
A CHEMICAL LANGUAGE MODEL FOR MOLECULAR TASTE PREDICTION

Yoel Zimmermann[†] Leif Sieben[‡] Henrik Seng[‡] Philipp Pestlin[‡] Franz Görlich[‡]

Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland
{yzimmermann, lsieben, heseng, ppestlin, fgoerlich}@ethz.ch

Kvant AI Labs, Zurich, Switzerland
{yz, ls, hs, pp, fg}@kvantlabs.ai

[†]All authors have contributed equally, ordering is reverse alphabetical

ABSTRACT

Determining molecular taste remains a significant challenge in food science. Here, we present FART (Flavor Analysis and Recognition Transformer), a chemical language model capable of predicting molecular taste from chemical structure. Trained on the largest public dataset (15,025 compounds) of molecular tastants to date, FART is the first model capable of parallel predictions across four fundamental taste categories: sweet, bitter, sour, and umami. FART achieves an accuracy above 91% for parallel taste prediction and outperforms previous state-of-the-art binary classifier models that specialize on predicting one taste class. Its transformer architecture allows for interpretability through gradient-based visualization of molecular features. The model identifies key structural elements driving taste properties and demonstrates utility in analyzing known tastants as well as novel compounds. By making both the model and the dataset publicly available, we provide the food science community with tools for rapid taste prediction, potentially accelerating the development of new flavor compounds and enabling systematic exploration of taste chemistry.

Keywords Machine learning · Molecular taste prediction · Transformer model · Interpretable artificial intelligence

1 Introduction

Flavor sensation is a complex phenomenon in which concentration, prior perceptions, cultural and immediate context, individual physiology, the combination of various tastants, as well as other sensations such as smell or texture play an important role [1–3]. Determining the taste of even a single molecule in isolation remains a time- and labor-intensive process involving human panelists or electronic tongues [4, 5]. Computational tools capable of predicting the taste of a molecule *in silico* could meaningfully accelerate this task and assist in the discovery of novel tastants.

A molecule elicits a taste perception by interacting with taste receptors in the mouth as determined by its steric and electronic properties [6]. As a result, the taste of a molecule should be a function of its underlying chemical structure. Over the past decade, machine learning methods have repeatedly demonstrated the ability to computationally predict molecular properties when provided with sufficient data [7]. By learning the relationship between molecular data and chemical structure, these models can predict properties even for compounds for which no data was initially available.

In recent decades, a number of machine learning approaches have been suggested for molecular taste prediction, as has been helpfully reviewed elsewhere [8–10]. Very recently, large language models such as OpenAI’s GPT-3.5 and GPT-4 have been used for taste prediction [11, 12] although it is unclear how well these proprietary models generalize to molecules not contained in their training data [13]. Fundamentally, the performance of machine learning methods is limited by the quality and size of the dataset available. Here we curated the largest, public dataset of 15,025 molecules and their associated taste labels, which is fully accessible in accordance with the FAIR principles [14].

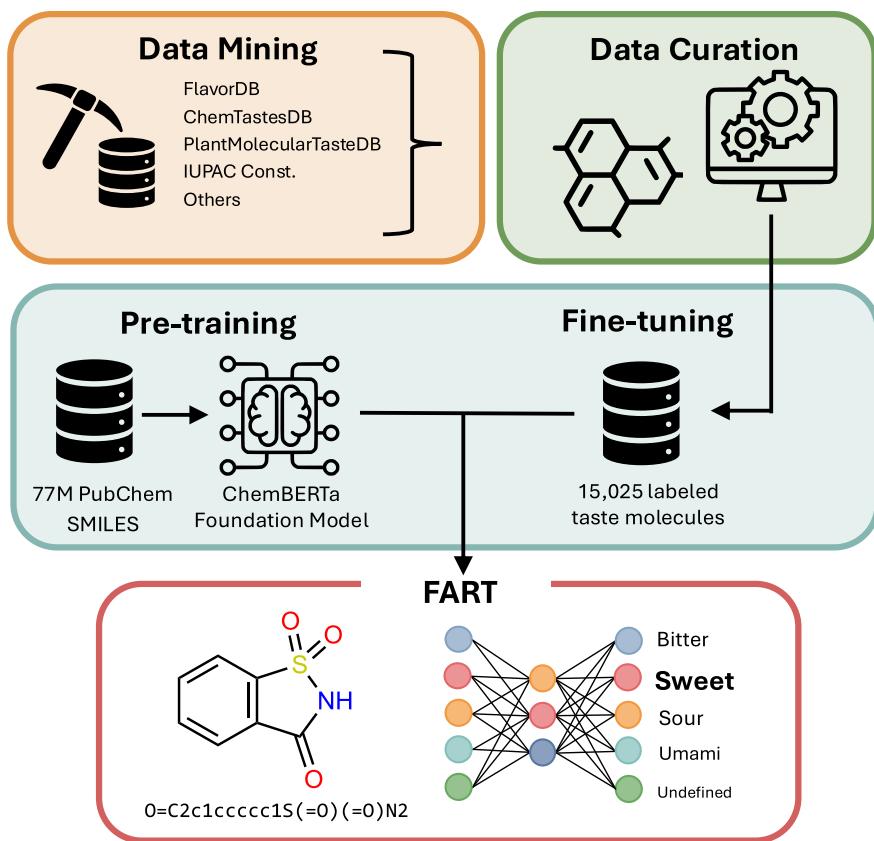


Figure 1: Schematic representation of the methodology behind the Flavor Analysis and Recognition Transformer (FART), showcasing the integration of data mining, data curation and fine-tuning.

