Classification of Tastants: A Deep Learning Based Approach

Prantar Dutta, Deepak Jain, Rakesh Gupta* and Beena Rai

Physical Sciences Research Area, Tata Research Development and Design Centre, TCS Research, 54-B, Hadapsar Industrial Estate, Pune- 411013, India

* *Correspondence to*: Rakesh Gupta

Email: gupta.rakesh2@tcs.com, Phone: + 91-20-66086422

Abstract

Predicting the taste of molecules is of critical importance in the food and beverages, flavor, and pharmaceutical industries for the design and screening of new tastants. In this work, we have built deep learning models to classify sweet, bitter, and umami molecules— the three basic tastes whose sensation is mediated by G protein-coupled receptors. An extensive dataset containing 1466 bitter, 1764 sweet, and 238 umami tastants was curated from existing literature. We analyzed the chemical characteristics of the molecules, with special focus on the presence of different functional groups. A deep neural network model based on molecular descriptors and a graph neural network model were trained for taste prediction. The class imbalance due to fewer umami molecules was tackled using special sampling techniques. Both models show comparable performance during evaluation, but the graph-based model can learn task-specific representations from the molecular structure without requiring handcrafted features. We further explain the deep neural network predictions using Shapley additive explanations. Finally, we demonstrated the applicability of the models by screening bitter, sweet, and umami molecules from a large food database. This study develops an in-silico

approach to classify molecules based on their taste by leveraging the recent progress in deep learning, which can serve as a powerful tool for tastant design.

Keywords: Tastant, Sweet, Bitter, Umami, Multiclass Classification, Deep Learning, Graph Neural Network, SHAP

1 Introduction

Taste is a sensory modality that governs the interaction of humans with the food they eat. The gustatory system, which is responsible for the sense of taste, differentiates healthy nutrients from harmful toxins, ensuring survival and a high quality of life. The interplay of five basic tastes – sweet, bitter, sour, salty, and umami, constitutes our taste experience [1]. The brain associates each taste with an underlying chemical characteristic. Sweetness and umaminess indicate the presence of carbohydrates and proteins, respectively, while sourness is linked with acids, often due to food spoilage. The unpleasant bitter sensation informs us to avoid the ingestion of rotten food and toxic substances, although safe edibles like cocoa and coffee can be bitter too. Saltiness relates to minerals like sodium that are essential for regulating body fluids. In combination with olfaction (sense of smell) and somatosensation (sense of touch, temperature, and pain), gustation determines our overall perception of flavors [2].

Taste prediction and the design of tastants play a crucial role in food and beverages, flavor, and pharmaceutical industries. For example, the search for safe artificial low-calorie sweeteners with similar chemosensory profile as sucrose is still an open research problem. The demands for the specialized flavors in the consumer product landscape is evolving rapidly to keep up with

the latest trends, and medicinal chemists are constantly looking for taste modulators to combine with oral drug formulations. Although taste perception shows variability among individuals and demographics due to genetic [3], cultural [4], and medical [5] factors, specific biological mechanisms, which are common to all humans, exist for perceiving each of the basic taste qualities. Thus, rational approaches can be conceived to design and screen tastants based on their chemistry and interactions with the gustatory system.

Thousands of small protuberances called papillae cover the tongue, each of which contains hundreds of taste buds [6]. Each of the taste bud comprises of 50-100 taste cells having specialized sensing receptors. Tastants stimulate the receptor cells, leading to signal transmission to the gustatory cortex in the brain. Sweet, bitter, and umami receptors belong to the family of G protein-coupled receptors (GPCRs) [7]. For structure-based computational design and screening of tastant molecules, elucidation of tastant-receptor interactions and the subsequent conformational dynamics of the receptor complex play a critical role. However, experimental structures of taste receptors are still elusive and *in silico* techniques like homology modeling, molecular docking, and molecular dynamics have been applied as surrogates to facilitate virtual screening of tastants [8]. A detailed discussion on taste receptors and their molecular modeling can be found in the review by Pallante and co-workers [9]. But the absence of accurate receptor structures decreases the reliability of the results, and the computational cost prohibits high-throughput screening of large databases. Moreover, no simple correlation can be established between the tastant-receptor binding affinity and the taste quality or intensity, further complicating the task.

Ligand-based structure-property relationship models offer an exciting alternative to rigorous molecular modeling techniques. Recent advances in computing capabilities and machine learning (ML) algorithms, and availability of curated datasets, make it possible to map molecular features to taste qualities with high accuracy. Initial studies mostly focused on building sweet/non-sweet and bitter/non-bitter classification models using standard molecular descriptors from cheminformatics tools, wherein tastant databases constituted the positive set and random molecules were selected to create the negative set [10, 11, 12, 13, 14]. Later, efforts were made to address the sweet-bitter dichotomy by considering both the taste qualities together [15, 16, 17, 18]. A few studies also reported predictive models to quantify the taste intensity of molecules using properties like relative sweetness with respect to sucrose [19, 20, 21]. Recently, advanced techniques like ensemble learning, gradient boosting, and transformers were employed for sweeteners and bitterants as well, which outperformed traditional machine learning models [22, 23, 24, 25, 26]. Data-driven investigations with umami focused mostly on peptides— umami/non-umami classifiers were built using the UMP442 database containing 140 umami and 302 non-umami peptides [27, 28, 29]. A comprehensive discussion on databases and ML approaches related to tastants can be found in the recent review by Malavolta, Pallante, and co-workers [30]. Integrated data and structure-based modeling frameworks, combining structure-property relationship and molecular docking, were also proposed to identify potential sweeteners [31] and umami molecules [32].

Existing studies on computational taste prediction mostly considers sweet and bitter tastants only. However, the approaches can, in principle, be extended to predict multiple taste qualities simultaneously. The primary hurdle is the lack of large, curated datasets that can be exploited

to build predictive models. For example, not many umami molecules are known, although their perception mechanism is like sweet and bitter molecules (via GPCRs). The exceptional progress in molecular ML in recent years, especially deep learning (DL) and graph-based models, offers a range of tools to make better predictions using existing and limited data [33, 34]. Nonetheless, the enigmatic power of DL algorithms limits the interpretability and explainability of models, which leads to skepticism about deployment for industrial use.

In this work, we develop multiclass classifiers for differentiating between bitter, sweet, and umami molecules. We apply deep neural network (DNN), also known as multilayer perceptron, and graph neural network (GNN) models on an extensive dataset curated from multiple sources in literature. GNNs are especially attractive because molecules can be easily represented as graphs, with atoms as nodes and bonds as edges [35]. Additionally, GNNs can work without including expert handcrafted features that require domain knowledge. We further try to make sense of the results from our DNN models using state-of-the-art explainability methods. An analysis of the functional groups in the tastants is presented as well, with an aim to relate the taste quality with chemistry. Our work widens the scope and advances the applicability of in silico taste prediction using data-driven techniques by inheriting latest developments in DL, coupled with insights from chemistry.

2 Methods

2.1 Dataset

A collection of bitter, sweet, and umami molecules was curated using information from the ChemTastesDB database [36], the datasets made available by Tuwani and co-workers [16], and the UMP442 database (available at https://github.com/Shoombuatong/Dataset-Code/tree/master/iUmami). All three datasets include tastants from multiple repositories and earlier works on data-based modeling, the details of which can be found in the respective papers. ChemTastesDB has 2944 verified tastants, both organic and inorganic, belonging to nine classes, including the five basic tastes and four additional categories, namely tasteless, multitaste, non-sweet, and miscellaneous. We extracted the canonical simplified molecular input line entry system (SMILES) representations and the corresponding taste labels of sweet, bitter, and umami compounds from the database. Similarly, the sweet and bitter molecules used by Tuwani et al. to build the BitterSweet models were obtained. The UMP442 database contains the sequence information of 140 umami and 302 non-umami peptides. We obtained the umami peptides and converted the sequences into SMILES representation using the cheminformatics tool RDKit. The three sets of canonical SMILES were merged and standardized, and the duplicates were removed to create our initial raw tastant database. For some of the molecules, RDKit could not generate descriptors, and hence they were dropped. The final database contained 1466 bitter molecules, 1764 sweet molecules, and 238 umami molecules, making up a total of 3468 tastants. For practical applications of multiclass classification models, a control set comprising of molecules that exhibit neither of the three taste qualities is required. We prepared this control set using the salty, sour, and tasteless entries in ChemTastesDB, which made up a total of 238 molecules. Henceforth, all analysis and model

development are performed using the 3706 SMILES strings belonging to four classes – bitter, sweet, umami, and control.

