

DmcFlavor: Compound-informed prediction and interpretation of integrated multi-flavor profiles in foods and herbs

Tengteng Xu^{b,1}, Ke Zhao^{a,1}, Haoping Li^{a,1}, Na Lin^c, Yanqiong Zhang^c, Zhaochen Ma^c,
Weijie Li^c, Wanyi Guo^a, Lihan Wang^a, Jie Chen^d, Liang Liu^{a,*}, Hudan Pan^{a,*}, Yudong Liu^{a,*}

^a Chinese Medicine Guangdong Laboratory, Hengqin 519031, Guangdong, China

^b Shenzhen Medical Academy of Research and Translation (SMART), Shenzhen 518036, Guangdong, China

^c State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

^d National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, Guangdong, China

ARTICLE INFO

Keywords:

Multi-flavor prediction
Integrated flavor
Food therapy
Food-medicine homologous herbs
Deep learning

ABSTRACT

Food flavor is inherently multidimensional and arises from the complex combined effects of diverse compounds; however, dedicated computational tools for predicting integrated multi-flavor profiles of foods are still lacking. In this study, DmcFlavor was developed as the first framework for predicting and interpreting integrated multi-flavor profiles of foods. DmcFlavor employs Transformer-based modeling to perform both compound-level and mixture-level flavor prediction across five flavor attributes: bitter, pungent, salty, sour, and sweet. DmcFlavor was applied to tomato ripening analysis, revealing dynamic changes in integrated flavor profiles during ripening and quantitatively identifying major contributors to sourness and sweetness. It was further applied to characterize integrated flavor patterns across 18 food categories, including fishes, fruits, herbs and spices, and vegetables, and to explore potential flavor-health effect associations. This framework may facilitate integrated food multi-flavor evaluation, support flavor optimization in the food industry, and promote the systematic exploration of food therapy.

1. Introduction

As living standards improve, consumer expectations for food quality have increasingly shifted from basic nutrition to pleasurable and distinctive sensory experiences. Flavor has therefore become a key criterion for evaluating food quality, shaping consumer preference, and differentiating food products (K. Xu et al., 2025). Accordingly, accurate identification and prediction of flavor-related compounds are important for food research, functional food development, and rational product design. However, food flavor is rarely determined by a single molecule; instead, it arises from the combined contributions of multiple compounds with diverse structures, concentrations, and sensory properties (B. Chen et al., 2025). Therefore, accurate evaluation of integrated multi-flavor profiles is essential for understanding food quality and sensory characteristics.

Recent advances in artificial intelligence have provided a new

paradigm for structure-based flavor prediction by learning nonlinear relationships between molecular structure and sensory attributes (Wei et al., 2026). Existing models mainly include binary predictors for specific tastes, such as bitter-focused BERT4Bitter (Charoenkwan, Nantasenamat, Hasan, Manavalan, & Shoombuatong, 2021) and iBitter-Fuse (Charoenkwan, Nantasenamat, Hasan, Moni, et al., 2021), and sweet-focused ChemSweet (Qian et al., 2025) and BoostSweet (Lee et al., 2022). Multi-flavor models, including FART (Zimmermann et al., 2025) and VirtuousMultiTaste (Androutsos et al., 2024), have further extended taste prediction beyond binary classification. These studies demonstrate the ability of deep learning to capture structure-taste relationships and reduce reliance on sensory evaluation. However, most models remain limited to a single compound and do not directly address real food systems, in which multiple compounds jointly determine integrated flavor profiles. In addition, quantitative constituent compound information and sensory activation thresholds are often

* Corresponding authors.

E-mail addresses: lliu@gzucm.edu.cn (L. Liu), hdpan@gzucm.edu.cn (H. Pan), kabutoun1@163.com (Y. Liu).

¹ These authors offer an equal contribution to this study.

unavailable or difficult to obtain across foods, making integrated multi-flavor modeling highly difficult.

Food–medicine homologous herbs are edible plants with established traditional medicinal value. With increasing interest in dietary regulation and food therapy for health maintenance, these herbs provide a valuable link between food composition, sensory properties, and biological function (Jiang et al., 2025). Unlike common foods, which lack unified and standardized flavor annotations, medicinal–food homologous herbs have well-documented flavor attributes in traditional pharmacopeial systems, offering structured references for learning mixture-level flavor characteristics. In these herbs, flavor attributes are not merely sensory descriptors but are often associated with health effects; for example, bitter, pungent, salty, sour, and sweet attributes may reflect differences in chemical composition and potential bioactivity. Recent food knowledge resources, such as FoodAtlas, have begun to investigate associations among food composition, chemical bioactivity, and health-related information (F. Li et al., 2026). Integrating multi-flavor prediction with health-effect information may therefore help interpret the functional properties of food-derived and medicinal–food homologous herbs.

In this study, we developed DmcFlavor, a compound-informed computational framework for predicting and interpreting integrated flavor profiles of foods and herbs. DmcFlavor uses Transformer-based modeling to perform flavor prediction at both compound and mixture levels. At the compound level, multiple molecular representations were evaluated, and the Transformer-based model was selected to predict five flavor attributes: bitter, pungent, salty, sour, and sweet. The resulting compound-level flavor features were then used for mixture-level modeling. At the mixture level, each food or herb was represented as an unordered set of constituent compounds, and Set Transformer was used to learn integrated flavor profiles from compound composition. This design enables DmcFlavor to predict the overall flavor characteristics of complex foods and herbs while capturing the contributions of individual compounds. DmcFlavor was subsequently applied to tomato ripening analysis, large-scale food flavor profiling, and flavor–health effect association analysis of compounds and medicinal–food homologous herbs. This framework provides a computational strategy for integrated food multi-flavor evaluation and the systematic exploration of food therapy.

2. Materials and methods

2.1. Data collection and preprocessing

The flavor data of single compounds were collected from multiple public databases, including FooDB (www.foodb.ca), BitterDB (Ziaikin et al., 2025), VirtualTaste (Fritz et al., 2021), ChemTastesDB (Rojas et al., 2022), PlantMolecularTasteDB (Gradinaru et al., 2021), PungentDB (Z. Chen et al., 2024), and PubChem (Kim et al., 2025). To construct a reliable five-class flavor dataset, only compounds with clear single-flavor annotations corresponding to bitter, pungent, salty, sour, or sweet were retained. Compounds with multiple or conflicting flavor annotations were excluded. When multiple annotations were available, records with strong sensory descriptors, such as “very,” “extensive,” or “extremely,” were preferentially retained. All compounds were standardized using PubChem IDs and represented by canonical SMILES. The final compound-level dataset comprised 3186 compounds, including 1054 bitter, 333 pungent, 614 salty, 463 sour, and 722 sweet compounds (Table S1).

For mixture-level modeling, herb flavor annotations were obtained from the *Chinese Pharmacopeia 2025 edition* (CP2025), and constituent compound information was collected from BATMAN-TCM 2.0 (Kong et al., 2024), and HIT 2.0 (Yan et al., 2022). Of the 616 herbs documented in CP2025, 558 with available constituent compound information were retained for model development, comprising 69,553 herb–compound records. Because individual herbs may have multiple

documented flavor attributes, these annotations were retained as multi-label records. Duplicate compound records from different sources were merged, and entries with identical compound compositions but different names were removed. For external validation, an independent dataset of 211 herbs comprising 11,735 herb–compound records was constructed using constituent compound information from BATMAN-TCM 2.0 and flavor annotations from ETCM 2.0 (Y. Zhang et al., 2023). In addition, an application dataset comprising 826 foods and 64,720 constituent compounds was obtained from FooDB for large-scale food flavor profiling (Table S2).

After preprocessing, both compound-level and mixture-level datasets were independently divided into training, validation, and test sets at a ratio of 8:1:1. The training sets were used for model fitting, the validation sets for hyperparameter optimization and model selection, and the test sets for internal performance evaluation. The independent external validation dataset was used exclusively to evaluate model generalization. The FooDB food dataset was used for downstream application of the final mixture-level model.

2.2. Model architecture

2.2.1. Compound-level flavor prediction

DmcFlavor first characterized compound-level flavor profiles across five categories: bitter, pungent, salty, sour, and sweet. To capture flavor-related structural information, canonical SMILES were used to generate three complementary molecular representations: ECFP4 fingerprints, molecular graphs, and SMILES sequences. Based on these representations, single-modality prediction models and multimodal fusion models were constructed and systematically evaluated.

For fingerprint-based modeling, ECFP4 fingerprints with different bit lengths were calculated using RDKit and used as inputs for random forest (RF), CatBoost, and fully connected neural network (FCNN) classifiers. For graph-based modeling, each molecule was represented as a molecular graph, with atoms encoded as nodes and chemical bonds as edges, and a GATv2-based graph neural network was used to extract topology-dependent molecular features. For sequence-based modeling, canonical SMILES were encoded using a ChemBERTa-based Transformer model to capture contextual chemical information. In addition to these single-modality models, multimodal fusion models were constructed by concatenating features derived from ECFP4 fingerprints, molecular graphs, and SMILES sequences, followed by fully connected layers for classification. All models were independently trained and evaluated over 10 repeated runs, and the best-performing model was selected to generate compound-level flavor features. The final output was a five-dimensional probability vector representing the predicted likelihood of bitter, pungent, salty, sour, and sweet attributes for each compound.

