate interactions between molecular structures and odor. Existing methods face significant challenges in capturing the complex relationship between molecular structure and odor. Traditional atomic-level features and handcrafted fingerprints fail to adequately represent these interactions due to their limited expressiveness. Furthermore, class imbalance in odor descriptors exacerbates model bias [Saini and Ramanathan, 2022], hindering effective prediction.

To address the aforementioned issues, we propose a Hierarchical Multi-Feature Mapping Network (HMFNet). Specifically, it consists of a fine-grained Local Multi-Hierarchy Feature Extraction module (LMFE) and a Global Multi-Hierarchy Feature Extraction module (GMFE). LMFE performs deep feature extraction on single-structure matrices, such as atomic, bond, and molecular structure data, to capture fine-grained local features that are more beneficial for odor prediction. To better extract discriminative features from atomic information, we design a Harmonic Modulated Feature Mapping (HMFM), which enhances the model's efficiency in utilizing molecular features by dynamically learning feature importance and applying frequency modulation. This improves the model's ability to handle the complex relationships between molecules and odors. GMFE learns global features from the molecular graph's topological structure, molecular fingerprints, and global chemical properties, leveraging global information to improve discriminative power for odor prediction and further enhancing the model's capability to manage complex molecule-odor relationships. Additionally, we integrate components such as label correlation

*Corresponding author.

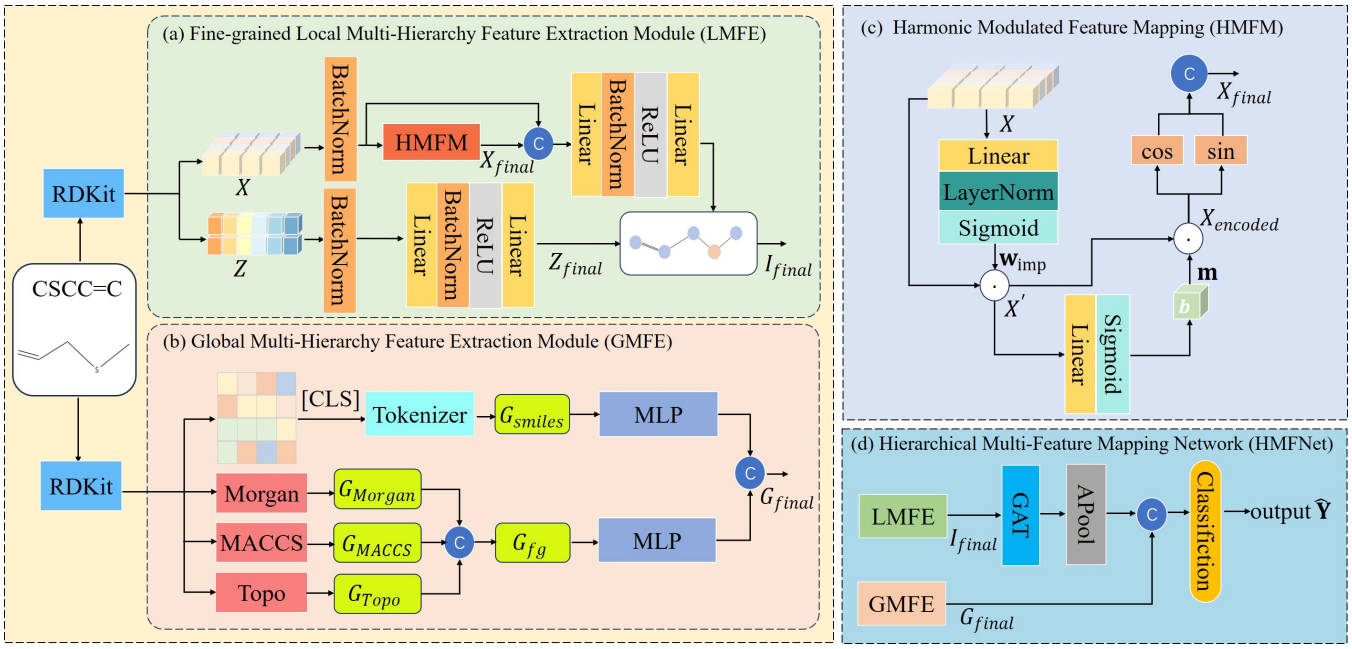


Figure 1: (a): The concrete structure of fine-grained Local Multi-Hierarchical Feature Extraction; (b): The concrete structure of Global Multi-Hierarchy Feature Extraction; (c): The concrete structure of Harmonic Modulated Feature Mapping; (d): The architecture of the proposed overall framework.

to design a Chemically-Informed Loss (CIL), specifically tailored to address the class imbalance problem in molecular odor prediction, thereby improving the model's performance. Our main contributions are as follows:

- We propose LMFE that performs deep feature extraction to capture fine-grained features critical for odor prediction. To further enhance the extraction of discriminative atomic features, we design HMFM. This module dynamically learns feature importance and applies frequency modulation, improving the model's ability to capture complex odor-molecular relationships.
- We propose an improved loss function, Chemically-Informed Loss, which incorporates multiple components. This multi-faceted design addresses issues of class imbalance, improves the model's focus on minority classes, and fosters better learning of label co-occurrence relationships.
- We propose HMFNet, a multi-hierarchical framework that enhances odor prediction by integrating GMFE into LMFE, effectively capturing feature information from local to global levels. Experimental results confirm that, guided by the CIL, our approach achieves state-of-the-art performance.