2.2 Featurization and Data Preprocessing

We generated molecular descriptors using RDKit for building DNN models. The Descriptors module of RDKit returns a list of 2-D 208 features for each molecule, which represent the structural, physical, and chemical information of the molecule as numerical values. They include a wide variety of molecular properties and fragment counts. However, not all features are relevant for a particular dataset or task and including them can lead to poor model quality. First, we removed all the constant and quasi-constant features in the dataset, as well as those features which could not be calculated for a significant number of molecules using RDKit. Then, we dropped the features for which less than 20 % of the molecules had non-zero values to prevent overfitting. These data cleaning steps reduced the feature set to 118. Subsequently, we performed correlation analysis of the features using Pearson correlation coefficient, which measures the strength of linearity between two variables. Highly correlated features introduce bias in ML models, leading to poor performance and generalization. In our dataset, if two features had a correlation coefficient greater than 0.7, one of them was randomly dropped. Finally, our clean and processed dataset consisted of 41 features for the 3706 molecules.

The dataset was split into training and test sets with a train-test ratio of 85:15 for model building and evaluation. Multiple training and test sets with different random states were created to ensure the reliability of results (see Supporting Information). We applied min-max

scaling and one-hot encoding to transform the features and taste labels, respectively. All preprocessing was performed with the Scikit-learn package.

For building the GNN model, we obtained the SMILES strings of the molecules from the database for conversion to graph objects, keeping the training and test sets same as previously discussed. In the GNN framework, molecules are treated as undirected graphs. Each heavy atom (non-hydrogen) in a molecule is considered as a node, and we compute the following node features: one-hot encoding of the element, degree of the atom, whether the atom is aromatic or not, number of attached hydrogen atoms, and the implicit valence. The bonds are homologous to graph edges with the following edge features: one-hot encoding of the bond order or aromaticity, whether the bond is part of a ring, and whether the bond is conjugated. In addition, an adjacency matrix is generated for each molecule which contains information about the neighbors of all the atoms. The atom and node features are same as those used by Duvenaud et al. in their paper on graph convolutions on which our GNN model is based [37].

2.3 Model Development

We first built a DNN model to classify the 3706 tastants in our database into four taste classes— bitter, sweet, umami, and control. The architecture consisted of the input layer, two hidden layers, and the output layer. Both the hidden layers were made up of 128 neurons each and activated by a rectified linear unit (ReLU) function. To reduce overfitting, the dropout technique was employed after the first hidden layer with a probability 0.3. Four output neurons, with SoftMax activation, predicted the probability of molecules belonging to each of the three classes. We chose the class with the highest probability as the model output for taste

prediction. The input data was fed to the DNN model in batches of 32. The Adam optimizer with a learning rate of 0.0001 and the categorical cross-entropy loss function were used to train the model. 15 % of the training data was kept aside for validation, and the model with lowest validation loss obtained during the training process of 200 epochs was saved as the best model. We arrived at the chosen architecture after running multiple experiments with different number of hidden layers and a range of values for the number of units in each hidden layer, along with various dropout strategies, to maximize validation accuracy and ensure minimal overfitting. Finally, the model was evaluated on the test set by computing the overall accuracy, confusion matrix, and classwise precision, recall, and F1 scores. The Keras API of TensorFlow 2 was used to implement the DNN model.

As our dataset includes far fewer umami and control compounds than sweet or bitter, the problem of class imbalance arises. Imbalanced datasets can lead to inferior performance for the minority class, even though the overall accuracy may be quite high. We attempt to solve this problem using the synthetic minority oversampling technique (SMOTE), a popular method of data augmentation which generates synthetic datapoints based on the original data but not its duplicates [38]. SMOTE randomly selects instances of the minority classes, finds their five nearest neighbors of the same class and chooses one at random, and synthetically generates samples as a convex combination of the two points in the feature space. We applied the 'not majority' settings for SMOTE, wherein all classes except the majority class are resampled.

We further built a GNN model based on convolutional neural networks operating directly on molecular graphs, as proposed by Duvenaud and co-workers [37]. The architecture consists of two identical convolution blocks, each of which is made of a graph convolution layer and a

graph pooling layer, with batch normalization applied between the two layers. The convolution blocks update the per-atom feature vectors in a non-linear way by incorporating information from its bonds and adjacent atoms. A channel width of 128 and ReLU activation function is used for the graph convolution layer and max-pooling is used for aggregating the neighborhood information of an atom. Finally, a graph gather layer combines the node-level feature vectors into a single graph-level feature vector that represents the entire molecule. This feature vector is then passed through a dense layer of 150 neurons to the output layer that predicts the desired probabilities of the four taste classes. The dropout technique was used after each layer with probability 0.1 to reduce overfitting. The applied layers, node and edge features, and model mechanisms are the same as the original paper; we finetuned the channel width, dense layer units, and other parameters for our task. Figure 1 presents a schematic of the information flow in the GNN architecture. Additional details regarding the model architecture can be found in the original paper. Like DNN training, the input data was fed to the model in batches of 32, the categorical cross entropy loss function was employed, and 15 % of the training data was kept aside for validation. The model was trained for 50 epochs and the hyperparameters were optimized using random search to minimize the validation loss. We implemented the GNN model using the DeepChem framework [39].

The class imbalance problem is particularly challenging for GNNs to tackle. We experimented with oversampling using SMILES enumeration. 7 variants of each SMILES string were generated for the umami and control molecules in the training set and augmented with the original training data. This additional data made the number of umami and control samples of the same order as bitter and sweet. It is to be noted that although the initial inputs for the augmented

date are different, GNNs are, by nature, permutation invariant. Hence, the learnt representation after the two convolutional blocks is same for all the SMILES variant of a molecule, which effectively makes the data transformation a case of oversampling.

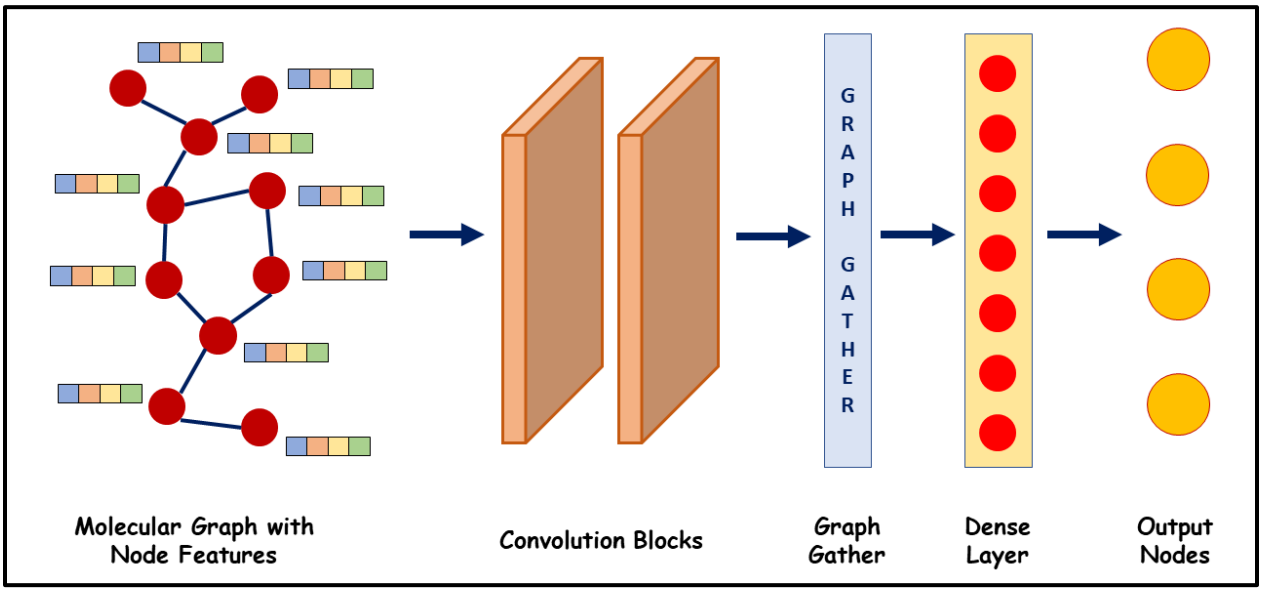

Figure 1: Flow of information in the graph neural network architecture. The input molecular graph with its node features is processed by two convolution blocks. A graph gather layer combines the per-atom representations to generate a molecule-level fingerprint vector, which is processed by a dense layer. The four output nodes predict the probability of a molecule being either bitter, sweet, umami, or control.