Using this dataset, we developed a chemical language model using the transformer architecture [15], named Flavor Analysis and Recognition Transformer (FART), which compares favorably to baseline machine learning methods in a multi-class setting. While FART is trained to predict four taste classes in parallel, it consistently outperforms models specifically designed for binary classification tasks, e.g. sweet/non-sweet or bitter/non-bitter. Central to the success of transformer models is the pre-training, fine-tuning paradigm [16, 17]. In natural language processing (NLP) for which the transformer architecture was originally developed, a model is first pre-trained on a large corpus of text allowing it to learn general language structures and semantic relationships. This broad foundation is then refined through fine-tuning on smaller, domain-specific datasets, adapting the model to excel at specialized tasks. Analogously, FART is based on a pre-trained foundation model for chemistry, ChemBERTa [18, 19], which was then fine-tuned with the dataset presented here, see Fig. 1. Instead of natural language, a chemical language model takes SMILES (Simplified Molecular Input Line Entry System) [20] as input, a text-based notation that encodes chemical structures into linear strings. FART is hence capable of predicting taste from chemical structure alone, obviating the need for further data collection from either experiments or computations.

Unlike previous approaches, which were limited to predicting one to two taste classes, FART is capable of classifying molecules across four of the basic human taste categories: sweet, bitter, umami and sour. The category of "salty" was excluded as it is predominately elicited by salts containing sodium and chloride ions. Unlike other taste modalities, the perception of saltiness is largely determined by physical properties such as ion size and solubility rather than molecular structure [21]. Instead, we opted to include a category of "undefined" tastants, which includes tasteless as well as molecules with ambiguous taste profiles.

To be most useful for food scientists, such a machine learning model should not behave as a black box and allow for at least some rationalization of its predictions. The transformer architecture offers opportunities for interpretability through an analysis of its attention mechanism [15, 22] or gradient-based interpretability methods. The latter evaluates how changes in the input (such as atomic structure) impact the model's predictions, attributing the result to specific features [23]. FART is the first interpretable, chemical language model for taste prediction of small molecules across all

four common taste categories. In the future, robust and interpretable taste prediction could advance drug formulation and rational food design as well as help elucidate the interaction of molecules with their respective taste receptors.

2 Results

The performance of the models was evaluated on a subset of 15% (or 2,254 molecules) of the dataset, which had been excluded from the training. All models were evaluated on the same test set for various metrics, see Table 1. While accuracy measures how a model performs across all five taste categories, other metrics need to be calculated on a per-class basis. For the class "sweet" for example we would evaluate *precision* (proportion of predicted sweet molecules which are truly sweet), *recall* (proportion of truly sweet molecules that were also predicted sweet), the *F1 score* (which combines recall and precision into one number) and the area under the receiver operating characteristic (AUROC, which measures a model's ability to distinguish between classes). As FART is specifically tasked with parallel predictions across four taste categories, we decided to report overall performance using unweighted (macro) averages across the five taste categories. This means that a model cannot compensate bad performance on a minority class (in our case umami) with exceptionally good performance on the other, more common classes. More detailed results for each taste class individually and their weighted averages can be found in the supplementary information (SI).

We compared FART to tree-based classifier models (XGBoost, Random Forest) which are considered strong baselines for classification tasks compared to more complex model architectures such as transformer models [24]. This benchmarking relative to simpler, tree-based classifiers, which often perform exceptionally well with tabular data, is necessary to assess whether the added complexity of the transformer architecture is justified.

Table 1: Performance comparison between the trained transformers and random-forest models across several metrics. All scores are given as unweighted averages across taste classes which emphasizes minority classes, in our case umami. Area under the receiver operating characteristic (AUROC) values are calculated as one-vs-rest for each taste class. Support refers to the number of compounds out of the test set which were assigned a prediction by the model.

Model	Accuracy	Unweighted average				
		Precision	Recall	F1 Score	AUROC	Support
XGBoost: fingerprints (fp)	0.8988	0.8169	0.7400	0.7661	0.8520	100%
XGBoost: fp+descriptors	0.8962	0.8842	0.7402	0.7779	0.8513	100%
Balanced Random Forest: fp	0.7972	0.5845	0.7322	0.6014	0.8391	100%
FART	0.8860	0.7720	0.6873	0.7118	0.9639	100%
FART augmented	0.8940	0.8789	0.7388	0.7737	0.9744	100%
FART augmented + confidence	0.9155	0.9000	0.7617	0.7956	0.9806	94%

2.1 Baseline Tree-Based Classifiers

Notably, the balanced random forest method, which under-samples the majority class to ensure that all classes are equally represented during training, did not perform better even when considering the more favorable unweighted averages. XGBoost [25] combined with a larger albeit unbalanced dataset yielded superior prediction performance. Unlike the transformer models, which are trained on text-based representations of molecules (i.e. SMILES), the baseline classifiers were trained using a vector representation of molecules known as Morgan fingerprints [26]. Concatenating these fingerprints with an additional set of 15 molecular descriptors that were previously found to be highly correlated with taste [27] did not improve performance. This observation is relatively unsurprising given that these descriptors are nearly exclusively based on adjacency matrices and therefore encode little additional information compared to the atom-radius based Morgan fingerprint.