## 2 Related Work

Molecular odor prediction has become a key research focus across chemistry, neuroscience, and computer science. With the advent of machine learning, many studies now employ computational methods to predict the olfactory properties of molecules. Early research explored the relationship between molecular structure and odor through chemical parameters. For example, PaDEL-Descriptor [Yap, 2011] calculates 797 molecular descriptors and 10 types of fingerprints, including electro-topological state descriptors and molecular volume, which are essential for quantitative structure–activity relationship (QSAR) studies. Despite its extensive library, PaDEL-Descriptor faces challenges in processing speed and its ability to handle large molecules. To overcome these limitations, Mordred [Moriwaki *et al.*, 2018] introduced a more advanced descriptor calculation tool, capable of computing over 1,800 2D and 3D molecular descriptors. Mordred is at least twice as fast as PaDEL and can handle large molecular descriptors that other tools cannot. With its high performance, ease of use, and comprehensive descriptor library, Mordred has become a key resource in cheminformatics, especially for structure–property relationship studies. While descriptor-based feature extraction remains important, machine learning techniques are increasingly driving advancements in molecular odor prediction. Graph neural networks (GNNs) [Wu *et al.*, 2020] have proven effective in modeling the complex relationship between molecular structure and odor perception. More recently, Lee et al. [Lee *et al.*, 2023] employed GNNs to create an odor mapping that preserves perceptual relationships, facilitating quality prediction for uncharacterized odor molecules. In prospective validation on 400 unseen odor samples, the POM model's odor profiles were closer to the training group's mean than the median, confirming its reliability as a prediction tool. This model outperformed traditional cheminformatics methods, effectively encoding the structure–odor relationship. This approach not only enhances our understanding of odor prediction but also advances molecular odor prediction.

# 3 Methodology

## 3.1 Overview

Molecular odor prediction aims to predict odor descriptors based on molecular structures, offering significant benefits in fragrance design, environmental monitoring, and personalized health products. However, existing methods face challenges in capturing the complex relationships between molecular structures and odor properties, as well as dealing with data imbalance, which affects model performance.

To address these issues, we propose a Hierarchical Multi-Feature Mapping Network (HMFNet), the detailed structure is shown in Figure 1. Our method includes a local to global molecular feature extraction technique, which enhances feature richness by incorporating atomic-level, bond-level, and global descriptors. This allows the model to better capture the complex relationships between molecular structure and odor, improving prediction accuracy. For a detailed description, refer to Sections 3.2 and 3.3. We also introduce a novel molecular feature mapping method, Harmonic Modulated Feature Mapping (HMFM), which dynamically adjusts feature contributions through importance learning and frequency modulation. This enables the model to effectively capture complex relationships, improving the overall predictive performance of molecular odor characteristics, the specifics are discussed in Section 3.2. Lastly, we design a Chemically-Informed Loss (CIL) function to address class imbalance and inter-label dependencies. By incorporating structural similarity constraints, label correlation integration, and adaptive energy adjustments, CIL enhances the model's ability to handle imbalanced datasets, particularly improving predictions for minority odor descriptors, detailed description in section 3.4.

## 3.2 Fine-grained Local Multi-Hierarchy Feature Extraction Module (LMFE)

To capture fine-grained local features that are more beneficial for odor prediction, we design a fine-grained Local Multi-Hierarchy Feature Extraction module (LMFE). LMFE performs deep feature extraction on single-structure matrices such as atomic and bond data. Specifically, based on RDKit [Wong *et al.*, 2024], for each atom $v$, its feature representation is $X_v^A$, where $A$ represents the characteristic number of atoms. For each chemical bond $o$, its feature representation is $Z_o^P$, where $P$ represents the characteristic number of chemical bonds.

In molecular odor prediction tasks, existing methods [Saini and Ramanathan, 2022; Schicker *et al.*, 2023] struggle to learn non-smooth objective functions and address the issue of mixed feature dimensions, make traditional feature mapping methods insufficient for effectively capturing these multidimensional relationships. To overcome this issue, we propose a novel Harmonic Modulated Feature Mapping (HMFM) method based on feature importance learning and frequency modulation. To achieve this, we introduce a feature importance layer and a frequency modulation layer. The modulation, combined with base frequencies, forms periodic and phase encoding, effectively capturing the complex relationships between molecular features and odors.

Specifically, for each atomic feature, we learn its relative importance in odor prediction through the feature importance layer, enabling the model to adaptively adjust each feature's impact on the final prediction. Given the input feature $X \in \mathbb{R}^{N\times A}$, where $N$ is the batch size, the feature importance weight $\mathbf{w}_{\text{imp}} \in \mathbb{R}^{N\times A}$ is calculated through the following steps:

$$\mathbf{w}_{\text{imp}} = \sigma\left(LN\left(Linear(X)\right)\right) \tag{1}$$

where $\sigma$ is the Sigmoid activation function, $Linear(\cdot)$ is a linear transformation, and $LN(\cdot)$ is layer normalization. The resulting $\mathbf{w}_{\text{imp}}$ represents the learned importance of each atomic feature.

Next, the weighted features $X'$ are obtained by element-wise multiplication of the feature $X$ and the importance weights $\mathbf{w}_{\text{imp}}$:

$$X' = \mathbf{w}_{\text{imp}} \odot X \tag{2}$$

where $\odot$ denotes the element-wise multiplication.