3 Results and Discussion

3.1 Exploratory Data Analysis

The final set of 3706 tastants, belonging to any one of the three taste qualities— bitter, sweet, or umami, or the control set, was analyzed for characteristics, patterns, and insights. Figure 2

(a) shows the density distribution plot of molecular weights of the tastants. Molecular weight relates to the size of molecules, and hence affects ligand binding to a receptor. We observe that most tastants have molecular weights within 1000 Daltons, and thus can be considered as small molecules, although a few molecules with higher weights exhibit taste qualities as well. The molecular weights are normally distributed, with umami tastants more likely to be heavier than sweet and bitter compounds. The molecules in the control set, which are sour, salty, or tasteless, show a normal distribution of molecular weights as well, and the range and expected values are roughly similar to those of the sweet, bitter, and umami tastants. Figure 2 (b) shows the density distribution of the octanol-water partition coefficient logP, a measure of hydrophobicity. We find a good mix of hydrophilic (logP < 0) and hydrophobic (logP > 0) molecules in the sweet, bitter, and control categories, while most umami molecules are hydrophilic. Di Pizio et al. reported in a study with limited datapoints (677 bitter and 312 sweet) that bitter compounds have higher hydrophobicity than sweet ones, while sweet compounds have a wider size range [40]. However, our work, based on a superset of their data, reveals that such conclusions may not be drawn with certainty. The distributions of the two key properties, molecular weight and logP, for all the classes, demonstrate that the control set is chemically similar to the other tastants, and hence, the classification problem is well formulated. Figures 2 (c), (d), and (e) highlight the hydrogen bond-forming tendencies of the three classes of tastants. Hydrogen bond stabilizes protein-ligand complexes and thus crucially affects the binding affinity. In agreement with our expectations, most tastant molecules have hydrogen bond donors and acceptors, which enable them to interact with the residues in the binding pockets of the receptors.

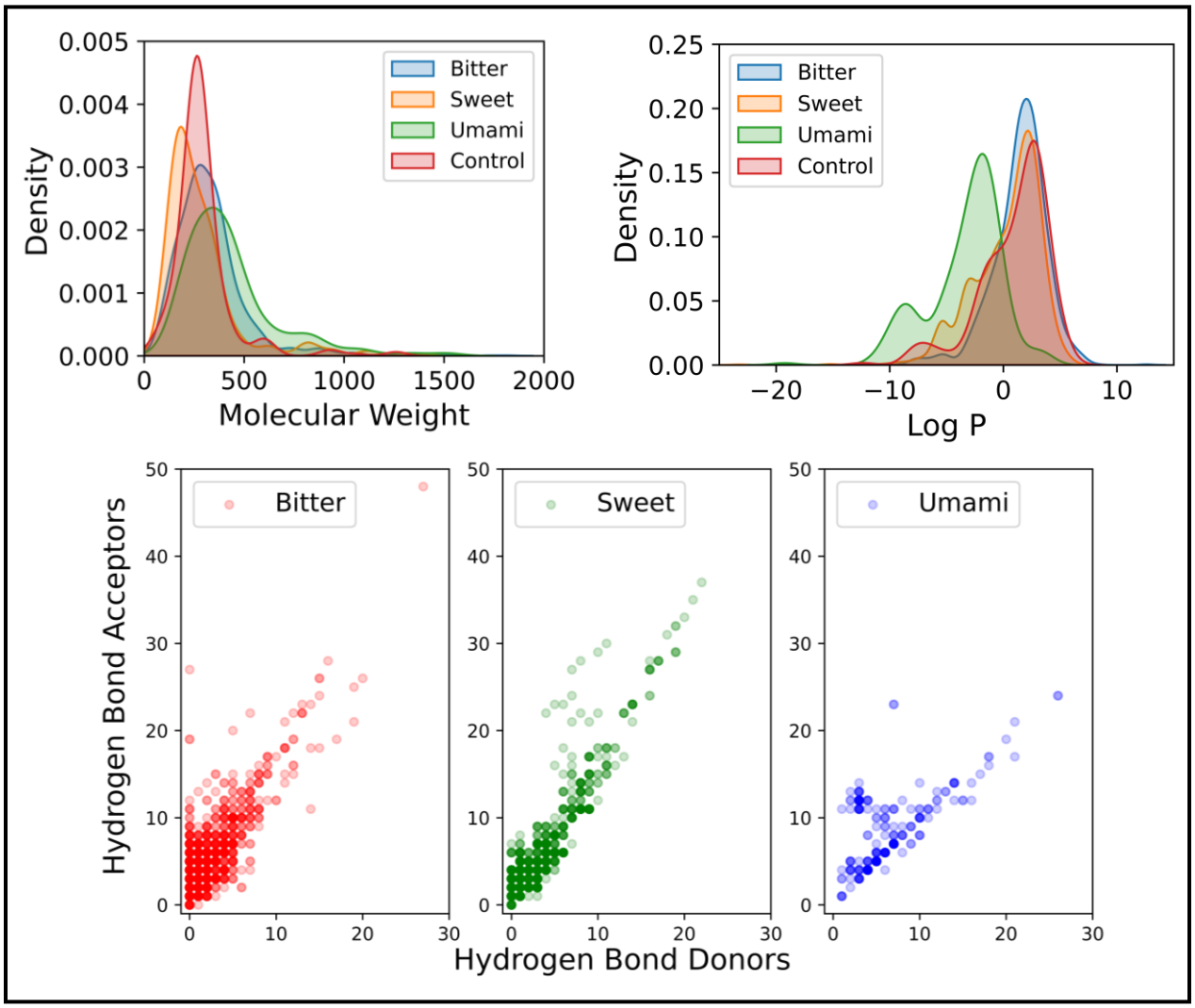

Figure 2: Dataset characteristics based on key molecular properties. Density distribution of (a) molecular weight and (b) octanol-water partition coeffcient logP of bitter, sweet, and umami compounds in the dataset. Hydrogen bond donors and acceptors in the (c) bitter, (d) sweet, and (e) umami molecules. A single point in (c), (d), and (e) can be an overlap of multiple tastants – transperancy rendering has been used to show the extent of overlap.

To further explore the dataset visually, we performed principal component analysis (PCA) to reduce the high dimensional data. PCA is an unsupervised learning technique that transforms a large set of correlated variables into a smaller set of uncorrelated variables, while maintaining the variation of the original dataset. Figure 3 (a) shows the relationship between the first and

second principal components. As evident from the PCA plot, the tastants have ample structural diversity. But the overlap between the three taste qualities is significant and no trivial way to separate them is apparent. Furthermore, the control set also overlaps with all the three taste qualities. Thus, the multiclass classification task is highly complex, and needs non-linear functions to distinguish between the molecules of different tastes, making neural networks the suitable choice for model building. The first principal component explains 22 % of the total variance, and the second one explains around 14 %. We also generated t-distributed stochastic neighbor embedding (t-SNE) plots for our dataset, as shown in Figure 3 (b). t-SNE is also an unsupervised dimensionality reduction technique, more powerful than PCA for visualizing complex data in two-dimensional space [41]. It minimizes the divergence between the distribution that measures pairwise similarities of input objects and the distribution that measures pairwise similarities of the corresponding embeddings. We assessed the performance of the algorithm for different values of the two key hyperparameters, perplexity and learning rate, by visually comparing the generated plots after optimizing for 2000 iterations. A perplexity of 100 and a learning rate of 800 was found to be suitable. t-SNE concurs with PCA regarding the structural diversity of the tastants in the dataset and the difficulty of the classification task. Although small clusters of similar tasting compounds can be observed, the overlap between the classes is high. No clear distinction between the taste qualities is apparent from the t-SNE visualization. It is to be noted that cluster size and distance between clusters in t-SNE plots bear no significance.