Model parameters were optimized using the AdamW optimizer. Predictive performance was evaluated using accuracy (ACC), Matthews correlation coefficient (MCC), receiver operating characteristic area under the curve (ROC-AUC), macro-averaged precision, macro-averaged recall, and macro-averaged F1 score.

2.2.2. Mixture-level flavor prediction

For foods and herbs, flavor perception arises from complex combinations of multiple compounds rather than from individual compounds alone. However, quantitative composition information and sensory activation thresholds are often unavailable or difficult to obtain across diverse foods and herbs. To address this challenge, DmcFlavor developed two set-based learning strategies for mixture-level flavor prediction: attention-based multiple instance learning (Attention MIL) and Set Transformer. In both strategies, each food or herb was represented as an unordered set of constituent compounds, and each compound was encoded using a Transformer-derived embedding, a five-flavor probability vector, or their fused representation.

Attention MIL was used to estimate the relative contributions of in-

dividual compounds to the overall flavor profile. In this framework, each food or herb was treated as a bag of constituent compounds, and the attention mechanism dynamically assigned an importance weight to each compound based on its learned feature representation. The normalized attention weight of the i -th compound was calculated as:

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^N \exp(e_j)}$$

where a_i denotes the attention weight assigned to the compound i , e_i represents its attention score, and N is the total number of compounds in the mixture. The mixture-level representation was then obtained by attention-weighted aggregation:

$$z = \sum_{i=1}^N a_i h_i$$

where h_i represents the feature vector of the i -th compound, and z denotes the aggregated mixture-level representation used for flavor prediction.

To capture compositional flavor patterns among constituent compounds, a Set Transformer architecture was further employed. Unlike Attention MIL, which mainly estimates compound-level contribution weights, Set Transformer uses self-attention to update each compound representation based on its relationships with other compounds in the same mixture. This allows the model to learn set-level compositional patterns in the compound embedding space. The self-attention operation was defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

where Q , K , and V denote the query, key, and value matrices, respectively, and d represents the feature dimension. After self-attention-based feature interaction and pooling, the resulting set-level representation was used to predict a five-dimensional probability distribution corresponding to bitter, pungent, salty, sour, and sweet attributes.

Mixture-level models were trained using the AdamW optimizer with binary cross-entropy loss with logits. Model performance was evaluated using Top-1, Top-2, and Top-3 accuracy, mean reciprocal rank (MRR), and mean rank.

2.3. Food multi-flavor profiling

To evaluate the applicability of DmcFlavor in food flavor analysis, tomato ripening was used as a representative case study. Tomato flavor changes during ripening are driven by dynamic metabolite variation. Therefore, tomato fruits collected at 0, 3, 6, and 9 days after harvest, corresponding to the mature green, breaker, pink, and red-ripe stages, respectively, were analyzed. DmcFlavor was applied to predict dynamic changes in the five-flavor profiles of tomato metabolites across ripening stages and to quantitatively identify the major contributors to sourness and sweetness.

DmcFlavor was further applied to large-scale food flavor profiling using FooDB constituent compound information. In total, 64,720 FooDB compounds were predicted using the compound-level flavor prediction module, and the mixture-level flavor probabilities were used to infer the integrated flavor profiles of 826 foods. These foods were grouped into 18 categories, including fishes, baking goods, beverages, cereals and cereal products, cocoa and cocoa products, coffee and coffee products, confectioneries, eggs, fats and oils, fruits, gourds, herbs and spices, milk and milk products, nuts, pulses, soy, teas, and vegetables. This analysis enabled systematic comparison of predicted multi-flavor characteristics across major food categories.

2.4. Flavor-health effect association analysis

To explore potential associations between predicted flavor characteristics and health-related effects, DmcFlavor was applied at both the compound and mixture levels. At the compound level, flavor profiles were predicted for 12,042 DrugBank compounds (Knox et al., 2024) and compared across eight pharmacological categories: analgesics, anti-asthmatic agents, anti-inflammatory agents, antipyretics, antirheumatic agents, bronchodilator agents, dietary supplements, and laxatives. Potential physicochemical and structural confounding was further assessed by comparing major physicochemical properties between DrugBank and FooDB compounds, performing physicochemical-adjusted regression analyses of pharmacological category-bitterness associations, and analyzing generic Bemis-Murcko scaffold composition. Detailed procedures for these analyses are provided in the Supplementary Material, Section 1.

At the mixture level, ginger and ginkgo seed were selected as representative food-medicine homologous herbs. DmcFlavor was used to infer their integrated flavor profiles from constituent compounds, and compound-level flavor predictions were mapped to reported health-effect annotations from FooDB to explore potential associations between flavor characteristics and reported health effects.

3. Results and discussion

3.1. Compound-level flavor prediction by DmcFlavor

With increasing consumer demand for high-quality food flavor, accurate prediction and efficient identification of flavor-related compounds have become important drivers of innovation in food research and product development (Wang et al., 2026). However, the integrated flavor of foods and herbs generally arises from the combined contributions of multiple constituents with diverse chemical structures and sensory properties. DmcFlavor was therefore designed to extend flavor prediction from individual compounds to complex food and herb mixtures based on their constituent compound composition.

As the first step of this framework, a compound-level flavor prediction module was established to characterize flavor-related molecular features and provide transferable representations for downstream mixture-level modeling. As shown in Fig. 1, compounds with clear single-flavor annotations were curated from multiple public databases and classified into five categories: bitter, pungent, salty, sour, and sweet. The resulting model-development dataset comprised 3186 labeled compounds, including 1054 bitter, 333 pungent, 614 salty, 463 sour, and 722 sweet compounds. Canonical SMILES were used as standardized structural representations to generate fingerprint-, graph-, and sequence-based molecular features. The compound-level models predicted a five-dimensional probability vector for each compound, representing its probabilities of bitter, pungent, salty, sour, and sweet attributes. These compound-level representations were subsequently transferred to mixture-level modeling of foods and herbs.

Fingerprint-based flavor prediction was first evaluated using ECFP4 fingerprints with different bit lengths in combination with three classifiers: RF, CatBoost, and FCNN. As shown in Fig. S1, predictive performance varied according to both fingerprint length and classifier architecture. Among the evaluated models, CatBoost showed the most consistent overall performance across ACC, MCC, ROC-AUC, macro-averaged precision, macro-averaged recall, and macro-averaged F1 score, with the 768-bit ECFP4 fingerprint achieving the best overall performance. These findings indicate that ECFP4 fingerprints effectively capture substructural information associated with flavor characteristics, while their predictive performance is influenced by feature dimensionality and the learning algorithm.

Different molecular representation strategies were subsequently compared, including ECFP4 fingerprints, molecular graphs, SMILES sequences, and multimodal fusion features. Among the single-modality

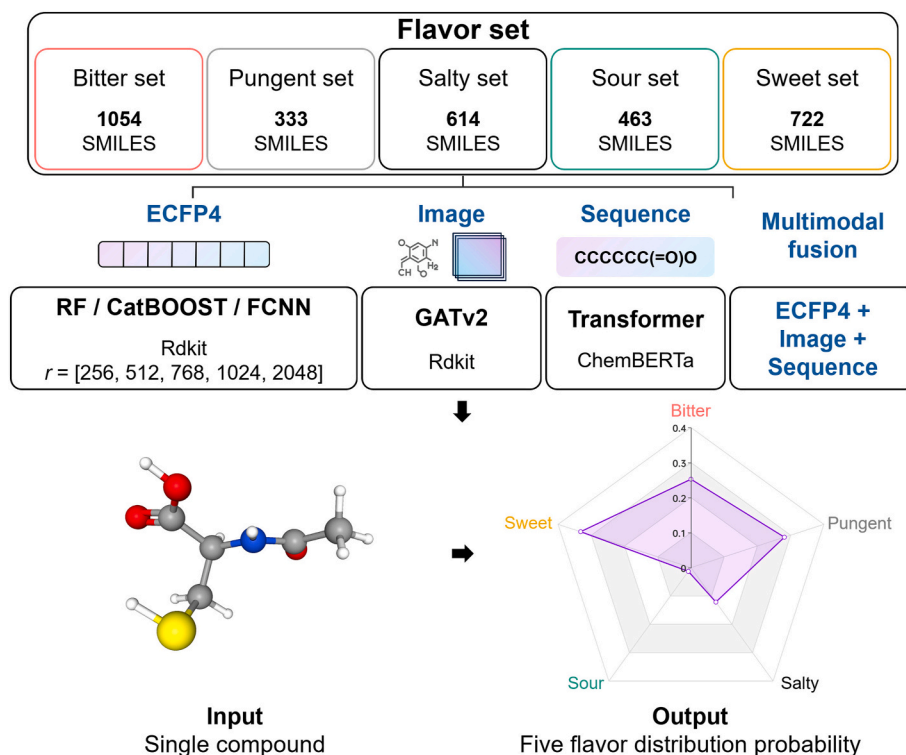


Fig. 1. Framework for constructing the compound-level flavor prediction module of DmcFlavor.

models, the Transformer-based model achieved the best overall performance (Fig. 2A), indicating the effectiveness of contextual representations derived from SMILES for compound-level flavor prediction. Transformer architectures can capture contextual dependencies among molecular tokens and thereby encode structural patterns that may be relevant to flavor characteristics. Multimodal fusion models were also evaluated but did not consistently outperform the Transformer-based model (Fig. 2B), suggesting that direct integration of heterogeneous molecular representations does not necessarily provide additional predictive information and may introduce redundant features. Based on its overall performance, the Transformer-based model was selected as the final compound-level model and feature generator for subsequent analyses.