To apply different frequency responses to various features, we design a frequency modulation mechanism. By learning to modulate atomic features, we dynamically adjust the frequency of each feature, enabling dynamic adaptation between feature importance and frequency. Specifically, we apply the frequency modulation layer on the input features $X'$. This layer generates a modulation coefficient $\mathbf{f} \in \mathbb{R}^{N\times D}$, where $D$ represents the output feature dimension. The frequency modulation coefficient is computed as:

$$\mathbf{f} = \sigma\left(Linear(X')\right) \tag{3}$$

The obtained frequency modulation coefficient $\mathbf{f}$ is then multiplied element-wise with a base frequency coefficient $\mathbf{b}$ to obtain the modulated frequency coefficients $\mathbf{m}$:

$$\mathbf{m} = \mathbf{b} \odot \mathbf{f} \tag{4}$$

The base frequency coefficient $\mathbf{b}$ is pre-calculated using the formula $\mathbf{b} = 2\pi\sigma' \frac{t}{D}$, where $\sigma'$ represents the standard deviation, and $t$ denotes the index of the feature dimension.

Once the modulated frequency coefficients are obtained, we combine them with the weighted feature $X'$ to generate periodic and phase encodings. Specifically, the encoding result is calculated through the following steps:

$$X_{encoded} = \mathbf{m} \odot X' \tag{5}$$

Then, the cosine and sine values of $X_{encoded}$ are computed: $\cos(X_{encoded}), \sin(X_{encoded})$.

Finally, the feature mapping $X_{final}$ is obtained by concatenating the cosine and sine values along the feature dimension:

$$X_{final} = concat(\cos(X_{encoded}), \sin(X_{encoded}), \dim = -1) \tag{6}$$

where $concat$ denotes the concatenation operation.

The final features, $X_{final}$ and the bond features $Z$, are passed through two encoders to form the molecular graph representation $I_{final}$. This representation is then processed through a GAT Network [Veličković *et al.*, 2017] and a pooling layer to obtain the enhanced features $I'_{final}$.

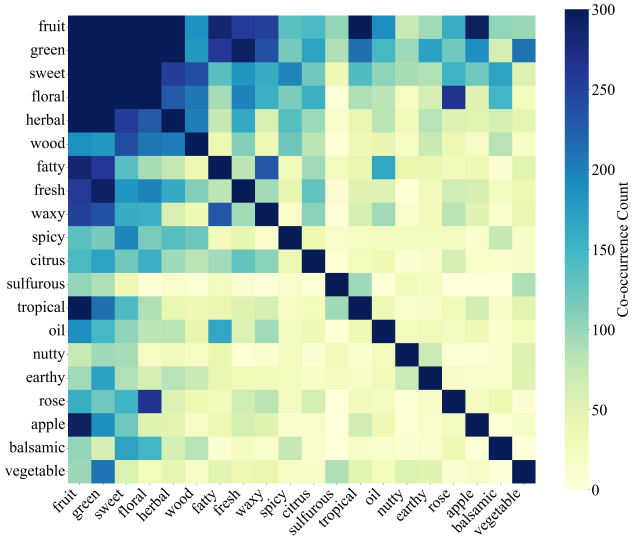


Figure 2: Co-ocurrence matrix of Top 20 odor descriptors.

### 3.3 Global Multi-Hierarchy Feature Extraction Module (GMFE)

Additionally, to further enhance the model's ability to handle complex molecule-odor relationships, we design a Global Multi-Hierarchy Feature Extraction module (GMFE). GMFE learns global features from the molecular fingerprints and global chemical properties, effectively leveraging global information to improve the model's discriminative power for odor prediction. We compute three types of molecular fingerprints based on RDKit: Morgan fingerprint [Rogers and Hahn, 2010], MACCS fingerprint [Durant *et al.*, 2002], and Topological fingerprint [Nilakantan *et al.*, 1987]. The feature vectors for Morgan, MACCS, and Topological fingerprints are denoted as $G_{Morgan}$, $G_{MACCS}$, $G_{Topo}$. By concatenating these fingerprint vectors, we obtain the overall molecular fingerprint representation:

$$G_{fg} = concat(G_{Morgan}, G_{MACCS}, G_{Topo}) \tag{7}$$

where $concat$ denotes the concatenation operation.

We utilize a Transformer-based [Vaswani, 2017] approach for the SMILES strings to obtain a global chemical property representation of the molecule. The molecules are labeled and embedded into vectors, with the resulting feature representation denoted as $G_{smiles}$.

The resulting $G_{fg}$ and $G_{smiles}$ are passed through two identical MLP layers to obtain the final $G_{final}$ is then combined with the enhanced $I'_{final}$ to form the final representation.

### 3.4 Chemically-Informed Loss (CIL)

To address data imbalance in molecular odor prediction, we propose a Chemically-Informed Loss function comprising four key components: weighted binary cross-entropy (BCE) loss [Mao *et al.*, 2023], chemical property energy loss, sample-level multi-label constraint loss, and label correlation loss. Each component is designed based on a deep understanding of molecular features, aiming to overcome the limitations of existing methods in handling complex chemical information and multi-label prediction. To address the class imbalance in odor descriptors, we introduce a weighted BCE loss. Specifically, the weight $w_j$ for each class is computed based on the ratio of positive to negative samples and is adjusted dynamically during training:

$$w_j = \frac{W_{\text{neg},j}}{W_{\text{pos},j}}, \quad w_j \in [0.1, 10] \tag{8}$$

where $W_{\text{pos},j}$ and $W_{\text{neg},j}$ represent the numbers of positive and negative samples for the odor descriptor $j$. To prevent training instability caused by very large or very small weights, the weights are limited to the range $[0.1, 10]$.

$$\mathcal{L}_{\text{basis}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} w_j \left[Y_{i,j}\log(\hat{Y}_{i,j}) + (1-Y_{i,j})\log(1-\hat{Y}_{i,j})\right] \tag{9}$$

where $N$ is the batch size, and $M$ is the number of classes (i.e., the number of odor descriptors), $Y_{i,j} \in \{0,1\}$ indicates whether molecule $i$ has the odor descriptor $j$ (1 means present, and 0 means absent). $\hat{Y}_{i,j}$ represents the predicted probability of the model, which indicates the likelihood that sample $i$ has the odor descriptor $j$.