2.2 Transformer Models

Fine-tuning on the curated dataset led FART to a performance already on par with the best tree-based models showing that fine-tuned transformer models are capable of taste prediction given sufficient data, see Table 1. A common technique to synthetically expand the dataset to improve generalizability is to use SMILES augmentation. By using the fact that the same molecule can be represented by multiple SMILES strings, it is possible to generate such additional non-canonical SMILES for each data point [28]. SMILES augmentation can help improve performance more generally but is particularly useful in ensuring the model generalizes well to input SMILES which are not canonicalized. Indeed, the performance of the unaugmented iteration of FART dropped markedly (by about 6 to 10 percentage points across

Table 2: Comparison of the FART models with previously published work using state-of-the-art binary classifiers as given in [9]. Despite being trained for multi-class prediction, FART outperforms state-of-the-art methods specifically trained on predicting only sweet/non-sweet, bitter/non-bitter, sour/non-sour and umami/non-umami respectively. The size of the test set (n) is also reported. RF: Random forest. MLP: Multi-layer perceptron.

	Reference	Model Name	Classifier	n	Accuracy	F1	AUROC
Sweet/Non-Sweet	Tuwani et al. 2019 [30]	BitterSweet	AdaBoost	161	0.834	0.856	0.883
	Fritz et al. 2021 [31]	VirtualSweet	RF	403	0.893	0.888	0.951
	Bo et al. 2022 [32]	SweetMLP-Fingerprint	MLP	444	0.900	—	0.940
	Lee et al. 2022 [33]	BoostSweet	Consensus	459	0.899	0.907	0.961
	Yang et al. 2022 [34]	ChemSweet	RF	241	0.920	—	0.971
	This work	FART augmented	Transformer	2254	0.926	0.944	0.978
	This work	FART confidence	Transformer	2129	0.938	0.954	0.984
Bitter/Non-Bitter	Tuwani et al. 2019 [30]	BitterSweet	RF	154	0.819	0.838	0.880
	Fritz et al. 2022 [31]	VirtualBitter	RF	323	0.898	0.882	0.956
	Charoenkwan et al. 2021 [35]	BERT4Bitter	Transformer	128	0.922	—	0.964
	Bo et al. 2022 [32]	BitterMLP-Descriptor	MLP	446	0.820	—	0.940
	This work	FART augmented	Transformer	2254	0.958	0.778	0.951
	This work	FART confidence	Transformer	2129	0.970	0.830	0.965
Sour/Non-Sour	Fritz et al. 2021 [31]	VirtualSour	RF	133	0.977	0.842	0.994
	This work	FART augmented	Transformer	2254	0.980	0.906	0.994
	This work	FART confidence	Transformer	2129	0.986	0.935	0.997
Umami/Non-Umami	This work	FART augmented	Transformer	2254	0.998	0.5	0.989
	This work	FART confidence	Transformer	2129	0.998	0.5	0.989

metrics) when non-canonical SMILES were used as input whereas the performance of the augmented model remains essentially unchanged, see Supplementary Tables 10 and 11 as well as Supplementary Figure 3.

Training FART on this augmented dataset resulted in improved performance and made predictions much more robust towards different SMILES of the same molecule. Furthermore, this augmentation allowed us to construct a confidence metric in which a prediction is only accepted when a consensus is reached across 10 different SMILES representations for the same molecule, see also section 3.2.3. Notably, data augmentation cannot be used to address data imbalance in the training set. We found that when augmenting in an imbalanced manner, i.e. mostly augmenting our minority class of umami and not augmenting our majority class of sweet, FART greedily classified all canonicalized SMILES as sweet. Performance as measured by a one-vs-rest AUROC consistently improved for FART models compared to XGBoost baselines, which suggests that FART learned an expressive latent representation of the input molecules in which taste classes can be more easily distinguished.

Molecular tastants are not confined to a single label however, the archetypical example being bittersweet compounds [29]. In the FART dataset, such compounds led to repeated entries with the same canonicalized SMILES for each taste associated. Indeed, of the 409 molecules in the dataset with multiple tastes associated, 152 (37%) only have "undefined" as an additional label and 165 (40%) are labeled as bittersweet. Because FART outputs a probability distribution across the taste labels, one could expect the probability to be higher for the two (and in rare cases three) original labels compared to the other taste classes. In other words, a bittersweet molecule should have higher probabilities for both bitter and sweet while ignoring the remaining classes. We find that in most cases FART strongly favors one out of the two or more labels, which is unsurprising given the model was also trained and evaluated on single labeled data points, see SI for further information.

The advantage of including an "undefined" class for taste prediction is that FART is capable of flagging tasteless molecules as well as compounds that are plausibly outside the domain of application, i.e. molecular structures that are too distinct from the train set to allow for meaningful predictions. For example, instead of forcing sodium chloride erroneously into one of the four taste labels that we have collected data on, FART classifies it as undefined. Similarly, FART classifies saturated alkanes, which are not water-soluble, as undefined as well.

2.3 Comparison to Binary Classifiers

Compared to state-of-the-art classifiers tasked with binary taste prediction (sweet/non-sweet, bitter/non-bitter, sour/non-sour or umami/non-umami), FART consistently performs better despite having been trained on the more challenging task of multi-class prediction, see Table 2. Note that while the performance metrics were re-calculated for FART based on the respective binary prediction tasks, the test set FART was evaluated on is significantly larger (2254 vs below 500) and more diverse than the homogeneous test sets compiled for binary prediction such as sweet/non-sweet. Therefore,

FART has to learn a much broader chemical space and has to embed information for five rather than two labels. Despite this, FART performs superior compared to these state-of-the-art binary classifiers.