After model development and evaluation, the compound-level prediction module was applied to 64,720 constituent compounds collected from 826 foods in FooDB to characterize their individual flavor attributes before mixture-level integration. The predicted distribution showed that bitter was the most prevalent flavor category among these compounds, followed by salty, sweet, pungent, and sour (Fig. 2C, Table S3).

3.2. Mixture-level flavor prediction by DmcFlavor

After establishing the compound-level prediction module, DmcFlavor was extended to mixture-level flavor prediction. A major challenge in food flavor modeling is the lack of unified and standardized

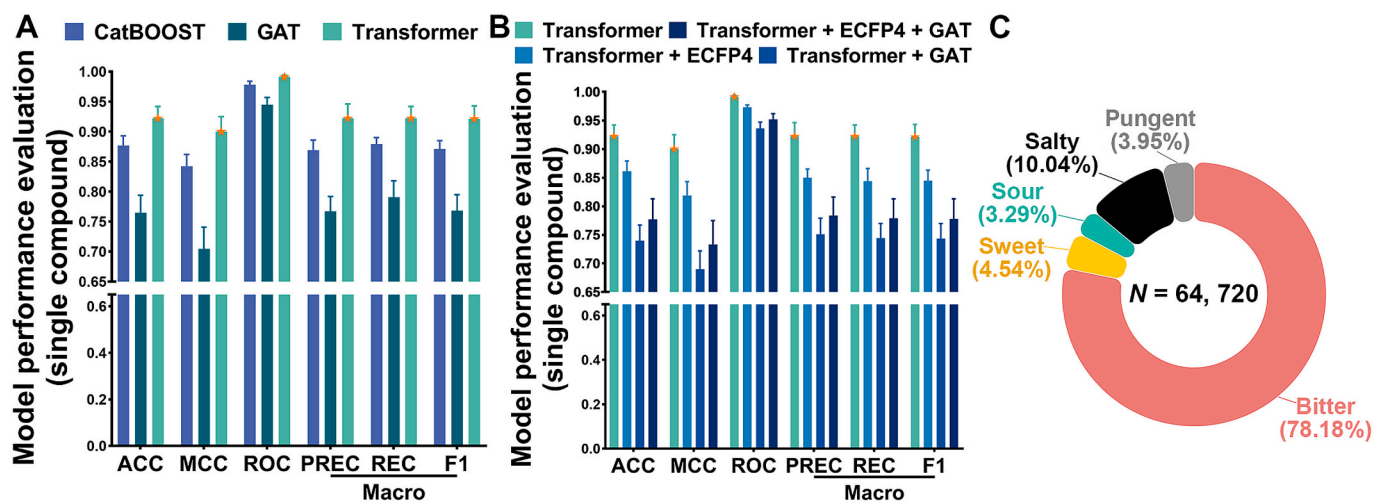


Fig. 2. Performance evaluation of compound-level flavor prediction models. (A) Performance comparison of representative single-modality models. (B) Performance of multimodal fusion models integrating ECFP4 fingerprints, graph-based features, and Transformer-derived sequence features. (C) Predicted flavor distribution of 64,720 FooDB compounds across five flavor categories.

annotations for complex food flavors. In contrast, CP2025, a national authoritative pharmacopeia, provides well-documented flavor annotations for medicinal herbs. Importantly, many medicinal herbs are also traditionally consumed as foods, providing a natural connection between pharmacopeial flavor annotations and food-related applications. The traditional five-flavor system has a sensory basis, with flavor attributes originally identified through direct tasting of medicinal materials. Therefore, the systematically documented flavor annotations in CP2025 provide a valuable reference for mixture-level flavor modeling. Accordingly, a labeled mixture dataset comprising 558 herbs, including 98 food–medicine homologous herbs, and 69,553 herb–compound records was constructed using flavor annotations from CP2025 and constituent compound information from BATMAN-TCM 2.0 and HIT 2.0 (Table S4).

Another challenge is that the relative proportions of constituent compounds and their sensory activation thresholds are often unavailable or difficult to obtain across foods and herbs. Therefore, concentration- or threshold-dependent rules are difficult to apply to large-scale mixture-level flavor prediction. To address this limitation, each food or herb was represented as an unordered set of constituent compounds, allowing DmcFlavor to learn flavor patterns directly from compound composition. As shown in Fig. 3A, each compound was encoded using features derived from the compound-level model, including a 384-dimensional Transformer-derived embedding, a five-dimensional flavor probability vector, or their 389-dimensional fused representation.

Based on these compound-level representations, two set-based learning strategies were evaluated. Attention MIL dynamically assigned importance weights to individual compounds and aggregated them into a mixture-level representation, enabling estimation of compound-specific contributions without predefined concentration information. In parallel, Set Transformer modeled compound–compound relationships through self-attention and learned set-level compositional patterns in the embedding space. These strategies enabled DmcFlavor to predict integrated flavor profiles from complex compound compositions, with the final output represented as a five-dimensional probability distribution for bitter, pungent, salty, sour, and sweet attributes (Fig. 3A).

The performance of mixture-level models was evaluated using both internal and external validation. For external validation, an independent dataset comprising 211 herbs and 11,735 herb–compound records was constructed using constituent compound information from BATMAN-TCM 2.0 and documented flavor annotations from ETCM v2.0. Compared with direct weighted averaging, set-based learning models achieved better performance in internal validation, highlighting the importance of learning compound-specific contributions and compositional patterns for food and herb flavor prediction (Fig. 3B). Among these models, Set Transformer achieved the best overall performance and was therefore selected as the final mixture-level prediction model. Its superior performance suggests that self-attention-based modeling of compound–compound relationships is particularly effective for