Additionally, we introduce an "energy" [Choi *et al.*, 2023] function related to molecular odor features, which sets a target energy for each odor descriptor, constraining the model's learning process to ensure its predictions align with chemical properties and physical laws. Specifically, if the predicted probability of molecule $i$ for odor descriptor $j$ is denoted as $\hat{Y}_{i,j}$, the energy $E_{energy}(j)$ of odor descriptor $j$ is defined as the average prediction probability of this descriptor across the entire sample set:

$$E_{energy}(j) = \frac{1}{N}\sum_{i=1}^{N}\hat{Y}_{i,j} \tag{10}$$

We introduce a constraint loss based on chemical property energy, with energy $m_{\text{in}}$ and $m_{\text{out}}$. The target energy $m_{\text{in}}$ is set to 1 for samples with odor descriptors, and $m_{\text{out}}$ is set to 0 for those without. These targets are optimized using a label co-occurrence matrix $C_{\text{co-occurrence}}$, reflecting the frequency of odor descriptor co-occurrence. Descriptors that frequently co-occur are assigned higher energy targets, improving the model's understanding of their interrelationships. The energy target formula is:

$$m_{\text{in}} = 1 + c \cdot \text{diag}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i^T Y_i\right) \tag{11}$$

$$m_{\text{out}} = c \cdot \text{diag}\left(\frac{1}{N}\sum_{i=1}^{N} (1-Y_i)^T (1-Y_i)\right) \tag{12}$$

Here, we conducted comparative experiments on the hyperparameters $c = 0.2$ with detailed results provided in Appendix A.2. $c$ controls the extent to which label co-occurrence relationships influence the adjustment of energy targets. This

| Method | | | | | | Evaluation Metrics | |
|---|---|---|---|---|---|---|---|
| Node | Edge | Fingerprint | Token | HMFM | CIL | F1 score | AUROC |
| ✓ | × | × | × | × | × | 0.3400 | 0.9239 |
| ✓ | ✓ | × | × | × | × | 0.4167 | **0.9356** |
| ✓ | ✓ | × | × | ✓ | × | <u>0.4393</u> | <u>0.9337</u> |
| ✓ | ✓ | × | × | ✓ | ✓ | **0.4757** | 0.9233 |
| ✓ | ✓ | ✓ | × | × | × | 0.4385 | 0.9207 |
| ✓ | ✓ | ✓ | ✓ | × | × | 0.4400 | 0.9221 |
| ✓ | ✓ | ✓ | ✓ | ✓ | × | <u>0.4508</u> | <u>0.9266</u> |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **0.4861** | **0.9316** |

Table 1: Ablation study results of key components. The best performance is highlighted in bold and the follow-up is highlighted in underlined.

ensures that the energy target adjustment is neither excessively amplified (avoiding excessively high energy targets) nor too small (which would weaken the impact of energy adjustment on model training), $Y_i$ represents the label vector for molecule $i$, and $\mathrm{diag}(\cdot)$ denotes the extraction of diagonal elements from the matrix, which indicates the co-occurrence frequency of different descriptors.

The chemical property energy loss is:

$$\mathcal{L}_{\text{class}} = \sum_{j=1}^{M} \left[ \sum_{i:Y_{i,j}=1} \max(0, E_{energy}(j) - m_{\text{in}})^2 \right] + \sum_{j=1}^{M} \left[ \sum_{i:Y_{i,j}=0} \max(0, m_{\text{out}} - E_{energy}(j))^2 \right] \tag{13}$$

To further enhance multi-label prediction performance, we designed a sample-level multi-label constraint loss. The expected energy for a sample is adjusted based on the label count:

$$E_{\text{expected}}(i) = e_1 + e_2 \cdot \sum_{j=1}^{M} Y_{i,j} \tag{14}$$

Here, $e_1 + e_2 = 1$ are hyperparameters. $e_1$ represents the baseline expected energy for each sample, ensuring that the model does not generate extreme energy targets due to an insufficient or excessive number of labels. $e_2$ is a modulation factor that ensures the increase in label count smoothly influences the expected energy of the sample, preventing an excessive number of labels from leading to overly large expected energies. The detailed results provided in Appendix A.2.

The loss is calculated based on the difference between the sample's predicted energy and the expected energy:

$$\mathcal{L}_{\text{sample}} = \frac{1}{N} \sum_{i=1}^{N} \left[ \max(0, E_{\text{expected}}(i) - \sum_{j=1}^{M} \hat{Y}_{i,j})^2 \right] \tag{15}$$

We introduce the label correlation loss, designed to minimize the discrepancy between the predicted correlation and the true label correlation. The correlation between labels is measured using the inner product of the label matrix, while the predicted correlation is computed through the inner product of the predicted outputs:

$$\mathcal{L}_{\text{col}} = \| \frac{1}{N} \sum_{i=1}^{N} \hat{Y}_i \hat{Y}_i^T - \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^T \|_2^2 \tag{16}$$

Here, $\hat{Y}_i$ represents the predicted value for the $i$-th molecule, $Y_i$ represents the true value for the $i$-th molecule, $\| \cdot \|_2^2$ represents the squared L2 norm, which is equivalent to the squared Euclidean distance.