2.4 Interpretability

To better understand how FART arrives at its predictions, we implemented an interpretability technique based on integrating the gradients of the neural network's output along the path from a baseline input (the zero embedding) to the input at hand [23], see SI. For FART, this enables the visualization of atoms and functional groups that contribute positively toward the prediction of a given label (e.g., "sour") in green, while those that detract from this prediction are highlighted in red. We found that in many cases this analysis reproduces previously known patterns in food chemistry.

To test this interpretability framework, we deployed FART on six compounds with known labels as shown in Figure 2, where two (compounds **4** and **6**) were not in the dataset collected in this work. For molecules labeled as sour for example, FART typically highlights the acid group, as is the case for *p*-anisic acid (**1**). Notably, even upon a small chemical change of essentially one carbon atom, where the acid group is esterified to yield methyl anisate (**2**), FART reassigned a new label which can also be seen in the integrated gradients for the new compound. The methyl group of the ester is now relevant for the predicted sweet taste. This has some grounding in truth as esters broadly, and methyl anisate in particular, are common flavor and odor chemicals with typically fruity flavors and aromas, which would likely be labeled as sweet by a human panelist [36, 37].

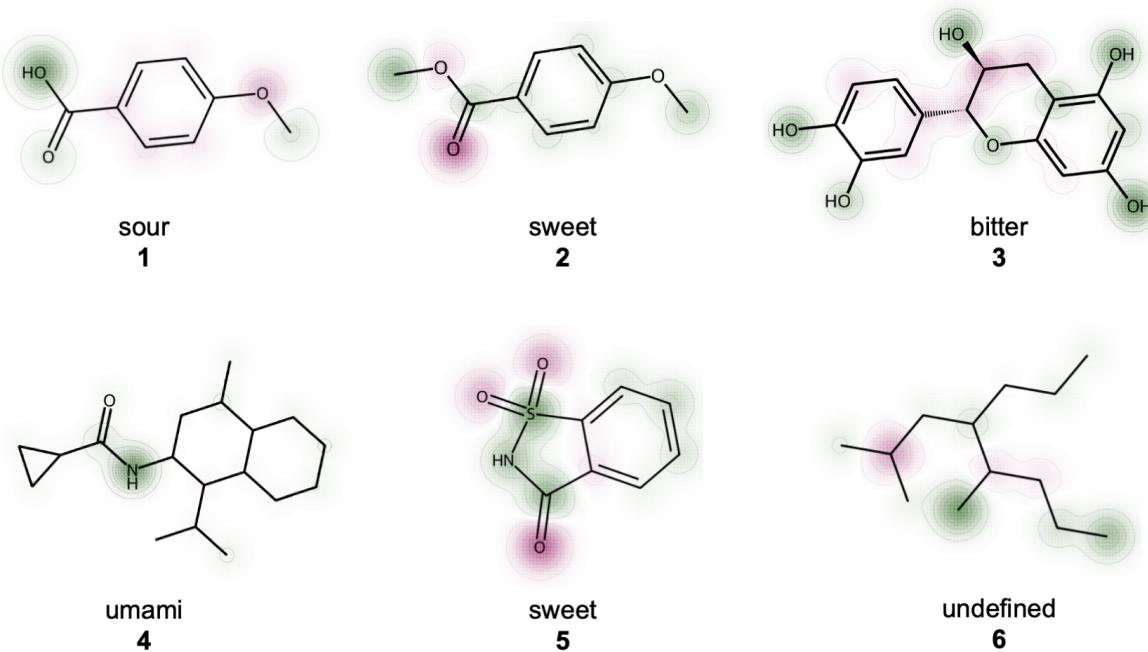


Figure 2: Using the integrated gradients method, atoms can be highlighted according to how important they were for classification. Positive contributions regarding the assigned label are highlighted in green, and negative ones in red. *p*-Anisic acid (**1**) is correctly classified as sour (pK_a 4.47) and yields methyl anisate (**2**) after esterification, which is found in star anise and has a sweet and fruity flavor. Catechin (**3**) is a bitter member of the polyphenol subgroup called flavonoids. The amide (**4**) is an analog of an umami flavor enhancer called cyclopropanecarboxylic acid (2-isopropyl-5-methyl-cyclohexyl)-amide. Saccharin (**5**) is a common artificial sweetener and 2,5-dimethyl-4-propyloctane (**6**) is a water unsoluble alkane that is not known to elicit any taste.

Similarly, FART gave higher weight to the polyphenol or more specifically the flavonoid scaffold of catechin (**3**) which is a shared substructure among many bitter-tasting compounds [38]. This approach to interpretability does also extend to molecules that were not part of our dataset, such as compound **4**, which is an analog of a known umami tastant, cyclopropanecarboxylic acid (2-isopropyl-5-methyl-cyclohexyl)-amide. Interestingly, the amide group, a feature of many umami molecules, particularly peptides, is overwhelmingly responsible for this classification. While sensible in the context of this molecule, this also highlights some of the limitations of this approach as an amide group alone is

not sufficient to elicit an umami taste. For the artificial sweetener saccharine (**5**), the sultam functional group with its sulfur-nitrogen bond is highlighted which is a pattern shared for example with another sweetener, cyclamate [39].