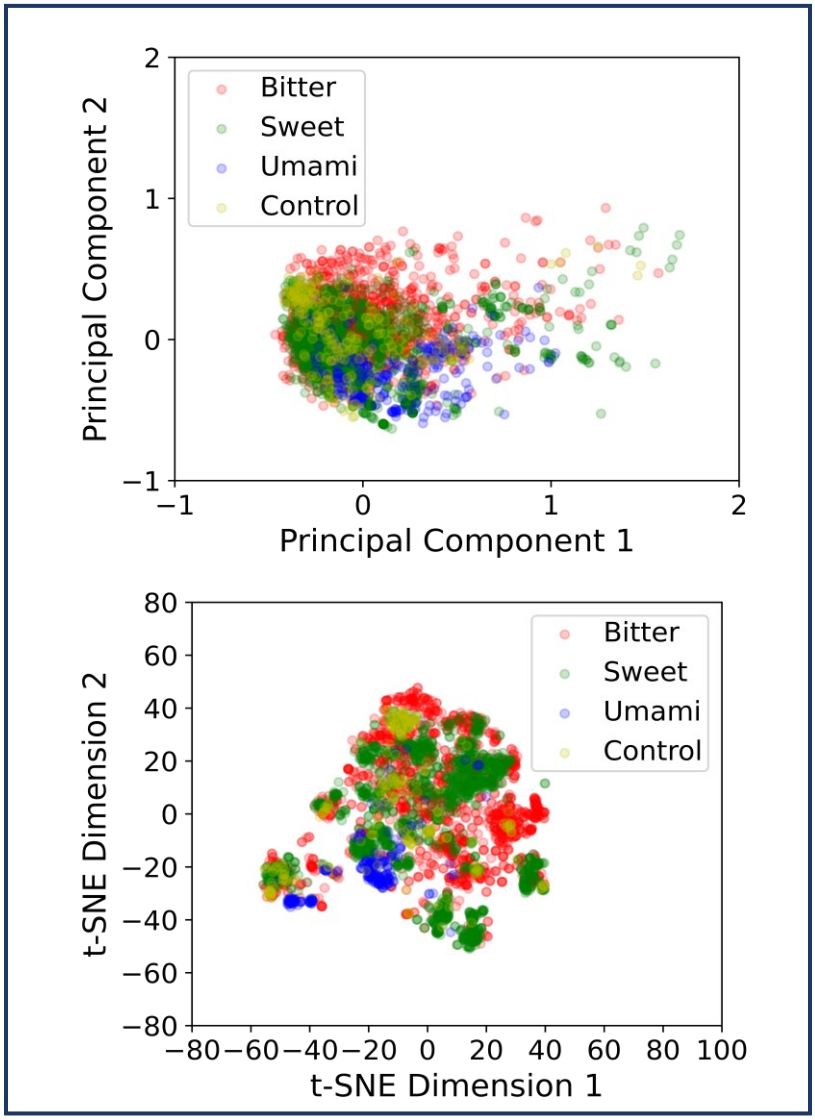

Figure 3: Unsupervised dimensionality reduction using (a) PCA and (b) t-SNE techniques. Molecules are colored according to their taste. Transperency rendering is applied to show overlap.

All results from our exploratory data analysis confirm the huge diversity of molecules, mostly within the small molecule chemical space with a few exceptions. Distributions of important molecular properties, as well as low-dimensional representations of the dataset, point to significant overlap between the three taste qualities, as well as the set of molecules used as

control. The structural similarities between many sweet and bitter compounds have been long known in the scientific community, with multiple cases of taste alteration on slight modifications in the structure [42]. Our work demonstrates this complication using data analytics and adds an added layer of complexity by including umami tastants within its scope.

3.2 Functional Group Analysis

To look deeper into the chemistry of tastants, we analyzed the functional groups present in the bitter, sweet, and umami compounds in our database. In-built modules in RDKit can compute the frequency of a predefined list of 85 substructures in a molecule, and we evaluated only those fragments for our analysis. Figure 4 shows the most common functional groups in the three classes of tastants along with their frequencies of occurrence. Carbonyl oxygen, which can belong to aldehydes, ketones, carboxylic acids, esters, amides, and other functional groups, is the most common fragment in both bitter (62.8 %) and sweet (65.1 %) molecules, and present in a majority (80 %) of umami compounds. In Figure 4, we consider only non-overlapping groups, and hence carbonyl is not shown— aldehyde, ketones, and other unique groups are treated as separate entities. It is visible from our analysis that many functional groups occur often in both bitter and sweet molecules, which agrees with our earlier discussion on the structural similarities between the two tastes. Apart from benzene, ether, aliphatic hydroxyl, and secondary amine, which are among the five most frequently occurring groups in both sweet and bitter compounds as shown in Figure 4, we also find carboxylic acid (12.4 % bitter and 29.2 % sweet), methoxy (16.8 % bitter and 17 % sweet), primary amine (9.1 % bitter 17 % sweet), tertiary amine (31.7 % bitter and 15.4 % sweet), and ester (23.9 % bitter and 32.5

% sweet) in many molecules of the two classes. About 45.8 % bitter and 15.3 % sweet compounds have bicyclic rings in their structure.

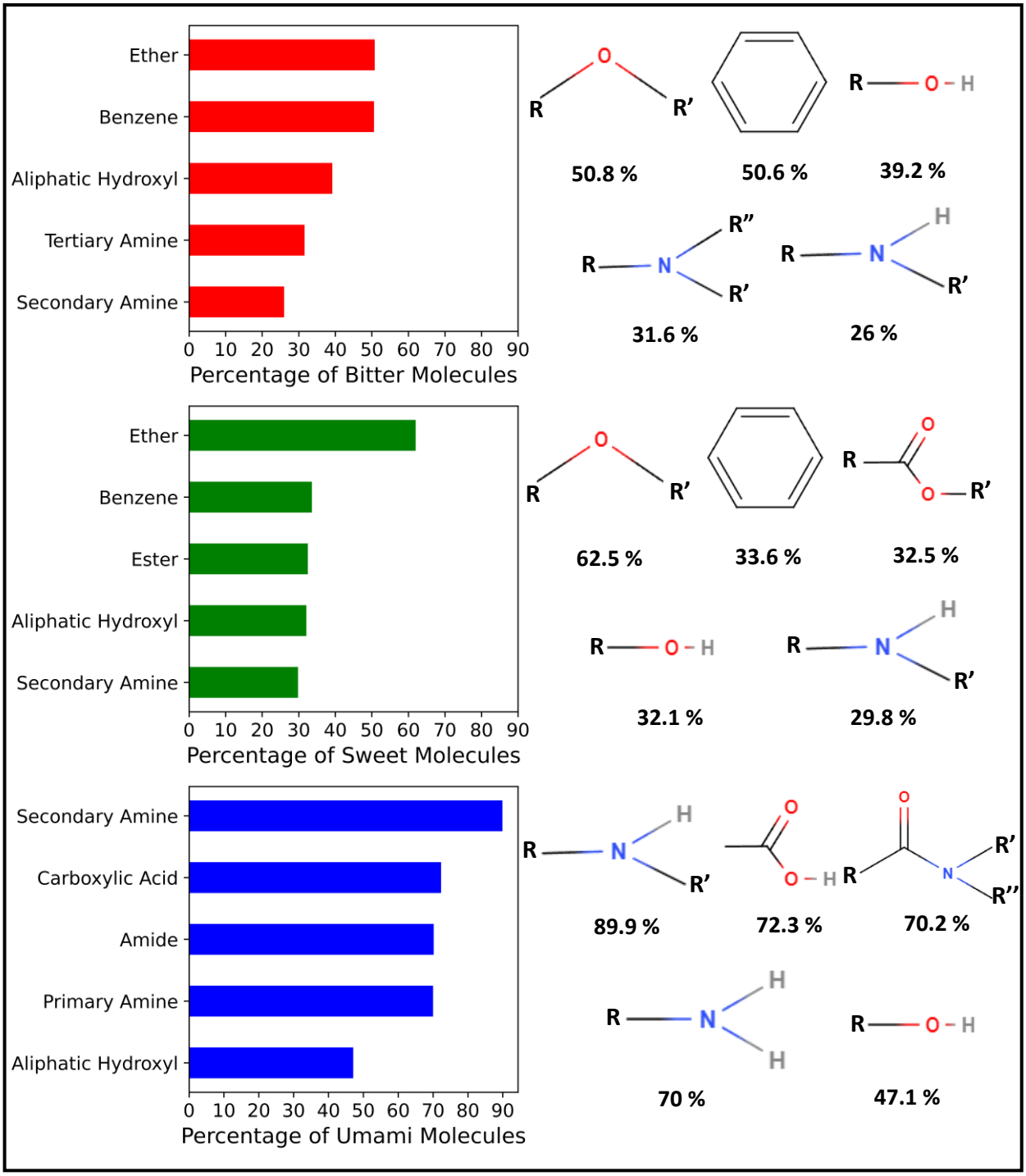

Figure 4: Most common functional groups in bitter, sweet, and umami compounds. The bar plots denote the percentage of compounds in each of the taste classes having the functional groups. The chemical structure of the functional groups and their exact percentage are provided beside each of the bar plots. Red, green, and blue in the bar plots correspond to bitter, sweet, and umami tastants, respectively.

Umami molecules are rich in nitrogen containing functional groups, as evident from Figure 4. Primary, secondary, and tertiary amines are present in 70 %, 89.9 %, and 37 % of umami compounds, respectively, the former two being among the five most common groups. Amide (70.2 %) and imidazole (25.63 %) exist widely as well. Imidazole and phosphate ester (23.9 %) strikingly distinguish umami as these two groups are scarcely found in sweet (0.3 % contain imidazole, 0.06 % contain phosphate ester) and bitter (1.9 % contain imidazole, 0.07 % contain phosphate ester) compounds. Like sweet and bitter, ether (27.3 %), carboxylic acid (72.3 %), and aliphatic hydroxyl (47 %) can be found frequently in umami molecules too. Sulfide group occur in 8.8 % of umami compounds, compared to 1.8 % and 3.6 % for bitter and sweet, respectively. As about 60 % of the umami compounds are UMP442 peptides, we observe such a high proportion of amines and amides. Salts of reactive metals like calcium, potassium, sodium, and magnesium make up slightly more than half of the non-peptide umami molecules in our database, with disodium salts being the most common.