Finally, the weighted sum of all loss terms constitutes the total loss function:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{basis}} + \lambda_2 \mathcal{L}_{\text{class}} + \lambda_3 \mathcal{L}_{\text{sample}} + \lambda_4 \mathcal{L}_{\text{col}} \tag{17}$$

Here, we conducted comparative experiments on the hyperparameters $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ with detailed results provided in Appendix A.2.

| Method | F1 score | AUROC |
|---|---|---|
| GCN [Kipf and Welling, 2016] | 0.3701 | 0.9271 |
| GCN+HMFM | 0.3910 | 0.9296 |
| GAT [Veličković *et al.*, 2017] | 0.3953 | 0.9274 |
| GAT+HMFM | 0.4066 | 0.9296 |
| MPNN [Gilmer *et al.*, 2017] | 0.4235 | 0.9304 |
| MPNN+HMFM | 0.4338 | 0.9314 |
| AFP [Xiong *et al.*, 2019] | 0.4429 | 0.9240 |
| AFP+HMFM | 0.4728 | 0.9255 |
| SMPGNN [Leeney and McConville, 2024] | 0.4448 | 0.9175 |
| SMPGNN+HMFM | 0.4550 | 0.9231 |
| GCast [Wang *et al.*, 2024b] | 0.4622 | 0.9288 |
| GCast+HMFM | 0.4677 | 0.9289 |
| GSAGE [Huang and Chen, 2024] | 0.3858 | 0.9284 |
| GSAGE+HMFM | 0.4187 | 0.9295 |
| GIN [Wang *et al.*, 2024a] | 0.3973 | 0.9303 |
| GIN+HMFM | 0.4158 | 0.9304 |
| HMFNet | **0.4861** | **0.9316** |

Table 2: Performance comparison with different models, as well as an evaluation of the performance of HMFM within mainstream deep learning models.

## 4 Experiments

In this section, we solve several key challenges in molecular odor prediction by exploring the following research questions:

- Q1: Can multi-level feature extraction effectively improve the performance of molecular odor prediction?

| Method | F1 score | AUROC |
|---|---|---|
| GCN+GRFF [Wacker and Filippone, 2022] | 0.3905 | 0.9265 |
| GCN+RFF [Mitra and Kaddoum, 2022] | 0.3833 | 0.9275 |
| GCN+PE [Yuan *et al.*, 2023] | 0.3807 | 0.9271 |
| GCN+LEE [Yamada, 2024] | 0.3908 | 0.9295 |
| GCN+HMFM | 0.3910 | **0.9296** |
| GCN+HMFM+CIL | **0.4560** | 0.9248 |
| MPNN+GRFF [Wacker and Filippone, 2022] | 0.4030 | 0.9308 |
| MPNN+RFF [Mitra and Kaddoum, 2022] | 0.4114 | **0.9322** |
| MPNN+PE [Yuan *et al.*, 2023] | 0.4181 | 0.9309 |
| MPNN+LEE [Yamada, 2024] | 0.4238 | 0.9299 |
| MPNN+HMFM | 0.4338 | 0.9314 |
| MPNN+HMFM+CIL | **0.4791** | 0.9306 |
| GAT+GRFF [Wacker and Filippone, 2022] | 0.4022 | 0.9246 |
| GAT+RFF [Mitra and Kaddoum, 2022] | 0.3655 | 0.9251 |
| GAT+PE [Yuan *et al.*, 2023] | 0.4023 | 0.9264 |
| GAT+LEE [Yamada, 2024] | 0.3887 | 0.9254 |
| GAT+HMFM | 0.4066 | 0.9286 |
| GAT+HMFM+CIL | **0.4692** | **0.9289** |

Table 3: Experimental results of harmonic modulated feature mapping and Chemically-Informed Loss.

- Q2: Does Harmonic Modulated Feature Mapping improve the performance of representative deep models?
- Q3: Does Chemically-Informed Loss improve molecular odor prediction?
- Q4: Does the proposed design achieve the best performance of molecular odor prediction at present?

## 4.1 Experiment Setting

**Datasets.** The dataset used in this study is sourced from the Leffingwell PMP 2001 [Leffingwell, 2005] and the GoodScents [Flavor, 2018]. This dataset is currently the largest of its kind, the detailed dataset can be found in Appendix A.1, the dataset is characterized by significant label imbalance, as evidenced by the long-tail distribution of odor descriptors. Moreover, previous studies emphasize the importance of considering label dependency information[Alvares-Cherman *et al.*, 2012]. Given that many odor descriptors occur infrequently, we present a co-occurrence matrix for the top 20 descriptors in Figure 2 for visual clarity.