Even in well-behaved cases, however, the resulting heatmap is only interpretable to a degree. In the case of saccharine for example, it is less clear why the oxygen atoms are particularly anti-correlated with sweetness. Similarly for the alkane **6**, it is rather the absence of any features that leads to it being classified as undefined. Nonetheless, these interpretability tools help rationalize FART's predictions and may aid in producing analogs by highlighting parts of the structure that can be altered without changing the label, see SI for an example use case.

3 Methods

3.1 Dataset

One aim of this work was to curate a large, high-quality dataset of molecular tastants from publicly available data. Every molecule was assigned a taste label: sweet, bitter, sour, umami, or undefined. The category of "undefined" contains molecules, not clearly assignable to the other categories, some of which may not have any taste associated with them. Multiple entries were created for compounds with more than one taste label.

The FART dataset aggregates data from six sources [40–44], see SI for further details. To visualize the distribution of the taste labels over the chemical space covered by the dataset, a t-SNE plot (Fig. 3a) was generated from Morgan fingerprints [26]. Fig. 3b highlights the strong imbalance of the taste classes of the dataset.

As ChemBERTa, the chemical language model used in this work, was pre-trained mostly on SMILES of small molecules, we curated our database accordingly with an average molecular weight of $374 \text{ Da} \pm 228$, see Fig. 3c as well as the SI for extended methodology. This notably excludes longer peptides which we do not consider in our approach. Previous work has explored this chemical space particularly in the context of umami and bitter prediction [45–47]. Entries with invalid SMILES, charged molecules, and multiple neutral fragments, e.g. SMILES which included solvents, were removed. Data curation was performed with the cheminformatics package RDKit [48]. All SMILES were canonicalized and subjected to the default RDKit standardization procedure. Duplicate removal based on canonicalized SMILES reduced the dataset by almost half to a final size of 15,025 entries. The large number of duplicates underlines the significant overlap among the databases used.

The curated dataset was further enriched by general information (PubChemID, IUPAC name, molecular formula, molecular weight, InChI, InChIKey), accessed through the PubChem API [49]. The FART dataset was made public in agreement with the FAIR principles [14] and can be accessed through several different interfaces to encourage its use by other research projects.

3.2 Models

3.2.1 Tree-Based Classifiers: XGBoost, Random Forest

Tree-based ensemble models such as random forest (RF) or XGBoost [25] are often considered strong baseline models in classification tasks for more complex model architectures, such as transformers, given their robust performance, efficient training, and low model complexity. A RF model is an ensemble learning method that trains multiple, in this case 150, decision trees during the training process, combining their predictions to improve accuracy and mitigate overfitting on the training data. A XGBoost model is an ensemble learning method that builds multiple decision trees sequentially, optimizing each tree to correct errors from the previous ones, thereby improving accuracy and mitigating overfitting through regularization techniques.

In this work, three different tree-based classifiers were trained including hyperparameter optimization. The two XGBoost models were trained either using 1024-Morgan (radius=2) fingerprints or these same fingerprints concatenated with 15 descriptors calculated using Mordred [50]. These predictors had been previously found to be particularly correlated with taste [27]. The models were evaluated with a multi-class logarithmic loss function, results are given in Table 1.

3.2.2 FART Models

For our transformer-based model, we utilized ChemBERTa [18, 19], which has been pre-trained on 77 million SMILES strings using a masked language modeling approach. Here, a percentage of the input is randomly masked, and the model is trained to predict the masked parts of the SMILES string. This allows the model to learn contextualized representations of chemical structures in a self-supervised manner, capturing both local and global molecular features. We fine-tuned ChemBERTa on our taste prediction dataset using a categorical cross-entropy loss function with a learning