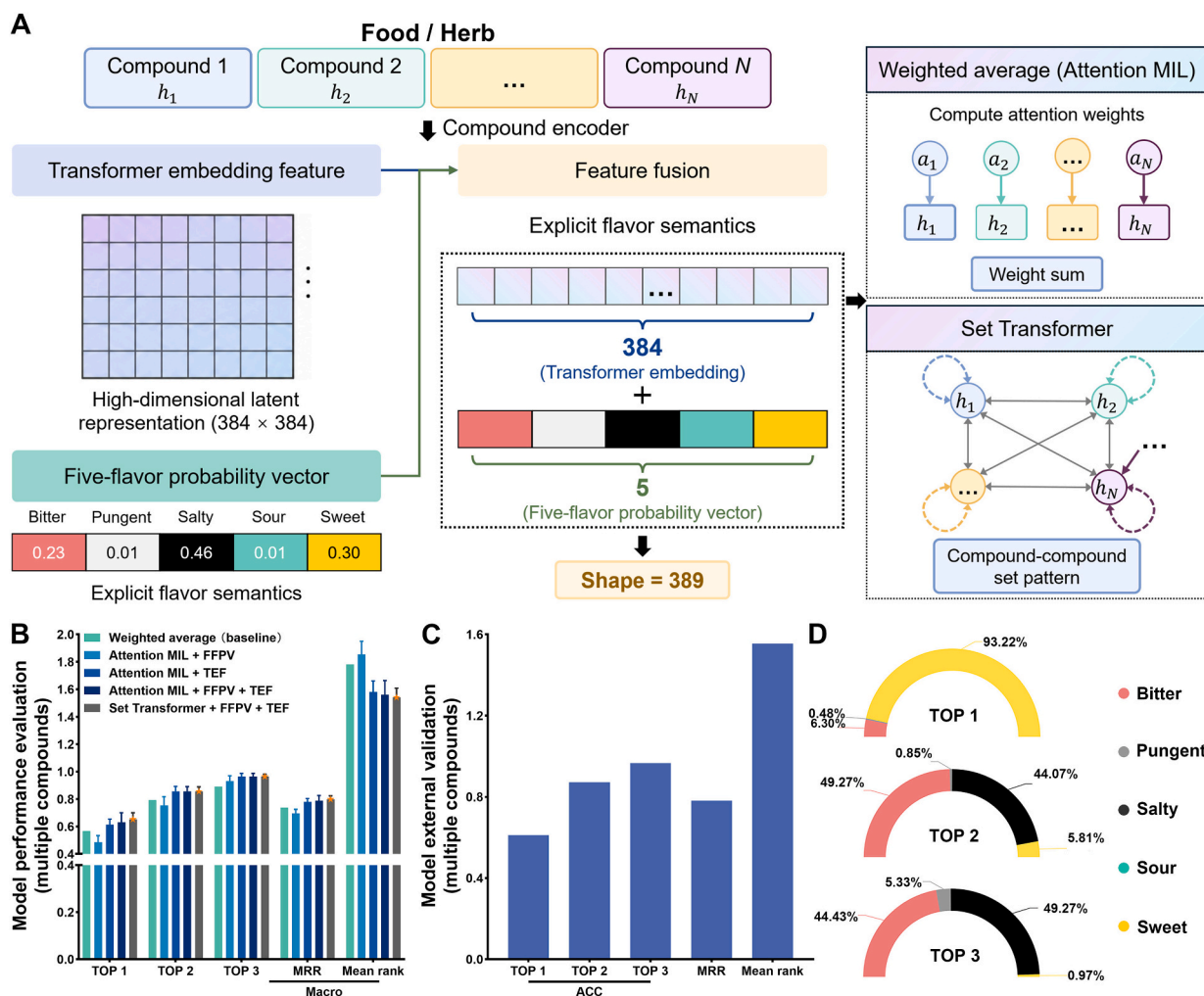


Fig. 3. Framework and performance evaluation of mixture-level flavor prediction in DmcFlavor. (A) Framework of the mixture-level prediction module. (B) Internal validation performance of mixture-level prediction models. (C) External validation performance of mixture-level prediction models. (D) Top-1, Top-2, and Top-3 predicted integrated flavor distributions of 826 FoodDB foods. TEF, Transformer embedding feature; FFPV, five-flavor probability vector.

capturing integrated flavor patterns in complex mixtures. The selected Set Transformer model was further evaluated on the external validation set, where DmcFlavor achieved a Top-1 ACC of approximately 0.62, while Top-2 and Top-3 ACC increased to approximately 0.87 and 0.97, respectively. The corresponding MRR was approximately 0.78, with a mean rank of approximately 1.55, indicating that documented flavor labels were generally ranked among the top predicted categories (Fig. 3C). These results demonstrate the generalization ability of DmcFlavor on independent herb mixtures and support its applicability to complex food and herb systems.

DmcFlavor was further applied to 826 foods comprising 64,720 constituent compounds from FooDB to predict their integrated flavor profiles. As shown in Fig. 3D, the Top-1, Top-2, and Top-3 predicted flavor distributions showed that sweetness was the dominant Top-1 flavor across foods. This result is notable because bitter compounds were the most abundant category at the compound level, whereas integrated food-level flavor profiles were dominated by sweetness. This shift from compound-level bitterness to mixture-level sweetness is consistent with common sensory perception of foods and suggests that DmcFlavor can distinguish isolated compound flavor tendencies from integrated flavor profiles emerging from complex food compositions.

3.3. Dynamic flavor profiling during tomato ripening and large-scale food flavor analysis

To further evaluate the predictive performance and interpretability of DmcFlavor in a real food system, tomato ripening data from Wang et al. (Xiao et al., 2026) were used as a representative case study

(Table S5). Tomato fruits collected at Day 0, Day 3, Day 6, and Day 9 represented progressive ripening stages from mature green to red-ripe (Fig. 4A). During ripening, tomato metabolites showed clear stage-dependent changes across amino acids, flavonoids, organic acids, and saccharides (Fig. 4B). Overall, saccharides increased during ripening, whereas organic acids decreased, reflecting the typical transition from acidic immature fruit to sweeter ripe fruit. Amino acids and flavonoids also exhibited dynamic changes, suggesting broader metabolic remodeling during postharvest ripening.

DmcFlavor predicted distinct flavor tendencies among metabolite classes (Fig. 4C). Amino acids were mainly associated with salty, sweet, and sour attributes; flavonoids were predominantly predicted as bitter; organic acids were mainly associated with sourness; and saccharides showed a strong sweet tendency. These predictions are consistent with known food chemistry characteristics and support the ability of DmcFlavor to assign chemically interpretable flavor attributes to food metabolites (Belloir et al., 2017; L. Li et al., 2023; Roland et al., 2013; Sood et al., 2025; X. Zhang et al., 2022).

The integrated tomato flavor profile changed progressively during ripening. Notably, baseline prediction without quantitative metabolite information produced a consistent overall flavor ranking, indicating that DmcFlavor could infer major flavor tendencies from compound composition alone. After incorporating ripening-stage metabolite abundance, the model captured temporal flavor changes more clearly (Fig. 4D, E). Sourness gradually decreased, sweetness gradually increased, and bitter, pungent, and salty attributes remained relatively stable. This trend is consistent with the sensory transition of tomato fruit from green and acidic to sweeter and more palatable at later ripening

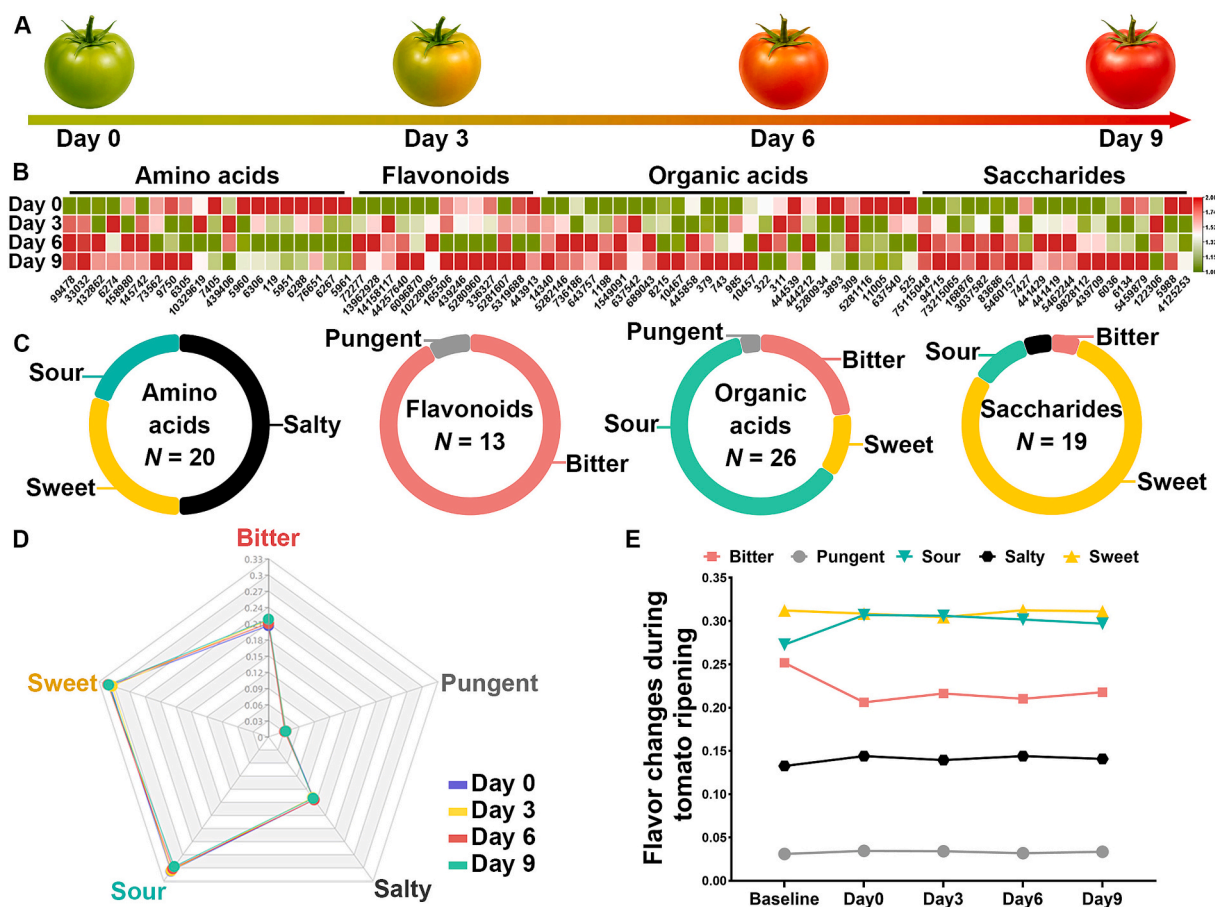


Fig. 4. Dynamic flavor profiling of tomato ripening using DmcFlavor. (A) Schematic representation of tomato ripening stages from Day 0 to Day 9. (B) Dynamic changes in representative tomato metabolites across ripening stages, including amino acids, flavonoids, organic acids, and saccharides. (C) Predicted flavor distributions of different metabolite classes. (D) Predicted integrated flavor profiles of tomato across ripening stages. (E) Dynamic changes in predicted flavor scores during tomato ripening.

stages (Y. Li et al., 2020), indicating that DmcFlavor can reflect biologically meaningful flavor changes when quantitative food composition data are available.