**Comparsion Setup.** To validate the effectiveness of each component in our multi-level feature extraction, we conducted ablation experiments on the atomic features, chemical bond features, fingerprint features, and Transformer-based string features. Additionally, to validate the superiority of HMFNet, we compare it against mainstream and state-of-the-art methods in this domain. We also conduct dedicated evaluations to assess the effectiveness of the proposed HMFM module. These models include Graph Convolutional Network (GCN) [Kipf and Welling, 2016], Graph Attention Network (GAT) [Veličković *et al.*, 2017], Attentive FP (AFP) [Xiong *et al.*, 2019], Substructure Matching Pretrained GNN (SMPGNN) [Leeney and McConville, 2024], Cross-scale Graph Propagation (GCast) [Wang *et al.*, 2024b], Graph Sample and Aggregation (GSAGE) [Huang and Chen, 2024], Graph Isomorphism Network (GIN) [Wang *et al.*, 2024a], and Message Passing Neural Network (MPNN) [Gilmer *et al.*, 2017]. To further demonstrate the superiority of HMFM, we compared it with four state-of-the-art feature mapping methods: Gaussian Random Fourier Features (GRFF) [Wacker and Filippone, 2022], Random Fourier Features (RFF) [Mitra and Kaddoum, 2022], Positional Encoding (PE) [Yuan *et al.*, 2023], and Laplacian Eigenvector Encoding (LEE) [Yamada, 2024]. Finally, we performed a study on the Chemically-Informed Loss (CIL) to prove its efficacy. By comparing our framework with existing methods, we demonstrated that our approach achieves the best performance in molecular odor prediction.

| Method | F1 score | AUROC |
|---|---|---|
| HMFNet+HIL [Kim *et al.*, 2024] | 0.3308 | 0.9174 |
| HMFNet+MTL [Liu *et al.*, 2022b] | 0.3784 | 0.9113 |
| HMFNet+BCE [Guo *et al.*, 2021] | 0.3845 | 0.9292 |
| HMFNet+ASL [Ridnik *et al.*, 2021] | 0.4297 | 0.9209 |
| HMFNet+AFL [Xie *et al.*, 2025] | 0.4632 | 0.9294 |
| HMFNet+CIL | **0.4861** | **0.9316** |

Table 4: Experimental results of the Hierarchical Multi-Feature Mapping Network with different loss functions.

**Evaluation Metrics.** We selected two indicators widely used in multi-label molecular odor prediction tasks, including F1 score and Area Under the Receiver Operating Characteristic Curve (AUROC). F1 score emphasizes the balance between precision and recall, which means it increases when the model becomes better at correctly identifying positive samples (i.e., reducing false positives and false negatives). AUROC, measures the model's overall ability to distinguish between the positive and negative classes.

## 4.2 Q1: Ablation Study of Hierarchical Feature Extraction

In this section, we demonstrate the effectiveness of the Hierarchical Feature Extraction through the ablation experiment of feature extraction. As shown in Table 1, when using only one feature type, the model's performance is limited. Specifically, relying solely on graph-based features achieves the highest AUROC but falls short in capturing the full complexity of odor prediction, as it lacks complementary information from molecular fingerprints and token embeddings. By combining graph-based features, molecular fingerprints, and token embeddings, this approach captures a wider range of molecular characteristics. Graph-based features provide structural context, fingerprints capture specific substructural elements, and token embeddings encode sequential relationships from the SMILES representation. This multi-layered feature extraction enables the model to better handle the complexity of odor prediction, enhancing its ability to generalize across diverse molecular structures.

## 4.3 Q2: Analysis of Harmonic Modulated Feature Mapping

This section aims to demonstrate the effectiveness of Harmonic Modulated Feature Mapping and verify the overall superiority of HMFNet. We compare it against mainstream and state-of-the-art methods in this domain. As shown in Table 2, the integration of HMFM into the baseline architectures resulted in substantial improvements. To provide a clearer comparison, we present bar charts of the F1 and AUROC scores, shown in Figure 3 and Figure 4, respectively. Secondly, we

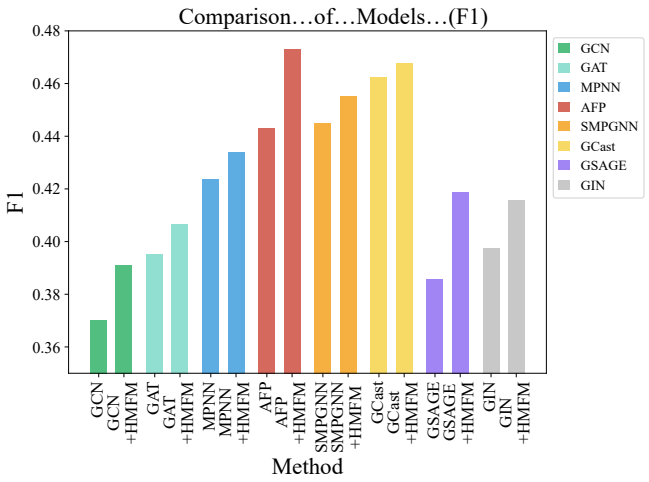


Figure 3: Comparison of F1 scores of histogram of Harmonic Modulated Feature Mapping on mainstream deep learning model.

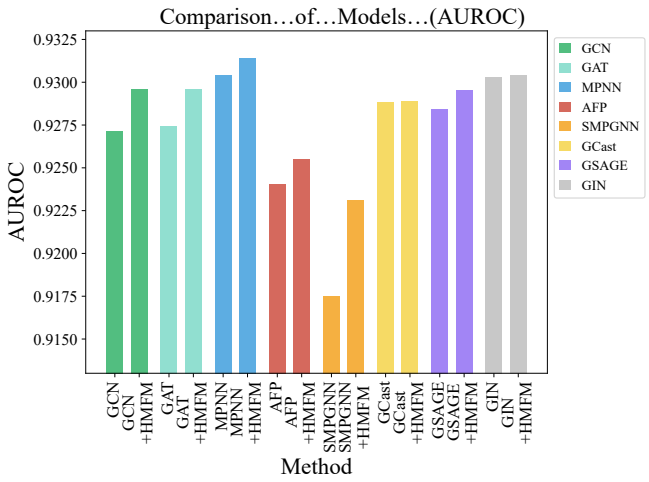


Figure 4: Comparison of histogram AUROC of Harmonic Modulated Feature Mapping on mainstream deep learning model.

compared it with several established feature mapping methods, as shown in Table 3.