Only five fragments, among the 85 calculated by RDKit, are not present in any of the molecules in our entire database. The bitter class shows the greatest diversity in terms of groups present in at least one compound. Even rare functional groups (present in less than 5 % of tastants for all three classes) occur more often in bitter molecules than in sweet or umami. This diversity can be attributed to the 25 T2Rs that can bind to a wide variety of ligands and elicit a taste

sensation. In contrast, we observe the least number of functional groups in umami molecules, which can possibly be the cause or consequence of the smaller set of known umami-tasting compounds. Overall, the functional group analysis sheds light on the chemical make-up of tastants and corroborates the inferences from exploratory data analysis regarding the complexity of the classification problem.

3.3 DNN Model Performance

The predicted taste classes from the DNN model were compared to the actual taste labels in our dataset to evaluate its performance. Figure 5 shows the confusion matrices of the DNN model with and without SMOTE for both the training and test sets. The hold-out test set contains 216 bitter, 270 sweet, 38 umami, and 32 control molecules, which roughly preserves the classwise distribution of tastants in the entire dataset. The overall prediction accuracies of the DNN model without SMOTE are 0.88 and 0.81 for the training and test sets, respectively. On applying SMOTE, the corresponding accuracies rise to 0.93 and 0.83. Although the values indicate some degree of overfitting in the model, it occurs principally due to the control set predictions. As the control set is a mixture of diverse molecules which do not belong to any particular category, the model tries to learn complex commonalities where there are none. In other words, the training and test are chemically dissimilar and the model just trains on noise. Only considering bitter, sweet, and umami molecules, the test set accuracies are 0.84 and 0.86 without and with SMOTE, respectively. An interesting observation is that the model rarely mislabels sweet or bitter compounds as umami. While these results are satisfactory, accuracy is

not always the appropriate measure of model performance in classification problems, especially with imbalanced datasets. We calculate more insightful metrics— precision, recall, and F1 score, for the three taste classes, as shown in Table 1. Precision is defined as the ratio of true positives to total predicted positives, while recall is the ratio of true positives to total actual positives. F1 score, which is the harmonic mean of precision and recall, strikes a balance between the two, and takes uneven class distribution into account. For multiclass classification tasks, especially with class imbalance, classwise F1 score, and confusion matrix gives a complete description of the prediction performance. We see that the F1 scores of all three classes improve on applying SMOTE. The precision and recall values are consequently higher as well. In particular, the umami F1 score changes more than the other two tastes, which was the main motivation for using SMOTE to resample the dataset.

As evident from Table 1, the model greatly overfits on the control set due to training on noise as described earlier, which is an unavoidable outcome of multiclass classification problems using ML. But using a control set for prediction problems like ours is indispensable, since the three categories of concern are not exhaustive. But, for practical applications, great model performance for the control set is not required. We desire a model that can reliably identify bitter, sweet, and umami tastants from large molecule databases of food and flavor compounds, natural products, and drug-like compounds. Figure 5 and Table 1 demonstrate that the DNN model can perform that task with high accuracy and precision. The confusion matrices also show us that many sweet molecules are mislabeled as bitter and vice versa. The model is not able to entirely resolve the complexity of classification due to the structural similarity between sweet and bitter compounds. Although direct comparisons with existing literature is

not possible due to the novelty of the multiclass problem and disparity between datasets, the accuracies and F1 scores are comparable to those obtained in the simpler sweet/non-sweet and bitter/non-bitter classification models, many of which further suffer from additional limitations like small dataset size, random negative set, unverified taste information, and lack of chemical diversity [11, 12, 13, 16]. Hence, our results indicate that neural networks can tackle complex classification problems in the biochemical domain by learning representations, provided that properly curated datasets are available.

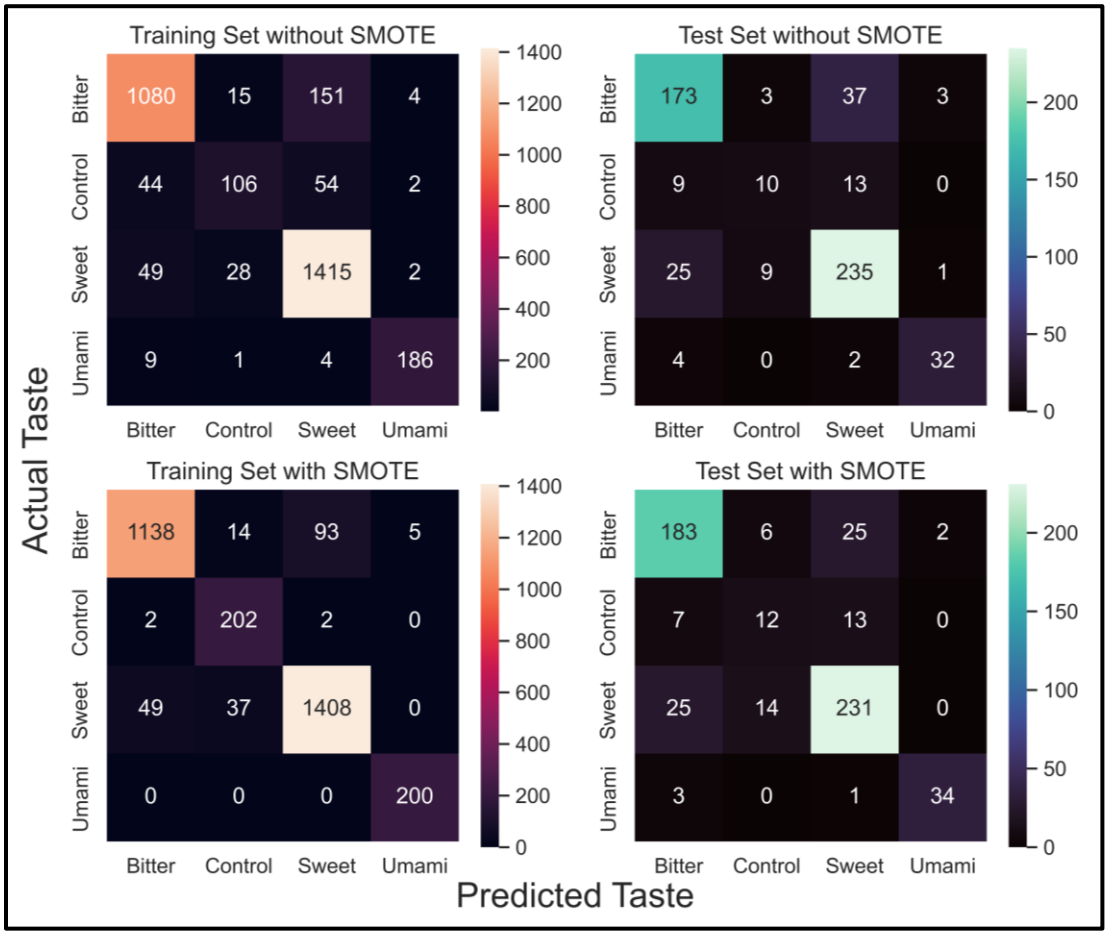

Figure 5: Confusion matrices based on the predictions of the DNN model with and without SMOTE for training and test sets. The numbers in each cell denote the absolute number of datapoints that satisfy the condition of the cell.

Table 1: Classwise precision, recall, and F1 scores of the training and test sets for the DNN model with and without SMOTE

Sampling	Dataset	Metric	Taste Class			
			Bitter	Sweet	Umami	Control
Without SMOTE	Training	Precision	0.91	0.87	0.95	0.71
		Recall	0.87	0.94	0.93	0.51
		F1 Score	0.89	0.91	0.94	0.60
	Test	Precision	0.82	0.82	0.89	0.45
		Recall	0.81	0.87	0.84	0.31
		F1 Score	0.81	0.84	0.86	0.37
With SMOTE	Training	Precision	0.95	0.93	0.97	0.80
		Recall	0.91	0.94	1.0	0.98
		F1 Score	0.93	0.94	0.99	0.88
	Test	Precision	0.84	0.86	0.94	0.38
		Recall	0.85	0.86	0.89	0.38
		F1 Score	0.85	0.86	0.92	0.38

3.4 Explaining DNN Predictions using SHAP

Despite exceptional predicting capabilities, neural networks are infamous for their lack of interpretability and explainability. The two terms are often used interchangeably, but a subtle yet crucial difference exists. Interpretability is the extent to which a cause-and-effect relationship can be determined to consistently predict how the output changes given a change in input. High interpretability often comes at the cost of performance, as it is difficult to establish cause-effect relationships beyond simple ML models like linear regression and decision trees. For the DNN model, we are concerned with explainability, which aims to understand the behavior of ML algorithms in human terms. Much effort in ML research has

been directed towards explainability in recent years, and we chose the Shapley additive explanations (SHAP) technique because of its unified framework for interpreting predictions [43]. SHAP uses a cooperative game-theoretic approach to compute the contribution of each feature towards the prediction and provides Shapley values as output. Specifically, we leverage the DeepExplainer method, which is suitable for neural networks. Figure 6 shows the average of absolute SHAP values over the entire test set for the ten key features based on the prediction of the DNN model with SMOTE. The features are ranked according to their overall relative importance, as explained by SHAP. The average SHAP values are computed separately for each taste class and control and plotted together by stacking. Detailed visualization for the three taste classes and individual SHAP values corresponding to each test example and each feature is shown in supporting information.