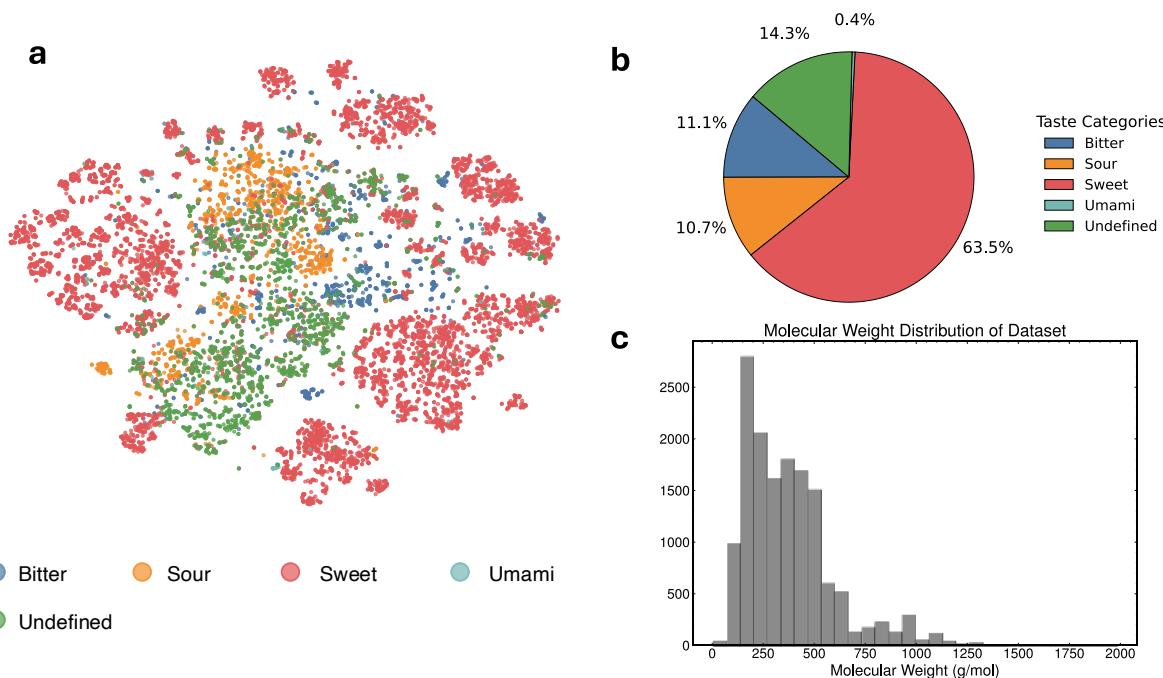


Figure 3: (a) t-SNE plot of the chemical space covered by the dataset. (b) Distribution across taste classes for the dataset, showing a strong imbalance for umami. (c) Molecular weight distribution for the dataset (mean 374 g/mol, std. deviation 228 g/mol).

rate of 10^{-5} and a batch size of 16. We experimented with three different training configurations: one model was trained for 20 epochs on the original dataset and a second model was trained for 2 epochs on a 10-fold augmented dataset. The performance of all models is summarized in Table 1.

3.2.3 Confidence Metric

Similar to linguistic synonymy, where multiple words share the same meaning, distinct SMILES representations can map to the same underlying molecular structure. To leverage this property, we generated an ensemble of 10 synonymous SMILES for each molecule in our dataset. We then performed inference on the entire ensemble, obtaining individual predictions for each SMILES variant.

To aggregate these results, we employed a voting procedure across the ensemble's predictions. Using the strictest threshold, where all 10 predictions had to agree for a label to be assigned by the model, we still retained predictions for 94% of the dataset while boosting the model's accuracy to above 91%. Both the transient augmentation and subsequent prediction on these SMILES adds to the computational time. The confidence metric suggested here is straightforward to implement and provides a robust indication of prediction reliability in addition to the interpretability framework.

4 Conclusion

The chemical language model FART, fine-tuned on a large corpus of labeled small molecule tastants, performs on par or superior compared to baseline tree-based ensemble methods. Enabled by the large training set of 15,025 labeled tastants collected in this work, FART is the first model capable of predictions across all four major taste categories (sweet, sour, bitter, umami) while accommodating other, including tasteless, molecules into a fifth, undefined category. While trained on multi-class taste prediction, FART outperforms state-of-the-art binary classifiers that specialize on predicting one taste class, e.g. sweet/non-sweet. FART gives robust and fast predictions based on chemical structure alone thereby avoiding the computational overhead of calculating physical or chemical descriptors.

To the best of our knowledge, the dataset presented in this work includes the large majority of high-quality labeled taste data publicly available today and may serve as a useful resource for future work on small molecule taste prediction. Nonetheless, the scarcity of data on umami compounds makes this a particularly challenging taste category also for

FART. Progress is already underway to build increasingly powerful chemical language models that could be fine-tuned using a similar approach as outlined here. However, we believe meaningful breakthroughs in taste prediction will not come from methodological advancements alone but requires the collection of larger, well-curated, and accessible datasets. Automated, experimental taste classification remains a large unmet need in the field of food chemistry today.

An advantage of using language models pre-trained on chemical data is that they typically generalize better [18, 51] to unseen molecules which is particularly relevant when applying these models for the discovery of novel small molecule tastants. In the future, FART and methods like it may be used to perform *in-silico* screenings of large chemical spaces to discover novel tastants with desirable properties similar to modern drug discovery approaches [52]. Future work should also focus on improving predictions for molecules with multiple associated tastes as well as making FART more sensitive toward stereochemistry. We appreciate that the use of synthetic chemicals to enhance or alter food flavor is not universally desirable. However, machine learning tools such as FART may also plausibly aid food scientists during the analysis of natural foodstuffs by shortlisting flavor-rich small molecules from a mixture of compounds. In this way, only a small number of promising compounds would need to be experimentally validated. Such an approach is highly dependent on the robustness of prediction and would motivate extending these predictions to proteins and other macromolecules as well.

Artificial intelligence has a role in assisting the food scientist of tomorrow. Taste prediction tools offer great utility for the discovery, development as well as analysis of small molecule tastants even if this does not obviate the need for thorough experimental validation. Machine learning tools like FART could open up new and exciting chemical spaces for food scientists going forward and would unquestionably bring molecular food science into the age of big data and deep learning.

Acknowledgements

The authors thank Prof. Dr. Kjell Jorner, Mr. Stefan Schmid, and Mrs. Lauriane Jacot-Descombes for an inspirational, inaugural rendition of the course "529-0150-00L Digital Chemistry" at ETH Zurich which kickstarted this work. We thank Mr. Stefan Schmid and Mrs. Carole Zermatten for their helpful comments on an earlier version of this manuscript.

Data availability

The data used and generated in our study is available at <https://github.com/fart-lab/fart.git>.