Compound-level contribution analysis further identified major flavor-driving metabolite classes and representative compounds associated with tomato ripening (Fig. S2A–D). Overall, amino acids contributed to multiple taste dimensions, particularly salty and sweet attributes. Organic acids mainly contributed to sourness, supporting their role in the acidic taste of tomato fruit. Saccharides were predominantly associated with sweetness and represented major drivers of the increased sweet profile during ripening, whereas flavonoids contributed mainly to bitterness. At the individual compound level, 1-O-feruloyl- β -D-glucose, L-mannopyranose, and dihydroferulic acid showed the strongest contributions to decreased sourness, while 1-O-feruloyl- β -D-glucose and L-mannopyranose contributed most prominently to increased sweetness. These findings indicate that DmcFlavor can not only predict the overall integrated flavor profile of a food system but also trace flavor variation to specific metabolite classes and candidate flavor-contributing compounds.

DmcFlavor was further applied to large-scale food flavor profiling based on FoodDB constituent compound information. The model inferred five-flavor distributions across 18 food categories, including fishes, baking goods, beverages, cereals and cereal products, cocoa and cocoa products, coffee and coffee products, confectioneries, eggs, fats and oils, fruits, gourds, herbs and spices, milk and milk products, nuts, pulses, soy, teas, and vegetables. The predicted flavor profiles varied markedly among food categories, indicating distinct category-specific flavor patterns. Baking goods were predominantly characterized by sweetness, with a relatively evident salty component (Fig. 5A). Beverages and coffee and coffee products showed pronounced bitter-dominant profiles (Fig. 5B, C), whereas fishes exhibited a distinct pattern characterized by relatively high pungent and sweet probabilities (Fig. 5D). Fruits showed a clear sweet-dominant profile (Fig. 5E), while herbs and spices displayed a mixed profile with relatively high sweet and bitter probabilities (Fig. 5F). Among the remaining food categories, sweetness was dominant in cereals and cereal products, confectioneries, eggs, gourds, milk and milk products, nuts, pulses, soy, teas, and vegetables, whereas cocoa and cocoa products and fats and oils showed relatively higher bitter probabilities (Fig. S3). Overall, sourness was generally less prominent across most food categories (Table S6).

These category-specific patterns are broadly consistent with known food flavor characteristics. For example, the higher bitter tendency in cocoa, coffee, tea, herbs, and spices may be associated with the enrichment of polyphenols, alkaloids, flavonoids, and other secondary

metabolites (Drewnowski, 2001; Shimizu et al., 2024; Soares et al., 2018; Y. Xu et al., 2022), whereas the dominant sweetness observed in many common food categories likely reflects the contribution of carbohydrates and other sweet-associated constituents (An et al., 2025; Sukumaran & Palayyan, 2022). These results indicate that DmcFlavor can capture both shared and category-specific flavor tendencies across structurally diverse foods, providing an interpretable overview of food flavor diversity from compound composition and offering potential support for flavor evaluation, flavor correction, and product optimization in the food industry.

3.4. Flavor–health effect association analysis

Flavor is increasingly recognized not only as a sensory attribute but also as a chemical feature potentially associated with biological functions. Many flavor-related compounds, particularly plant secondary metabolites, possess both sensory properties and biological activities (S. Liu et al., 2024). In traditional Chinese medicine, flavor attributes have historically been associated with functional properties, such as bitter with anti-inflammatory effects and pungent with pain-relieving actions (M. Li et al., 2022; Lin & Lin, 2011). A potential biological basis for such associations is suggested by the extraoral expression of taste receptors and the involvement of sensory pathways in diverse physiological processes (Egan, 2024). For example, bitter taste receptors have been implicated in airway defense and bronchodilation (Conaway, Huang, Hernandez-Lara, Kane, Penn, & Deshpande, 2024), whereas transient receptor potential channels activated by pungent compounds participate in nociceptive and inflammatory signaling (Arora et al., 2021). DmcFlavor therefore provides a computational framework for exploring potential flavor–health effect associations from complex chemical compositions.

To explore associations between predicted flavor characteristics and pharmacological functions, DmcFlavor was applied to 12,042 DrugBank compounds (Table S7). Their predicted flavor distribution was dominated by bitter, followed by salty, sweet, pungent, and sour, showing an overall pattern similar to that of FoodDB compounds (Fig. 6A). However, DrugBank and FoodDB compounds differed significantly in molecular weight, XLogP, polar surface area, hydrogen-bond donor and acceptor counts, rotatable bond count, and molecular complexity (all $P < 0.001$; Table S8). Given these domain-level physicochemical differences, subsequent analyses focused on within-DrugBank comparisons.

Flavor profiles were then compared across eight pharmacological categories: antipyretics, analgesics, anti-inflammatory agents, anti-rheumatic agents, anti-asthmatic agents, bronchodilator agents,

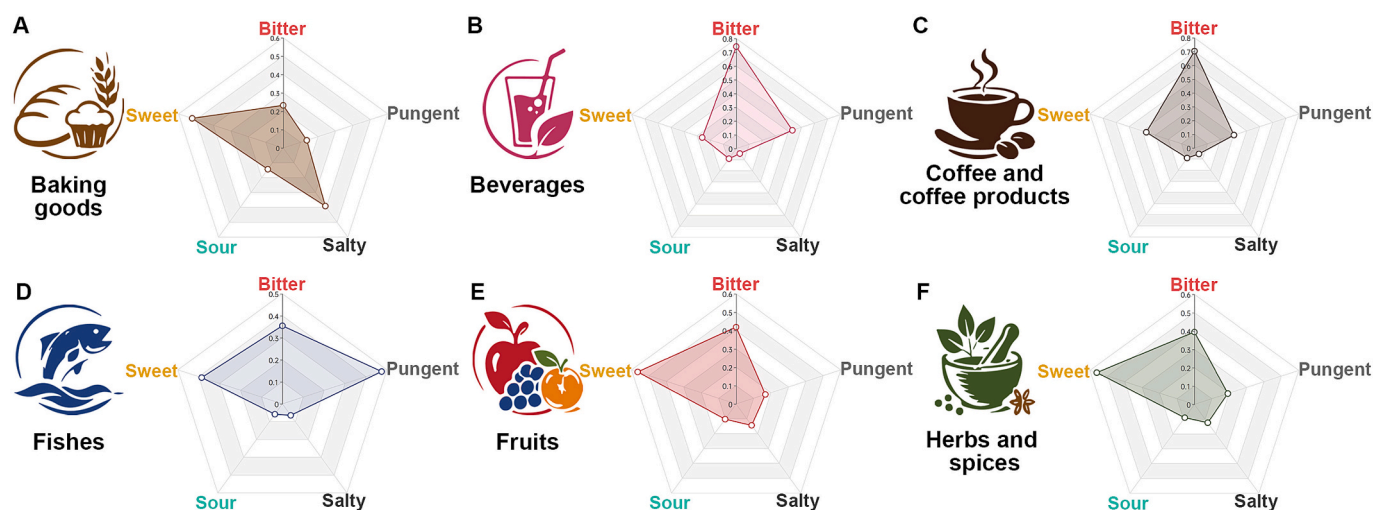


Fig. 5. Representative predicted five-flavor profiles across food categories. (A) baking goods, (B) beverages, (C) coffee and coffee products, (D) fishes, (E) fruits, and (F) herbs and spices.