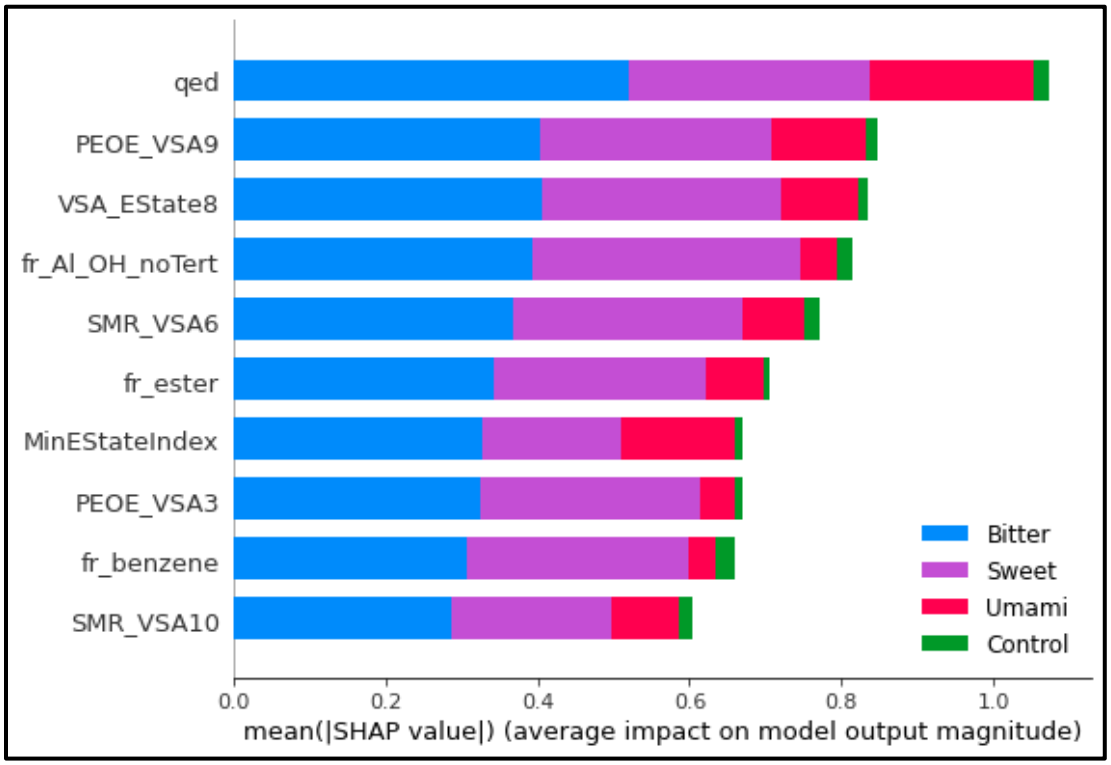

Figure 6: Bar plots of the mean absolute SHAP values for the test set of the ten important features, ranked in order of their relative importance, based on the predictions of the DNN model with SMOTE. Stacking is used to visualize the SHAP values of all three taste classes and control in a single chart.

Comparing the 20 main features (only ten are shown in Figure 6) of the DNN model with and without SMOTE (not shown in the manuscript), 13 are common to both, although their relative importance changes. Resampling with SMOTE helps the model to learn the relevant information to adequately represent all the classes. We observe that electrostatic (PEOE_VSAs), polarizability (SMR_VSAs), and electro-topological (EState_VSAs/VSAEStates) properties, along with drug-likeness and counts of fragments like benzene, aliphatic hydroxyl, and ester, are the crucial features that determine the taste of molecules. The SHAP explanation agrees with our current understanding of the physics of receptor-ligand interactions, where these properties (electrostatic, polarizability, etc.) determine the affinity of molecules towards a binding pocket. The DNN model discovers this physics without being explicitly programmed and makes predictions accordingly. However, we should avoid inferring a strong association between the binding physics and model mechanism because many of the initial features generated by RDKit were dropped using correlation analysis. Some properties like hydrophobicity/hydrophilicity (SlogP_VSAs) and Lipinski parameters (HeavyAtomCount, NumHDonors, NumHAcceptors, NumRotatableBonds, etc.) play an important role in receptor-ligand binding but were mostly removed randomly due to strong correlation with other features. SHAP analysis only explains how the model makes the decisions, which provides some physical insights and explainability of black-box neural networks, but cannot be interpreted as hard truth of the physical phenomena

itself. Our work demonstrates the capability of neural networks in learning complex structure-property relationships in molecules, while still being explainable to some extent with assistance from techniques like SHAP. Incorporating explainability within the realm of DL, especially in biochemical applications, can play a paramount role in bolstering its acceptance in industrial settings and bridging the gap with physics-based theoretical understanding of various phenomena.

3.5 GNN Model Performance

Similar to the DNN model, we compared the predictions of the GNN model with the actual taste labels and computed the performance metrics. Figure 7 shows the confusion matrices for the GNN model with and without oversampling. The overall prediction accuracies without oversampling are 0.91 and 0.80 for the training and test sets, respectively, which are comparable to the DNN model accuracies. Like DNN, the GNN model also mislabels many sweet molecules as bitter and vice versa. We experimented with different dropout probabilities and found 0.1 for all layers to be an optimum value that reduced the extent of overfitting without compromising on the test set performance. SMILES enumeration for handling class imbalance is especially helpful to create valid synthetic data that is different from the original data, when string-based featurization is used for building ML models. However, for graph-based featurization, different SMILES for the same molecule lead to node feature and adjacency matrices with different row ordering, which ultimately generates the same embedding as permutation invariancy is a precondition of all GNNs. But oversampling of the minority classes is also a valid resampling technique for dealing with class imbalance. The prediction accuracies with oversampling are 0.91 and 0.81 for the training and test set, respectively.

Like the DNN model, the GNN model too suffers from an extent of overfitting due to the control set for reasons discussed in section 3.3. Only considering bitter, sweet, and umami molecules, the test set accuracies without and with oversampling are 0.82 and 0.84, respectively. Table 2 shows the classwise precision, recall, and F1 score of the GNN model for the training and test sets. We observe that most metrics are comparable or slightly inferior to those of the DNN model. Oversampling slightly improves the predictions of the three taste classes during evaluation of the test set. Hence, the performances of both of our DL models are roughly similar on all metrics.