References

1. Jorge Regueiro, N. N. & Simal-Gándara, J. Challenges in relating concentrations of aromas and tastes with flavor features of foods. *Critical Reviews in Food Science and Nutrition* **57**, 2112–2127 (2017).
2. Small, D. M. Flavor is in the brain. *Physiology & Behavior* **107**. Special Issue : Flavor and Feeding, 540–552. ISSN: 0031-9384. <https://www.sciencedirect.com/science/article/pii/S0031938412001527> (2012).
3. Auvray, M. & Spence, C. The multisensory perception of flavor. *Consciousness and Cognition* **17**, 1016–1031. ISSN: 1053-8100. <https://www.sciencedirect.com/science/article/pii/S1053810007000657> (2008).
4. Tahara, Y. & Toko, K. Electronic Tongues – A Review. *IEEE Sensors Journal* **13**, 3001–3011 (2013).
5. Latha, R. S. & Lakshmi, P. K. Electronic tongue: An analytical gustatory tool. *Journal of Advanced Pharmaceutical Technology & Research* **3**. ISSN: 2231-4040. 10.4103/2231-4040.93556 (2012).
6. Lindemann, B. Taste reception. *Physiological Reviews* **76**. PMID: 8757787, 719–766. <https://doi.org/10.1152/physrev.1996.76.3.719> (1996).
7. Walters, W. P. & Barzilay, R. Applications of Deep Learning in Molecule Generation and Molecular Property Prediction. *Accounts of Chemical Research* **54**, 263–270. ISSN: 0001-4842. <https://doi.org/10.1021/acs.accounts.0c00699> (2021).
8. Malavolta, M. *et al.* A survey on computational taste predictors. *European Food Research and Technology* **248**, 2215–2235. ISSN: 1438-2385 (2022).
9. Rojas, C., Ballabio, D., Consonni, V., Suárez-Estrella, D. & Todeschini, R. Classification-based machine learning approaches to predict the taste of molecules: A review. *Food Research International* **171**, 113036. ISSN: 0963-9969 (2023).
10. Song, Y. *et al.* A Comprehensive Comparative Analysis of Deep Learning Based Feature Representations for Molecular Taste Prediction. *Foods* **12**. ISSN: 2304-8158 (2023).

11. Song, R., Liu, K., He, Q., He, F. & Han, W. Exploring Bitter and Sweet: The Application of Large Language Models in Molecular Taste Prediction. *Journal of Chemical Information and Modeling* **64**, 4102–4111. ISSN: 1549-9596. <https://doi.org/10.1021/acs.jcim.4c00681> (2024).
12. Qian, C., Tang, H., Yang, Z., Liang, H. & Liu, Y. Can Large Language Models Empower Molecular Property Prediction? 2023. arXiv: 2307.07443 [cs.LG]. <https://arxiv.org/abs/2307.07443>.
13. Zhang, A. K. *et al.* Language model developers should report train-test overlap 2024. arXiv: 2410.08385 [cs.LG]. <https://arxiv.org/abs/2410.08385>.
14. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. ISSN: 2052-4463. <https://doi.org/10.1038/sdata.2016.18> (2016).
15. Vaswani, A. *et al.* Attention is All you Need in *Advances in Neural Information Processing Systems* (eds Guyon, I. *et al.*) **30** (Curran Associates, Inc., 2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
16. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving Language Understanding by Generative Pre-Training.
17. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* 2019. arXiv: 1810.04805 [cs.CL]. <https://arxiv.org/abs/1810.04805>.
18. Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction. *arXiv Preprint*. eprint: 2010.09885 (2020).
19. Ahmad, W., Simon, E., Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa-2: Towards Chemical Foundation Models. *arXiv Preprint*. eprint: 2209.01712 (2022).
20. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. en. *Journal of Chemical Information and Computer Sciences* **28**, 31–36. ISSN: 0095-2338. <http://dx.doi.org/10.1021/ci00057a005> (Feb. 1, 1988).
21. Taruno, A. & Gordon, M. D. Molecular and Cellular Mechanisms of Salt Taste. *Annual Review of Physiology* **85**, 25–45. ISSN: 1545-1585 (2023).
22. Bibal, A. *et al.* Is Attention Explanation? An Introduction to the Debate in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (eds Muresan, S., Nakov, P. & Villavicencio, A.) (Association for Computational Linguistics, Dublin, Ireland, May 2022), 3889–3900. <https://aclanthology.org/2022.acl-long.269>.
23. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. <https://arxiv.org/abs/1703.01365> (2017).
24. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on tabular data? 2022. arXiv: 2207.08815 [cs.LG]. <https://arxiv.org/abs/2207.08815>.
25. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *CoRR abs/1603.02754*. arXiv: 1603.02754. <http://arxiv.org/abs/1603.02754> (2016).
26. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* **5**, 107–113. <https://doi.org/10.1021/c160017a018> (1965).
27. Androutsos, L. *et al.* Predicting multiple taste sensations with a multiobjective machine learning method. *npj Science of Food*.
28. Bjerrum, E. J. SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules. *CoRR abs/1703.07076*. arXiv: 1703.07076. <http://arxiv.org/abs/1703.07076> (2017).
29. Di Pizio, A., Shoshan-Galeczki, Y. B., Hayes, J. E. & Niv, M. Y. Bitter and sweet tasting molecules: It's complicated. *Neuroscience letters* **700**, 56–63 (2019).
30. Tuwani, R., Wadhwa, S. & Bagler, G. BitterSweet: Building machine learning models for predicting the bitter and sweet taste of small molecules. *Scientific Reports* **9**, 7155. ISSN: 2045-2322 (2019).
31. Fritz, F., Preissner, R. & Banerjee, P. VirtualTaste: a web server for the prediction of organoleptic properties of chemical compounds. *Nucleic Acids Research* **49**, W679–W684. ISSN: 0305-1048. eprint: <https://academic.oup.com/nar/article-pdf/49/W1/W679/38841873/gkab292.pdf>. <https://doi.org/10.1093/nar/gkab292> (Apr. 2021).
32. Bo, W. *et al.* Prediction of bitterant and sweetener using structure-taste relationship models based on an artificial neural network. *Food Research International* **153**, 110974. ISSN: 0963-9969. <https://www.sciencedirect.com/science/article/pii/S096399692200031X> (2022).
33. Lee, J., Song, S. B., Chung, Y. K., Jang, J. H. & Huh, J. BoostSweet: Learning molecular perceptual representations of sweeteners. *Food Chemistry* **383**, 132435. ISSN: 0308-8146. <https://www.sciencedirect.com/science/article/pii/S0308814622003971> (2022).