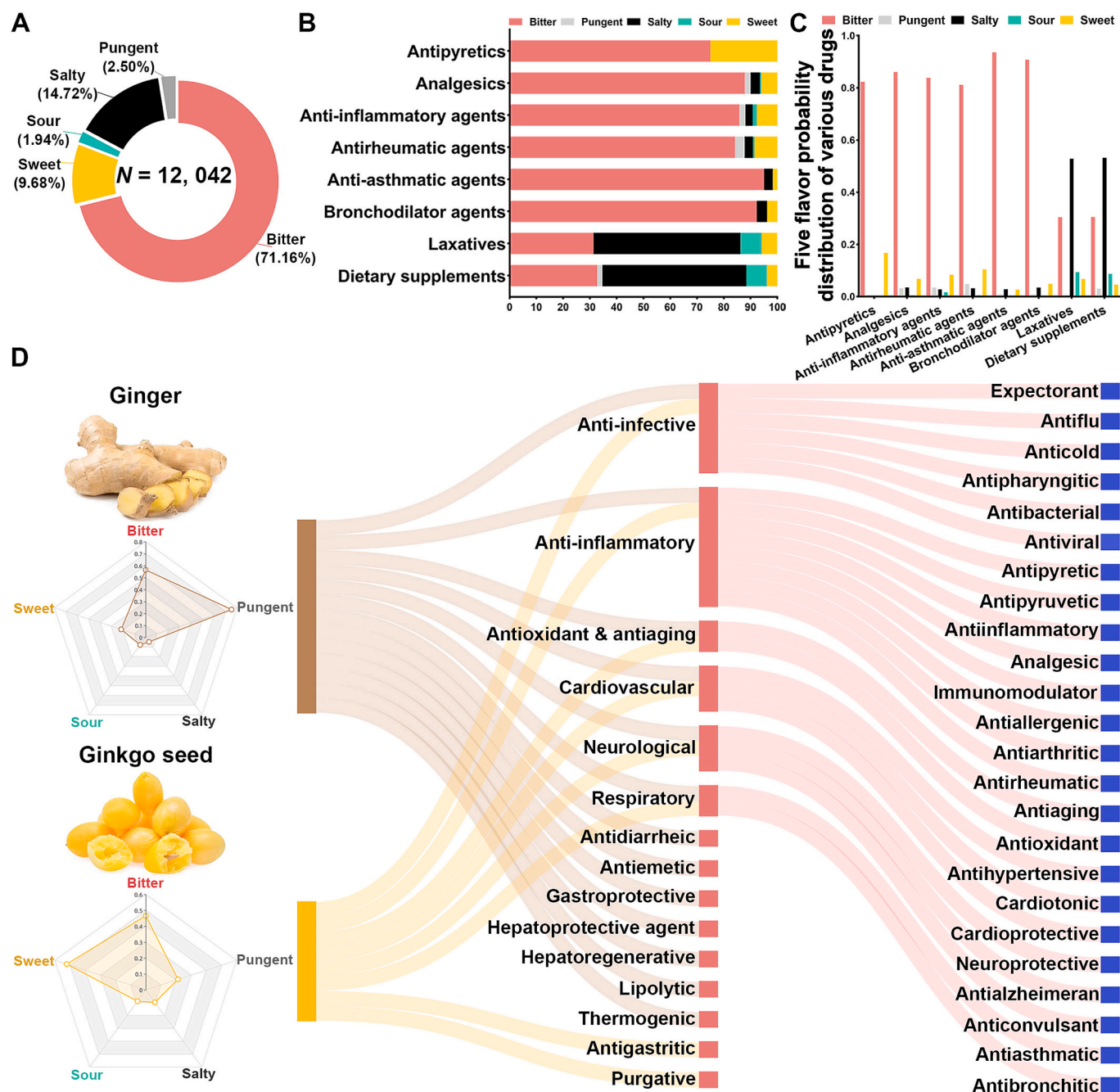


Fig. 6. Flavor-health effect association analysis based on DmcFlavor predictions. (A) Predicted five-flavor distribution of DrugBank compounds. (B) Flavor composition across 8 pharmacological categories. (C) Five-flavor probability distributions across the 8 pharmacological categories. (D) Flavor-health effect association analysis of ginger and ginkgo seed.

laxatives, and dietary supplements. Most categories exhibited bitter-dominant profiles, whereas laxatives and dietary supplements showed higher salty components (Fig. 6B and C). Given the predominance of bitterness, potential physicochemical confounding was evaluated using fractional logistic regression. In single-covariate models, pharmacological category-bitterness associations were generally directionally consistent after separate adjustment for individual physicochemical descriptors (Table S9). In the fully adjusted model incorporating all seven descriptors, anti-asthmatic agents ($P < 0.01$) and bronchodilator agents ($P < 0.05$) were significantly associated with higher predicted bitterness, whereas dietary supplements were associated with lower predicted bitterness ($P < 0.01$) (Fig. S4A). These findings suggest that the observed associations were not fully accounted for by the major

physicochemical characteristics examined. Potential structural contributions were further assessed using generic Bemis-Murcko scaffolds. Bitter and non-bitter compounds contained 210 and 28 unique scaffolds, respectively, with only 11 shared between the two groups (Fig. S4B). Scaffold enrichment analysis further identified distinct scaffolds significantly enriched in specific pharmacological categories (Table S10), indicating category-dependent differences in scaffold composition. Together, these results suggest potential associations between predicted flavor characteristics and pharmacological categories, although further investigation is needed to better understand these relationships.

Finally, ginger and ginkgo seed were selected as representative food-medicine homologous herbs to illustrate potential associations between predicted flavor characteristics and reported health effects.

FoodDB compound–health effect annotations were mapped to the corresponding DmcFlavor predictions (Fig. 6D). Ginger showed predominantly pungent and bitter characteristics, with corresponding compounds associated with reported anti-inflammatory and antiemetic effects (Anh et al., 2020; Shaukat et al., 2023). In contrast, ginkgo seed showed more prominent sweet and bitter characteristics, with corresponding compounds associated with reported gastroprotective, hepatoprotective, and purgative effects (Boateng, 2022; Y. Liu et al., 2022). These examples illustrate potential associations between compound-derived flavor characteristics and reported health effects.

3.5. Comparison with other models of multi-flavor prediction

To further contextualize the compound-level performance of DmcFlavor, we compared it with representative multi-flavor prediction models (Table S11). Because DmcFlavor is the first framework designed for mixture-level integrated multi-flavor prediction of foods and herbs, no existing model provides a directly comparable mixture-level benchmark. Therefore, the comparison was restricted to compound-level multi-flavor classification. FART showed strong performance in compound-level multi-flavor prediction, whereas VirtuousMultiTaste also focused on compound-level multi-flavor classification but could not be reproduced under the same experimental setting. Among the reproducible models, DmcFlavor achieved the best overall compound-level performance and, more importantly, further enabled mixture-level prediction from constituent compound composition. These results indicate that DmcFlavor provides competitive compound-level classification performance while extending flavor prediction to complex food and medicinal-food homologous herbs.

3.6. DmcFlavor web server

DmcFlavor is available as a web-based application at <https://www.tcmaait.com>, supporting both single-compound and mixture-level flavor prediction, result visualization, and data download (Fig. 7). Detailed

usage instructions are provided in the Supplementary Material, Section 2.

4. Conclusion

In this study, we developed DmcFlavor, a compound-informed computational framework for predicting and interpreting integrated flavor profiles of foods and herbs from chemical composition. DmcFlavor performs flavor prediction at both compound and mixture levels. At the compound level, multiple molecular representation strategies were compared, and the Transformer-based model achieved the best performance for predicting five flavor attributes: bitter, pungent, salty, sour, and sweet. At the mixture level, each food or herb was represented as an unordered set of constituent compounds, and Set Transformer was used to learn integrated flavor profiles from compound composition. This design enables DmcFlavor to extend flavor prediction from individual compounds to complex food and herb mixtures while capturing compound-specific contributions and set-level compositional patterns. Applications to tomato ripening, large-scale FoodDB food profiling, and medicinal-food homologous herbs demonstrated that DmcFlavor can reflect biologically meaningful flavor changes, characterize category-specific food flavor patterns, and provide interpretable associations between flavor-related chemical features and potential health functions.

Overall, DmcFlavor provides a scalable and interpretable strategy for integrated food flavor evaluation and the systematic exploration of food therapy. By identifying compounds associated with specific flavor attributes and their contributions to overall flavor changes, DmcFlavor may also support targeted flavor modulation during food processing and the screening of natural food additives or flavor-modulating compounds for improving undesirable sensory characteristics. Nevertheless, several limitations remain. In particular, the relative proportions of constituent compounds and their sensory activation thresholds were not fully incorporated, as such information is rarely available, difficult to standardize, and highly variable across diverse foods and herbs. A key challenge for further development is therefore to determine the