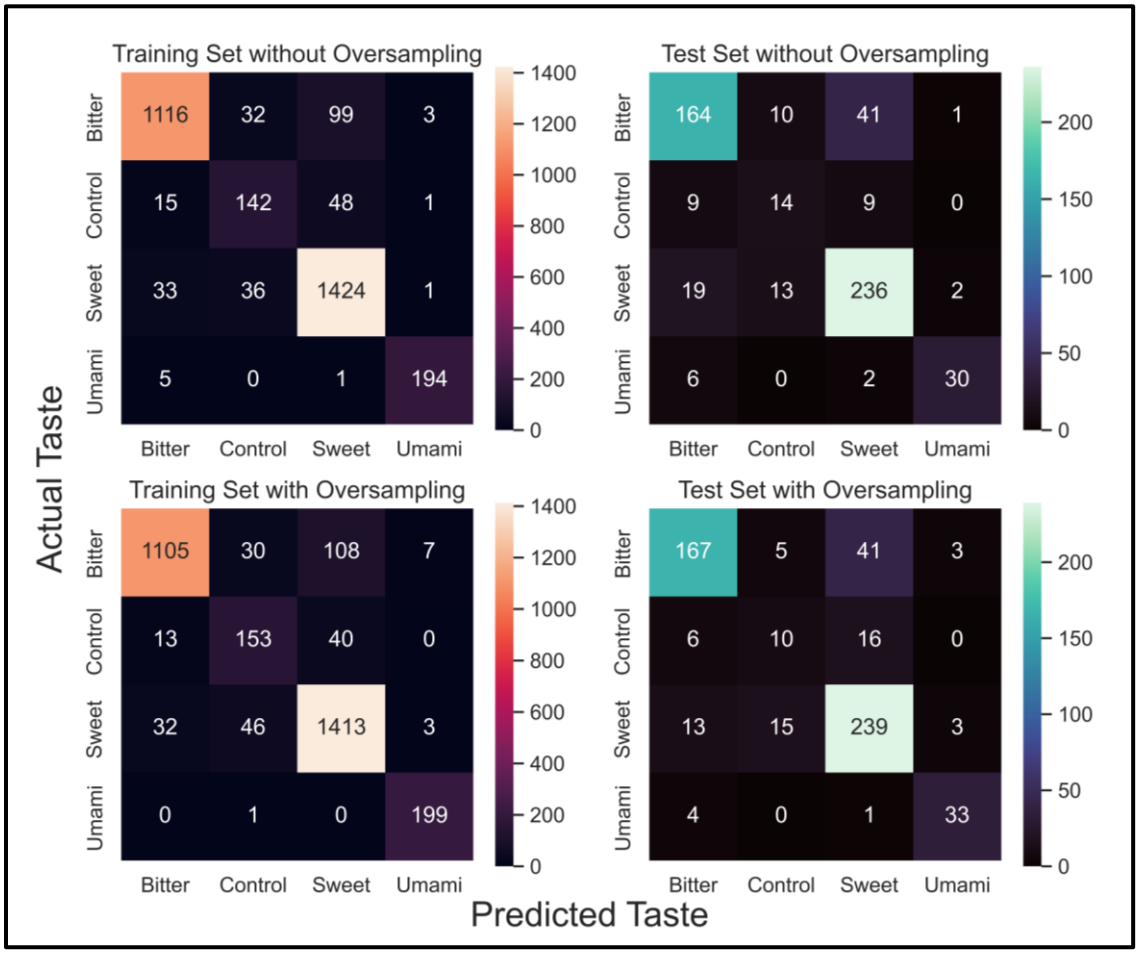

Figure 7: Confusion matrices based on the predictions of the GNN model with and without oversampling using SMILES enumeration for training and test sets. The numbers in each cell denote the absolute number of datapoints that satisfy the condition of the cell.

Table 2: Classwise precision, recall, and F1 scores of the training and test sets for the GNN model with and without oversampling

Sampling	Dataset	Metric	Taste Class			
			Bitter	Sweet	Umami	Control
Without Oversampling	Training	Precision	0.95	0.90	0.97	0.68
		Recall	0.89	0.95	0.97	0.69
		F1 Score	0.92	0.93	0.97	0.68
	Test	Precision	0.83	0.82	0.91	0.38
		Recall	0.76	0.87	0.79	0.44
		F1 Score	0.79	0.85	0.85	0.41
With Oversampling	Training	Precision	0.96	0.91	0.95	0.67
		Recall	0.88	0.95	0.99	0.74
		F1 Score	0.92	0.93	0.97	0.70
	Test	Precision	0.88	0.81	0.85	0.33
		Recall	0.77	0.89	0.87	0.31
		F1 Score	0.82	0.85	0.86	0.32

Although graph-based DL techniques have proved to be powerful tools for various molecular property prediction tasks, they often require large datasets to surpass the performance of traditional ML models. Otherwise, the likelihood of overfitting and poor test set performance is significant. Our tastant dataset only consists of a few thousand datapoints, which is typically not enough for GNNs. But we show that the graph convolutional approach proposed by Duvenaud et al. performs satisfactorily on the dataset and classifies bitter, sweet, and umami

molecules with high accuracy, precision, and recall. As discussed earlier, GNNs have the additional advantage of learning directly from the molecular structure and the associated chemical information like atom type, valance, aromaticity, etc., without the need for software-based featurization and expert feature cleaning. The number of descriptors, their values and calculation methodology, as well as the range of information can vary widely among the various commercial and free cheminformatics tools. Hence, if better or comparable performance is achieved, it is desirable to use GNNs in automated workflows over DNNs or traditional ML models. When larger datasets are obtained in the future from experiments, simulations, and data curations, GNNs are likely to outperform other alternatives. Our work demonstrates that even within the limitations of the current state of data availability, GNNs can be an attractive choice for screening and classifying tastants.

3.6 Applications of the DL Models

The principal aim of building ML models for molecules is to deploy them for practical applications for improving screening efficiency and reducing cost. The two deep learning-based classification models developed in this paper can be applied for screening bitter, sweet, and umami tastants. Existing works on data-driven taste predictors, as discussed in the introduction, can mostly handle one taste at a time. While higher accuracies can be obtained for binary classification problems, an array of different models is necessary to treat molecules of different taste qualities, which are trained on different datasets, have different applicability domains, and often at conflict with each other. For industrial use cases, which often deal with large

datasets, the ability to simultaneously identify multiple classes of tastants, albeit with slightly lower accuracy, is desired over highly accurate classifiers that identify molecules of a single taste only. Experimental validation is essential before manufacturing a novel molecule or formulation, and hence minor differences in the accuracies of ML models used for initial screening do not have any practical implications.

During the prediction phase of ML models, defining an applicability domain (AD) is important to selectively remove molecules whose structures are significantly dissimilar from the ones in the training set. The Organization of Economic Cooperation and Development has set guidelines for structure-activity relationship models, which we followed for our application. We employed two metrics – Euclidean distance and Tanimoto similarity coefficient to define the AD. If the median Euclidean distance in the feature space of a new molecule from its five nearest neighbors in the training set is greater than 1.5, then it is considered lying outside the AD of the models. The distance is calculated using all the 41 features used for building the DNN model, normalized by min-max scaling. Additionally, if the Tanimoto coefficient, computed using RDKit, of a new molecule with all the molecules in the training set is lesser than 0.6, then it is outside the AD. Both conditions are to be applied simultaneously to define the search space of the DL models for screening. It is to be noted that these thresholds and parameters are not strict and can be modified based on the end use and desired confidence.

Both the DNN model with SMOTE and the GNN model with oversampling were employed to screen the FooDB (https://foodb.ca/) database. FooDB contains around 70,000 food-related compounds, out of which 8957 were within our defined AD. The DNN model predicted 6700 molecules to be bitter, 1767 to be sweet, and 382 to be umami. The GNN model predicted 6492

molecules to be bitter, 2027 to be sweet, and 375 to be umami. A high number of FooDB entries are estimated to be bitter probably due to the flavonoids, alkaloids, and other bitter-tasting plant-derived compounds. But the proportion of bitter, sweet, and umami molecules as predicted by the models should not be considered as a reflection of the entire database. The number of compounds falling within our AD is about 13 % of the entire database, and hence the proportions cannot be generalized. However, based on requirements, the AD parameters can be made more lenient so that the models can be used to predict the taste of more molecules. For greater reliability of application, we prescribe taking the consensus of the DNN and GNN models to identify novel tastants. Comparing the predictions, we found that 6088 molecules were predicted to be bitter by both the models, 1452 to be sweet, and 288 to be umami.

4 Conclusions

In this paper, we present a data-driven approach for analysis and classification of tastants. Among the five basic tastes, bitter, sweet, and umami are sensed via GPCRs and hence are considered for a multiclass classification problem. We curated an extensive dataset of verified tastants from literature, which included a diverse class of molecules. The characteristics of the dataset including key chemical properties and functional groups are discussed in detail. Significant structural similarities between sweet and bitter molecules are observed from our data analysis, which revalidates the existing ideas on taste. To classify the molecules, we built and trained a descriptor-based DNN model and a graph-based GNN model. Both showed comparable performance in terms of multiple metrics. The GNN model has a notable advantage

of being able learn from the molecular structures without requiring handcrafted features and expert feature cleaning. As the number of umami molecules in the dataset is much lower compared to bitter and sweet, we applied special techniques to handle the class imbalance. Additionally, the SHAP method was utilized to explain the predictions of the DNN model. We observed that the neural network attributed more importance to features that correspond to physically relevant properties for molecule binding to a receptor. Finally, we defined an AD for model deployment in real-life scenarios and used both the models to simultaneously identify potential bitter, sweet, and umami compounds from the FooDB database.