Frequency modulation enhances feature representation by introducing high-frequency encoding, which increases the discriminative power for similar molecular structures in the embedding space. This approach enables the model to more precisely capture subtle chemical differences. Cyclic systems and repeating fragments are prevalent in molecular structures, and combining sinusoidal functions with frequency modulation provides a natural way to encode such periodic components. Compared to other encoding or modulation techniques, this method offers inherent advantages in representing complex molecule-odor relationships, improving feature discriminability, and modeling periodic structures.

### 4.4 Q3: Analysis of Chemically-Informed Loss

To evaluate the effectiveness of the proposed Chemically-Informed Loss (CIL), we compared it with present loss function, including Hierarchical Loss (HIL) [Kim *et al.*, 2024], Binary Cross-Entropy Loss (BCE) [Guo *et al.*, 2021], Asymmetric Loss (ASL) [Ridnik *et al.*, 2021], MultiTask Loss (MTL) [Liu *et al.*, 2022b], and Adaptive Focal Loss (AFL) [Xie *et al.*, 2025]. As shown in Table 3 and Table 4, the integration of CIL significantly enhances model performance. Designed to address key challenges in molecular odor prediction—such as class imbalance and label correlation—CIL demonstrates its robustness in improving predictive outcomes. Experimental results show that incorporating CIL consistently increases the F1 score across various base models. By embedding chemical information constraints, CIL not only enhances classification accuracy but also strengthens the model's ability to capture the nuanced relationships between molecular structures and odor descriptors. Furthermore, it achieves a balanced prediction for both majority and minority classes.

### 4.5 Q4: Overall Comparison

As shown in Table 1, The increase in F1 score with the addition of features and modules indicates that the model is improving its performance. Our method demonstrates a significant improvement in F1 score and achieves the best results in molecular odor prediction, as shown in Table 2. However, the highest AUROC (0.9356) was achieved when only node and edge features were used. This suggests that the model performs better at distinguishing between positive and negative classes when fewer features are involved. Based on the calculation principles of F1 score and AUROC (as detailed in Section 4.1), we analyze that AUROC primarily focuses on the model's ability to correctly rank samples. The inclusion of additional features (fingerprints and tokens) may have introduced noise or led to less confident distinctions between classes, which slightly reduced the model's ability to differentiate samples based on predicted probabilities. However, the continuous improvement in F1 score indicates that the model becomes better at correctly identifying positive samples, reflecting a stronger ability to discern odor-related molecular structures. This improvement comes at the cost of a slight reduction in its ability to rank samples effectively, which in turn impacts AUROC. Detailed ablation studies on specific components of HMFNet can be found in Appendix A.2.

## 5 Conclusion

In this paper, we present a novel framework for molecular odor prediction that effectively captures the relationships between molecular structures and their associated odors. By integrating local-to-global feature extraction with dynamic feature mapping and optimizing the loss function. This design enhances the model's ability to identify odor-related molecular patterns. Experimental results confirm the superior performance and robust transferability of our method.

## References

[Alvares-Cherman *et al.*, 2012] Everton Alvares-Cherman, Jean Metz, and Maria Carolina Monard. Incorporating label dependency into the binary relevance framework for multi-label classification. *Expert Systems with Applications*, 39(2):1647–1655, 2012.

[Choi *et al.*, 2023] Hyunjun Choi, Hawook Jeong, and Jin Young Choi. Balanced energy regularization loss for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15691–15700, 2023.

[Durant *et al.*, 2002] Joseph L Durant, Burton A Leland, Douglas R Henry, and James G Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of chemical information and computer sciences*, 42(6):1273–1280, 2002.

[Flavor, 2018] Fragrance Flavor. Food, and cosmetics ingredients information. *The Good Scents Company*, 2018.

[Gilmer *et al.*, 2017] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[Guo *et al.*, 2021] Zhichun Guo, Chuxu Zhang, Wenhao Yu, John Herr, Olaf Wiest, Meng Jiang, and Nitesh V Chawla. Few-shot graph learning for molecular property prediction. In *Proceedings of the web conference 2021*, pages 2559–2567, 2021.

[Huang and Chen, 2024] Kai Huang and Chen Chen. Subgraph generation applied in graphsage deal with imbalanced node classification. *Soft Computing*, pages 1–14, 2024.

[Ji *et al.*, 2023] Huizhuo Ji, Dandan Pu, Wenjing Yan, Qingchuan Zhang, Min Zuo, and Yuyu Zhang. Recent advances and application of machine learning in food flavor prediction and regulation. *Trends in Food Science & Technology*, 138:738–751, 2023.

[Keller *et al.*, 2017] Andreas Keller, Richard C Gerkin, Yuanfang Guan, Amit Dhurandhar, Gabor Turu, Bence Szalai, Joel D Mainland, Yusuke Ihara, Chung Wen Yu, Russ Wolfinger, et al. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355(6327):820–826, 2017.

[Kim *et al.*, 2024] Gibaeg Kim, SangHun Im, and Heung-Seon Oh. Hierarchy-aware biased bound margin loss function for hierarchical text classification. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 7672–7682, 2024.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Lee *et al.*, 2023] Brian K Lee, Emily J Mayhew, Benjamin Sanchez-Lengeling, Jennifer N Wei, Wesley W Qian, Kelsie A Little, Matthew Andres, Britney B Nguyen, Theresa Moloy, Jacob Yasonik, et al. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023.