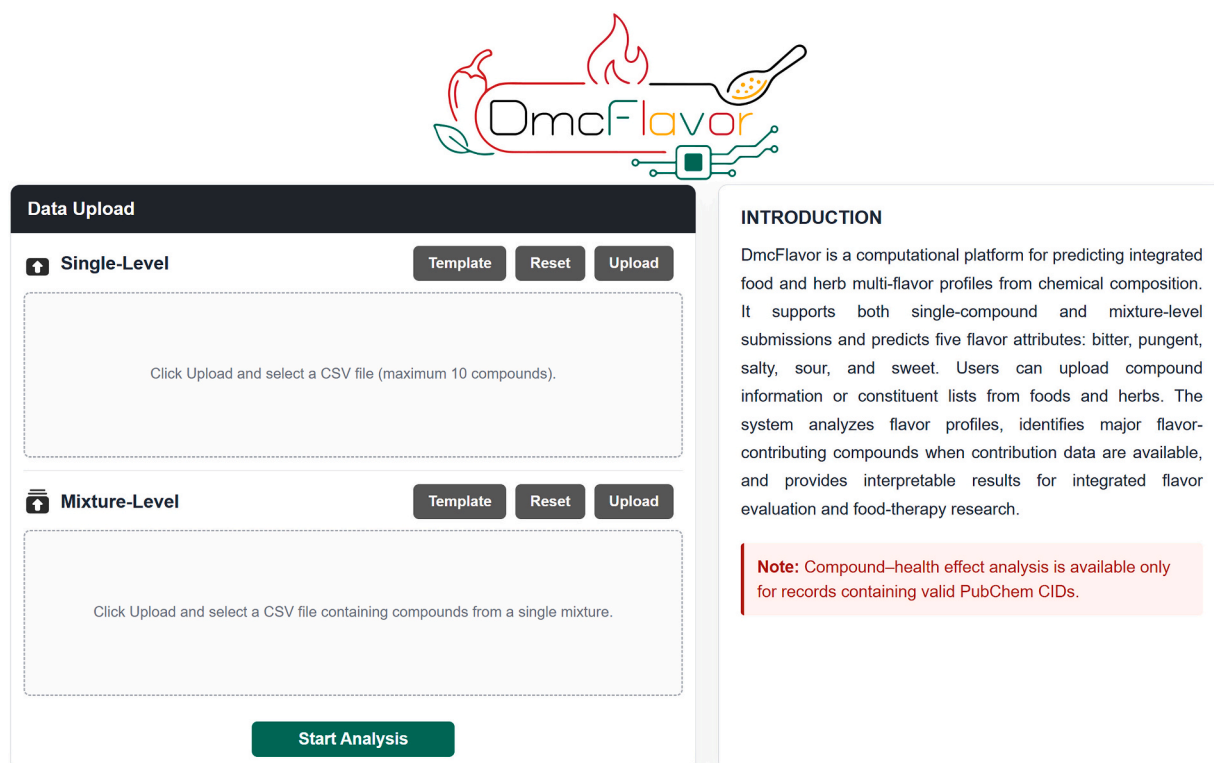


Fig. 7. Web interface of the DmcFlavor platform.

concentration levels at which individual compounds substantially influence desirable or undesirable flavor attributes. Future validation using standardized sensory-panel ratings or other experimentally measured sensory profiles would provide important independent evidence for evaluating the predicted integrated flavor profiles of DmcFlavor. Further integration of such sensory data with quantitative composition data, sensory thresholds, and food processing records could enable more accurate, concentration-dependent assessment of individual compound contributions and improve mixture-level flavor prediction.

The current framework is also limited to 5 sensory attributes and does not encompass the full spectrum of food flavor, particularly aroma-related properties. This scope largely reflects the availability and annotation systems of the reference datasets used for model development. Future extensions incorporating additional sensory dimensions, such as umami and kokumi, together with aroma-related properties, could provide a more comprehensive characterization of food sensory profiles and broaden the applicability of DmcFlavor to food design, personalized nutrition, and food therapy research.

CRediT authorship contribution statement

Tengteng Xu: Writing – original draft, Visualization, Data curation. **Ke Zhao:** Software. **Haoping Li:** Data curation. **Na Lin:** Project administration. **Yanqiong Zhang:** Project administration. **Zhaochen Ma:** Methodology. **Weijie Li:** Methodology. **Wanyi Guo:** Methodology. **Lihan Wang:** Resources. **Jie Chen:** Resources. **Liang Liu:** Writing – review & editing, Project administration. **Hudan Pan:** Writing – review & editing, Project administration. **Yudong Liu:** Writing – review & editing, Software, Methodology, Funding acquisition, Conceptualization.

Funding

This work was supported by the Guangdong Basic and Applied Basic Research Foundation, China (2026A1515011890), the Supported the scientific research start-up funds of Chinese Medicine Guangdong Laboratory (HQL2025SU039), the Traditional Chinese Medicine Research Project of Guangdong Provincial Bureau of Traditional Chinese Medicine, China (20264127, 20264133). Scientific and Technological Innovation Project of China Academy of Chinese Medical Sciences (C12026A03801).

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2026.151386>.

Data availability

Data will be made available on request.

References

- An, J. P., Wang, Y., Munger, S. D., & Tang, X. (2025). A review on natural sweeteners, sweet taste modulators and bitter masking compounds: Structure-activity strategies for the discovery of novel taste molecules. *Critical Reviews in Food Science and Nutrition*, 65(11), 2076–2099. <https://doi.org/10.1080/10408398.2024.2326012>
- Androustos, L., Pallante, L., Bompotas, A., Stojceski, F., Grasso, G., Piga, D., & Mavroudi, S. (2024). Predicting multiple taste sensations with a multiobjective

- machine learning method. *npj Science of Food*, 8(1), Article 47. <https://doi.org/10.1038/s41538-024-00287-6>
- Anh, N. H., Kim, S. J., Long, N. P., Min, J. E., Yoon, Y. C., Lee, E. G., & Kwon, S. W. (2020). Ginger on human health: A comprehensive systematic review of 109 randomized controlled trials. *Nutrients*, 12(1). <https://doi.org/10.3390/nu12010157>
- Arora, V., Campbell, J. N., & Chung, M. K. (2021). Fight fire with fire: Neurobiology of capsaicin-induced analgesia for chronic pain. *Pharmacology & Therapeutics*, 220, Article 107743. <https://doi.org/10.1016/j.pharmthera.2020.107743>
- Belloir, C., Neiers, F., & Briand, L. (2017). Sweeteners and sweetness enhancers. *Current Opinion in Clinical Nutrition and Metabolic Care*, 20(4), 279–285. <https://doi.org/10.1097/MCO.0000000000000377>
- Boateng, I. D. (2022). A critical review of current technologies used to reduce ginkgotoxin, ginkgotoxin-5'-glucoside, ginkgolic acid, allergic glycoprotein, and cyanide in Ginkgo biloba L. seed. *Food Chemistry*, 382, Article 132408. <https://doi.org/10.1016/j.foodchem.2022.132408>
- Charoenkwan, P., Nantasenamat, C., Hasan, M. M., Manavalan, B., & Shoombuatong, W. (2021). BERT4Bitter: A bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics*, 37(17), 2556–2562. <https://doi.org/10.1093/bioinformatics/btab133>
- Charoenkwan, P., Nantasenamat, C., Hasan, M. M., Moni, M. A., Lio, P., & Shoombuatong, W. (2021). iBitter-fuse: A novel sequence-based bitter peptide predictor by fusing multi-view features. *International Journal of Molecular Sciences*, 22(16). <https://doi.org/10.3390/ijms22168958>
- Chen, B., Lin, X., Liang, Z., Chang, X., Wang, Z., Huang, M., & Zeng, X. A. (2025). Advances in food flavor analysis and sensory evaluation techniques and applications: Traditional vs emerging. *Food Chemistry*, 494, Article 146235. <https://doi.org/10.1016/j.foodchem.2025.146235>
- Chen, Z., Wang, Z., Cao, Y., Shi, X., Xu, B., Li, X., & Qiao, Y. (2024). PungentDB: Bridging traditional Chinese medicine of medicine food homology and modern food flavor chemistry. *Food Chemistry: X*, 23, Article 101742. <https://doi.org/10.1016/j.fochx.2024.101742>
- Conaway, S., Jr., Huang, W., Hernandez-Lara, M. A., Kane, M. A., Penn, R. B., & Deshpande, D. A. (2024). Molecular mechanism of bitter taste receptor agonist-mediated relaxation of airway smooth muscle. *The FASEB Journal*, 38(14), Article e23842. <https://doi.org/10.1096/fj.202400452R>
- Drewnowski, A. (2001). The science and complexity of bitter taste. *Nutrition Reviews*, 59(6), 163–169. <https://doi.org/10.1111/j.1753-4887.2001.tb07007.x>
- Egan, J. M. (2024). Physiological integration of taste and metabolism. *The New England Journal of Medicine*, 390(18), 1699–1710. <https://doi.org/10.1056/NEJMr2304578>
- Fritz, F., Preissner, R., & Banerjee, P. (2021). VirtualTaste: A web server for the prediction of organoleptic properties of chemical compounds. *Nucleic Acids Research*, 49(W1), W679–W684. <https://doi.org/10.1093/nar/gkab292>
- Gradinaru, T. C., Petran, M., Dragos, D., & Gilca, M. (2021). PlantMolecularTasteDB: A database of taste active phytochemicals. *Frontiers in Pharmacology*, 12, Article 751712. <https://doi.org/10.3389/fphar.2021.751712>
- Jiang, Y., Wei, S., Ge, H., Zhang, Y., Wang, H., Wen, X., & Li, P. (2025). Advances in the identification methods of food-medicine homologous herbal materials. *Foods*, 14(4). <https://doi.org/10.3390/foods14040608>
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., & Bolton, E. E. (2025). PubChem 2025 update. *Nucleic Acids Research*, 53(D1), D1516–D1525. <https://doi.org/10.1093/nar/gkae1059>
- Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., & Wishart, D. S. (2024). DrugBank 6.0: The DrugBank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1), D1265–D1275. <https://doi.org/10.1093/nar/gkad976>
- Kong, X., Liu, C., Zhang, Z., Cheng, M., Mei, Z., Li, X., & Li, D. (2024). BATMAN-TCM 2.0: An enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. *Nucleic Acids Research*, 52(D1), D1110–D1120. <https://doi.org/10.1093/nar/gkad926>
- Lee, J., Song, S. B., Chung, Y. K., Jang, J. H., & Huh, J. (2022). BoostSweet: Learning molecular perceptual representations of sweeteners. *Food Chemistry*, 383, Article 132435. <https://doi.org/10.1016/j.foodchem.2022.132435>
- Li, F., Youn, J., Xie, K., Chan, T., Gupta, P., Yoo, A., & Tagkopoulos, I. (2026). A unified knowledge graph linking foodomics to chemical-disease networks and flavor profiles. *npj Science of Food*, 10(1), Article 33. <https://doi.org/10.1038/s41538-025-00680-9>
- Li, L., Yan, X., Chen, F., Zheng, L., Hu, Y., He, F., & Li, Q. (2023). A comprehensive review of the metabolism of citrus flavonoids and their binding to bitter taste receptors. *Comprehensive Reviews in Food Science and Food Safety*, 22(3), 1763–1793. <https://doi.org/10.1111/1541-4337.13129>
- Li, M., Guo, T., Lin, J., Huang, X., Ke, Q., Wu, Y., & Hu, C. (2022). Curcumin inhibits the invasion and metastasis of triple negative breast cancer via hedgehog/Gli1 signaling pathway. *Journal of Ethnopharmacology*, 283, Article 114689. <https://doi.org/10.1016/j.jep.2021.114689>
- Li, Y., Chen, Y., Zhou, L., You, S., Deng, H., Chen, Y., & Zhang, Y. (2020). MicroTom metabolic network: Rewiring tomato metabolic regulatory network throughout the growth cycle. *Molecular Plant*, 13(8), 1203–1218. <https://doi.org/10.1016/j.molp.2020.06.005>
- Lin, W. C., & Lin, J. Y. (2011). Five bitter compounds display different anti-inflammatory effects through modulating cytokine secretion using mouse primary splenocytes in vitro. *Journal of Agricultural and Food Chemistry*, 59(1), 184–192. <https://doi.org/10.1021/jf103581r>
- Liu, S., Grierson, D., & Xi, W. (2024). Biosynthesis, distribution, nutritional and organoleptic properties of bitter compounds in fruit and vegetables. *Critical Reviews in Food Science and Nutrition*, 64(7), 1934–1953. <https://doi.org/10.1080/10408398.2022.2119930>