34. Yang, Z.-F. *et al.* A novel multi-layer prediction approach for sweetness evaluation based on systematic machine learning modeling. *Food Chemistry* **372**, 131249. ISSN: 0308-8146. <https://www.sciencedirect.com/science/article/pii/S030881462102255X> (2022).
35. Charoenkwan, P., Nantasenamat, C., Hasan, M. M., Manavalan, B. & Shoombuatong, W. BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics* **37**, 2556–2562. ISSN: 1367-4803. <https://doi.org/10.1093/bioinformatics/btab133> (Feb. 2021).
36. FDA. *Substances Added to Food Inventory* <https://hfppappexternal.fda.gov/scripts/fdcc/index.cfm?set=FoodSubstances&id=METHYLANISATE> (2024).
37. Brault, G., Shareck, F., Hurtubise, Y., Lépine, F. & Doucet, N. Short-Chain Flavor Ester Synthesis in Organic Media by an *E. coli* Whole-Cell Biocatalyst Expressing a Newly Characterized Heterologous Lipase. *PLOS ONE* **9**, 1–9. <https://doi.org/10.1371/journal.pone.0091872> (Mar. 2014).
38. Drewnowski, A. The Science and Complexity of Bitter Taste. *Nutrition Reviews* **59**, 163–169. ISSN: 0029-6643. eprint: <https://academic.oup.com/nutritionreviews/article-pdf/59/6/163/24090113/nutritionreviews59-0163.pdf>. <https://doi.org/10.1111/j.1753-4887.2001.tb07007.x> (June 2001).
39. Chattopadhyay, S., Raychaudhuri, U. & Chakraborty, R. Artificial sweeteners – a review. *Journal of Food Science and Technology* **51**, 611–621. ISSN: 0975-8402. <https://doi.org/10.1007/s13197-011-0571-1> (2014).
40. Rojas, C. *et al.* ChemTastesDB: A curated database of molecular tastants. *Food Chemistry: Molecular Sciences* **4**, 100090. ISSN: 2666-5662 (2022).
41. Garg, N. *et al.* FlavorDB: a database of flavor molecules. *Nucleic Acids Research* **46**, D1210–D1216. ISSN: 0305-1048 (Oct. 2017).
42. Gradinaru, T.-C., Petran, M., Dragos, D. & Gilca, M. PlantMolecularTasteDB: A Database of Taste Active Phytochemicals. *Frontiers in Pharmacology* **12**. ISSN: 1663-9812 (2022).
43. Bayer, S. *et al.* Chemoinformatics View on Bitter Taste Receptor Agonists in Food. *Journal of Agricultural and Food Chemistry* **69**, 13916–13924 (2021).
44. Suess, B., Festring, D. & Hofmann, T. in *Flavour Development, Analysis and Perception in Food and Beverages* (eds Parker, J., Elmore, J. & Methven, L.) 331–351 (Woodhead Publishing, 2015). ISBN: 978-1-78242-103-0.
45. Charoenkwan, P. *et al.* iBitter-SCM: Identification and characterization of bitter peptides using a scoring card method with propensity scores of dipeptides. *Genomics* **112**, 2813–2822. ISSN: 0888-7543 (2020).
46. Charoenkwan, P., Yana, J., Nantasenamat, C., Hasan, M. M. & Shoombuatong, W. iUmami-SCM: A Novel Sequence-Based Predictor for Prediction and Analysis of Umami Peptides Using a Scoring Card Method with Propensity Scores of Dipeptides. *Journal of Chemical Information and Modeling* **60**, 6666–6678. ISSN: 1549-9596 (2020).
47. Charoenkwan, P. *et al.* iBitter-Fuse: A Novel Sequence-Based Bitter Peptide Predictor by Fusing Multi-View Features. *International Journal of Molecular Sciences* **22**. ISSN: 1422-0067 (2021).
48. Landrum, G. *et al.* *rdkit/rdkit: 2020_03_1 (Q1 2020) Release* version Release_2020_03_1. Mar. 2020. <https://doi.org/10.5281/zenodo.3732262>.
49. Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Research* **51**, D1373–D1380. ISSN: 0305-1048 (Oct. 2022).
50. Moriwaki, H., Tian, Y.-S., Kawashita, N. & Takagi, T. Mordred: a molecular descriptor calculator. *Journal of Cheminformatics* **10**, 4. ISSN: 1758-2946 (2018).
51. Honda, S., Shi, S. & Ueda, H. R. *SMILES Transformer: Pre-trained Molecular Fingerprint for Low Data Drug Discovery* 2019. arXiv: 1911.04738 [cs.LG]. <https://arxiv.org/abs/1911.04738>.
52. Wallach, I. *et al.* AI is a viable alternative to high throughput screening: a 318-target study. *Scientific Reports* **14**, 7526. ISSN: 2045-2322 (2024).