[Leeney and McConville, 2024] William Leeney and Ryan McConville. An investigation into using unsupervised metrics to optimise gnns for node clustering. *arXiv preprint arXiv:2402.07845*, 2024.

[Leffingwell, 2005] John C Leffingwell. Leffingwell & associates. *Chirality & Odour Perception. Available online: http://www. leffingwell. com/chirality/chirality. htm (accessed on 15 November 2023)*, 2005.

[Liu *et al.*, 2022a] Qi Liu, Dehan Luo, Tengteng Wen, Hamid GholamHosseini, Xiaofang Qiu, and Jingshan Li. Poi-3dgcn: Predicting odor intensity of monomer flavors based on three-dimensionally embedded graph convolutional network. *Expert Systems with Applications*, 199:116997, 2022.

[Liu *et al.*, 2022b] Shengchao Liu, Meng Qu, Zuobai Zhang, Huiyu Cai, and Jian Tang. Structured multi-task learning for molecular property prediction. In *International conference on artificial intelligence and statistics*, pages 8906–8920. PMLR, 2022.

[Mao *et al.*, 2023] Anqi Mao, Mehryar Mohri, and Yutao Zhong. Cross-entropy loss functions: Theoretical analysis and applications. In *International conference on Machine learning*, pages 23803–23828. PMLR, 2023.

[Mitra and Kaddoum, 2022] Rangeet Mitra and Georges Kaddoum. Random fourier feature-based deep learning for wireless communications. *IEEE Transactions on Cognitive Communications and Networking*, 8(2):468–479, 2022.

[Moriwaki *et al.*, 2018] Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, 10:1–14, 2018.

[Nilakantan *et al.*, 1987] Ramaswamy Nilakantan, Norman Bauman, J Scott Dixon, and R Venkataraghavan. Topological torsion: a new molecular descriptor for sar applications. comparison with other descriptors. *Journal of Chemical Information and Computer Sciences*, 27(2):82–85, 1987.

[Ridnik *et al.*, 2021] Tal Ridnik, Emanuel Ben-Baruch, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter, and Lihi Zelnik-Manor. Asymmetric loss for multi-label classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 82–91, 2021.

[Rogers and Hahn, 2010] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

[Saini and Ramanathan, 2022] Kushagra Saini and Venkatnarayan Ramanathan. Predicting odor from molecular structure: A multi-label classification approach. *Scientific reports*, 12(1):13863, 2022.

[Schicker *et al.*, 2023] Doris Schicker, Satnam Singh, Jessica Freiherr, and Andreas T Grasskamp. Owsum: algo-

rithmic odor prediction and insight into structure-odor relationships. *Journal of Cheminformatics*, 15(1):51, 2023.

[Sharma *et al.*, 2021] Anju Sharma, Rajnish Kumar, Shabnam Ranjta, and Pritish Kumar Varadwaj. Smiles to smell: decoding the structure–odor relationship of chemical compounds using the deep neural network approach. *Journal of Chemical Information and Modeling*, 61(2):676–688, 2021.

[Vaswani, 2017] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[Wacker and Filippone, 2022] Jonas Wacker and Maurizio Filippone. Local random feature approximations of the gaussian kernel. *Procedia Computer Science*, 207:987–996, 2022.

[Wang *et al.*, 2024a] Guishen Wang, Hui Feng, Mengyan Du, Yuncong Feng, and Chen Cao. Multimodal representation learning via graph isomorphism network for toxicity multitask learning. *Journal of Chemical Information and Modeling*, 64(21):8322–8338, 2024.

[Wang *et al.*, 2024b] Shixiong Wang, Wei Dai, and Geoffrey Ye Li. Distributionally robust beamforming and estimation of wireless signals. *arXiv preprint arXiv:2401.12345*, 2024.

[Wong *et al.*, 2024] Felix Wong, Satotaka Omori, Alicia Li, Aarti Krishnan, Ryan S Lach, Joseph Rufo, Maxwell Z Wilson, and James J Collins. An explainable deep learning platform for molecular discovery. *Nature Protocols*, pages 1–37, 2024.

[Wu *et al.*, 2020] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[Xie *et al.*, 2025] HongXin Xie, JianDe Sun, Yi Shao, Shuai Li, Sujuan Hou, YuLong Sun, and Jian Wang. Molecular odor prediction based on multi-feature graph attention networks. *arXiv preprint arXiv:2502.01430*, 2025.

[Xiong *et al.*, 2019] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16):8749–8760, 2019.

[Yamada, 2024] Hiroshi Yamada. Spatial smoothing using graph laplacian penalized filter. *Spatial Statistics*, page 100799, 2024.

[Yap, 2011] Chun Wei Yap. Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7):1466–1474, 2011.

[Yuan *et al.*, 2023] Qitong Yuan, Keyi Chen, Yimin Yu, Nguyen Quoc Khanh Le, and Matthew Chin Heng Chua. Prediction of anticancer peptides based on an ensemble model of deep learning and machine learning using ordinal positional encoding. *Briefings in Bioinformatics*, 24(1):bbac630, 2023.

[Zhang *et al.*, 2018] Lei Zhang, Haitao Mao, Linlin Liu, Jian Du, and Rafiqul Gani. A machine learning based computer-aided molecular design/screening methodology for fragrance molecules. *Computers & Chemical Engineering*, 115:295–308, 2018.