- Liu, Y., Xin, H., Zhang, Y., Che, F., Shen, N., & Cui, Y. (2022). Leaves, seeds and exocarp of *Ginkgo biloba* L. (Ginkgoaceae): A comprehensive review of traditional uses, phytochemistry, pharmacology, resource utilization and toxicity. *Journal of Ethnopharmacology*, 298, Article 115645. <https://doi.org/10.1016/j.jep.2022.115645>
- Qian, J., Wang, X., Song, F., Liang, Y., Zhu, Y., Fang, Y., & Dong, J. (2025). ChemSweet: An AI-driven computational platform for next-gen sweetener discovery. *Food Chemistry*, 463(Pt 3), Article 141362. <https://doi.org/10.1016/j.foodchem.2024.141362>
- Rojas, C., Ballabio, D., Pacheco Sarmiento, K., Pacheco Jaramillo, E., Mendoza, M., & Garcia, F. (2022). ChemTastesDB: A curated database of molecular tastants. *Food Chem (Oxf)*, 4, Article 100090. <https://doi.org/10.1016/j.fochms.2022.100090>
- Roland, W. S., van Buren, L., Gruppen, H., Driesse, M., Gouka, R. J., Smit, G., & Vincken, J. P. (2013). Bitter taste receptor activation by flavonoids and isoflavonoids: Modeled structural requirements for activation of hTAS2R14 and hTAS2R39. *Journal of Agricultural and Food Chemistry*, 61(44), 10454–10466. <https://doi.org/10.1021/jf403387p>
- Shaukat, M. N., Nazir, A., & Fallico, B. (2023). Ginger bioactives: A comprehensive review of health benefits and potential food applications. *Antioxidants (Basel)*, 12(11). <https://doi.org/10.3390/antiox12112015>
- Shimizu, T., Fushimi, T., Ohno, R., Yasuyuki, F., Aso, K., Jacobs, U. M., & Osakabe, N. (2024). Verification of the interaction between human bitter taste receptor T2R46 and polyphenols; Computational chemistry approach. *Current Research in Food Science*, 9, Article 100914. <https://doi.org/10.1016/j.crfs.2024.100914>
- Soares, S., Silva, M. S., Garcia-Estevez, I., Grobetamann, P., Bras, N., Brandao, E., & Meyerhof, W. (2018). Human bitter taste receptors are activated by different classes of polyphenols. *Journal of Agricultural and Food Chemistry*, 66(33), 8814–8823. <https://doi.org/10.1021/acs.jafc.8b03569>
- Sood, S., Methven, L., & Cheng, Q. (2025). Role of taste receptors in salty taste perception of minerals and amino acids and developments in salt reduction strategies: A review. *Critical Reviews in Food Science and Nutrition*, 65(18), 3444–3458. <https://doi.org/10.1080/10408398.2024.2365962>
- Sukumaran, S. K., & Palayyan, S. R. (2022). Sweet taste Signaling: The Core pathways and regulatory mechanisms. *International Journal of Molecular Sciences*, 23(15). <https://doi.org/10.3390/ijms23158225>
- Wang, J., Wang, S., Li, D., Fan, W., Jiang, Z., Zhang, J., & Chai, G. (2026). From machine learning to neuroimaging: A comprehensive review of flavor prediction and recognition techniques. *Food Research International*, 223(Pt 1), Article 117933. <https://doi.org/10.1016/j.foodres.2025.117933>
- Wei, C., Zhou, L., Liang, B., Chen, J., Wang, G., & Li, X. (2026). Artificial intelligence revolutionize food detection? Vision, olfaction and taste integrated with machine learning/deep learning in food detection. *Food Chemistry*, 499, Article 147377. <https://doi.org/10.1016/j.foodchem.2025.147377>
- Xiao, X.-K., Liu, W.-X., Lu, X., Chen, C.-R., Zhang, X.-Y., Pu, X.-J., & Wang, F. (2026). Integrated metabolomics and transcriptomics analyses provide insights into flavor and color regulation during tomato postharvest ripening. *Postharvest Biology and Technology*, 234, Article 114155. <https://doi.org/10.1016/j.postharvbio.2026.114155>
- Xu, K., Guo, Y., Wang, Y., Yang, Z., Han, H., Meng, F., & Wang, B. (2025). Applications of biosensing in assessing food flavor: Advancements, challenges, and prospects. *Food Chemistry*, 491, Article 145313. <https://doi.org/10.1016/j.foodchem.2025.145313>
- Xu, Y., Sun, Q., Chen, W., Han, Y., Gao, Y., Ye, J., ... Yang, Y. (2022). The taste-masking mechanism of chitosan at the molecular level on bitter drugs of alkaloids and flavonoid glycosides from traditional Chinese medicine. *Molecules*, 27(21). <https://doi.org/10.3390/molecules27217455>
- Yan, D., Zheng, G., Wang, C., Chen, Z., Mao, T., Gao, J., & Cao, Z. (2022). HIT 2.0: An enhanced platform for herbal ingredients' targets. *Nucleic Acids Research*, 50(D1), D1238–D1243. <https://doi.org/10.1093/nar/gkab1011>
- Zhang, X., Du, X., Li, Y. Z., Nie, C. N., Wang, C. M., Bian, J. L., & Luo, F. (2022). Are organic acids really related to the sour taste difference between Chinese black tea and green tea? *Food Science & Nutrition*, 10(6), 2071–2081. <https://doi.org/10.1002/fsn3.2823>
- Zhang, Y., Li, X., Shi, Y., Chen, T., Xu, Z., Wang, P., & Xu, H. (2023). ETCM v2.0: An update with comprehensive resource and rich annotations for traditional Chinese medicine. *Acta Pharmaceutica Sinica B*, 13(6), 2559–2571. <https://doi.org/10.1016/j.apsb.2023.03.012>
- Ziaikin, E., David, M., Uspenskaya, S., & Niv, M. Y. (2025). BitterDB: 2024 update on bitter ligands and taste receptors. *Nucleic Acids Research*, 53(D1), D1645–D1650. <https://doi.org/10.1093/nar/gkae1044>
- Zimmermann, Y., Sieben, L., Seng, H., Pestlin, P., & Gorlich, F. (2025). A chemical language model for molecular taste prediction. *npj Science of Food*, 9(1), 122. <https://doi.org/10.1038/s41538-025-00474-z>