Future directions in computational taste prediction include expanding the ideas we have established in this paper to all the five basic tastes, multitaste, and tastelessness. Achieving this would require curating datasets with a significant number of molecules belonging to all the different classes, as the performance of ML techniques strongly depends on the data. Currently, our models or, to the best of our knowledge, any other model in literature, cannot handle molecules exhibiting multiple taste qualities, and any such compound was not considered in our training and test sets. Estimating the relative taste intensities of molecules with respect to a reference is productive as well, but, except for sweetness, the progress is limited due to lack of experimental results. For industrial applications, simultaneously predicting one or more taste labels for a molecule, along with the taste intensity, will be of great economic advantage. Taste prediction can be further combined with cheminformatics approaches for oral bioavailability and toxicity analysis to screen databases for potential tastants with desired properties. Finally, an integrated computational framework can include a molecular docking or molecular dynamics module at the end of the pipeline to validate the screened molecules. This work contributes a

foundation to this framework and demonstrates that DL can be a powerful tool for food and flavor applications.

Acknowledgements

The authors would like to thank Mr. K Ananth Krishnan, CTO, Tata Consultancy Services and Dr. Gautam Shroff, Head of Research, Tata Consultancy Services, for their constant encouragement and support during this project.

Author Contributions

D.J. and R.G. conceived the project idea. P.D. and R.G. carried out the functional group analysis. P.D. and D.J. developed the deep learning models. All authors contributed to the interpretation and discussion of the results and preparation of this manuscript.

Conflict of Interest

The authors have no conflicts of interest to declare.

Supporting Information

Alternate train-test splits; Shapley additive explanations

References

[1] R. L. Doty and S. M. Bromley, "Taste," in *Encyclopedia of the Neurological Sciences*, Amsterdam, Elsevier, 2014, pp. 394-396.

[2] C. Spence, "Multisensory flavor perception," *Cell,* vol. 161, pp. 24-35, 2015.

[3] J. Diószegi, E. Llanaj and R. Ádány, "Genetic background of taste perception, taste preferences, and its nutritional implications: a systematic review," *Frontiers in Genetics,* vol. 10, p. 1272, 2019.

[4] S. Jeong and J. Lee, "Effects of cultural background on consumer perception and acceptability of foods and drinks: a review of latest cross-cultural studies," *Current Opinion in Food Science,* vol. 42, pp. 248-256, 2021.

[5] T. Hummel, B. N. Landis and K.-B. Hüttenbrink, "Smell and taste disorders," *GMS Current Topics in Otorhinolaryngology- Head and Neck Surgery,* vol. 10, 2011.

[6] B. P. Trivedi, "Gustatory system: The finer points of taste," *Nature,* vol. 486, pp. S2-S3, 2012.

[7] R. Ahmad and J. E. Dalziel, "G Protein-Coupled Receptors in Taste Physiology and Pharmacology," *Frontiers in Pharmacology,* vol. 11, 2020.

[8] G. Spaggiari, A. Di Pizio and P. Cozzini, "Sweet, umami and bitter taste receptors: State of the art of in silico molecular modeling approaches," *Trends in Food Science & Technology,* vol. 96, pp. 21-29, 2020.

[9] L. Pallante, M. Malavolta, G. Grasso, A. Korfiati, S. Mavroudi, B. Mavkov, A. Kalogeras, C. Alexakos, V. Martos, D. Amoroso, G. Di Benedetto, D. Piga, K. Theofilatos and M. Deriu, "On the human taste perception: Molecular-level understanding empowered by computational methods," *Trends in Food Science and Technology,* vol. 116, pp. 445-459, 2021.

[10] C. Rojas, R. Todeschini, D. Ballabio, A. Mauri, V. Consonni, P. Tripaldi and F. Grisoni, "A QSTR-Based Expert System to Predict Sweetness of Molecules," *Frontiers in Chemistry,* vol. 5, 2017.

[11] S. Zheng, W. Chang, W. Xu, Y. Xu and F. Lin, "e-Sweet: A Machine-Learning Based Platform for the Prediction of Sweetener and Its Relative Sweetness," *Frontiers in Chemistry,* vol. 7, 2019.

[12] W. Huang, Q. Shen, X. Su, M. Ji, X. Liu, Y. Chen, S. Lu, H. Zhuang and J. Zhang, "BitterX: a tool for understanding bitter taste in humans," *Scientific Reports,* vol. 6, 2016.

[13] A. Dagan-Wiener, I. Nissim, N. Ben Abu, G. Borgonovo, A. Bassoli and M. Y. Niv, "Bitter or not? BitterPredict, a tool for predicting taste from chemical structure," *Scientific Reports,* vol. 7, 2017.

[14] S. Zheng, M. Jiang, C. Zhao, R. Zhu, Z. Hu, Y. Xu and F. Lin, "e-Bitter: Bitterant Prediction by the Consensus Voting From the Machine-Learning Methods," *Frontiers in Chemistry,* vol. 6, 2018.

[15] P. Banerjee and R. Preissner, "BitterSweetForest: A Random Forest Based Binary Classifier to Predict Bitterness and Sweetness of Chemical Compounds," *Frontiers in Chemistry,* vol. 6, 2018.

[16] R. Tuwani, S. Wadhwa and G. Bagler, "BitterSweet: Building machine learning models for predicting the bitter and sweet taste of small molecules," *Scientific Reports,* vol. 9, 2019.

[17] G. Maroni, L. Pallante, G. Di Benedetto, M. A. Deriu, D. Piga and G. Grasso, "Informed classification of sweeteners/bitterants compounds via explainable machine learning," *Current Research in Food Science,* vol. 5, pp. 2270-2280, 2022.

[18] W. Bo, D. Qin, X. Zheng, Y. Wang, B. Ding, Y. Li and G. Liang, "Prediction of bitterant and sweetener using structure-taste relationship models based on an artificial neural network," *Food Research International,* vol. 153, 2022.

[19] J.-B. Chéron, I. Casciuc, J. Golebiowski, S. Antonczak and S. Fiorucci, "Sweetness prediction of natural compounds," *Food Chemistry,* vol. 221, pp. 1421-1425, 2017.

[20] A. Goel, K. Gajula, R. Gupta and B. Rai, "In-silico prediction of sweetness using structure-activity relationship models," *Food Chemistry,* vol. 253, pp. 127-131, 2018.

[21] M. Goel, A. Sharma, A. S. Chilwal, S. Kumari, A. Kumar and G. Bagler, "Machine learning models to predict sweetness of molecules," *Computers in Biology and Medicine,* vol. 152, 2023.

[22] J. Lee, S. B. Song, Y. K. Chung, J. H. Jang and J. Huh, "BoostSweet: Learning molecular perceptual representations of sweeteners," *Food Chemistry,* vol. 383, 2022.

[23] E. Margulis, A. Dagan-Wiener, R. S. Ives, S. Jaffari, K. Siems and M. Y. Niv, "Intense bitterness of molecules: Machine learning for expediting drug discovery," *Computational and Structural Biotechnology Journal,* vol. 19, pp. 568-576, 2021.

[24] G. De León, F. Eleonore, E. Fink, A. Di Pizio and S. Salar-Behzadi, "Premexotac: Machine learning bitterants predictor for advancing pharmaceutical development," *International Journal of Pharmaceutics,* vol. 628, 2022.

[25] P. Charoenkwan, J. Yana, N. Schaduangrat, C. Nantasenamat, M. M. Hasan and W. Shoombuatong, "iBitter-SCM: Identification and characterization of bitter peptides using a scoring card method with propensity scores of dipeptides," *Genomics,* vol. 112, pp. 2813-2822, 2020.

[26] P. Charoenkwan, C. Nantasenamat, M. M. Hasan, B. Manavalan and W. Shoombuatong, "BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides," *Bioinformatics,* vol. 37, pp. 2556-2562, 2021.

[27] P. Charoenkwan, J. Yana, C. Nantasenamat, M. M. Hasan and W. Shoombuatong, "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,* vol. 60, pp. 6666-6678, 2020.

[28] P. Charoenkwan, C. Nantasenamat, M. M. Hasan, M. A. Moni, B. Manavalan and W. Shoombuatong, "UMPred-FRL: A New Approach for Accurate Prediction of Umami Peptides Using Feature Representation Learning," *International Journal of Molecular Sciences,* vol. 22, 2021.

[29] L. Pallante, A. Korfiati, L. Androutsos, F. Stojceski, A. Bompotas, I. Giannikos, C. Raftopoulos, M. Malavolta, G. Grasso, S. Mavroudi, A. Kalogeras, V. Martos and D. Amoroso, "Toward a general and

interpretable umami taste predictor using a multi-objective machine learning approach," *Scientific Reports,* vol. 12, 2022.

[30] M. Malavolta, L. Pallante, B. Mavkov, F. Stojceski, G. Grasso, A. Korfiati, S. Mavroudi, A. Kalogeras, C. Alexakos, V. Martos, D. Amoroso, G. Di Benedetto and D. Piga, "A survey on computational taste predictors," *European Food Research and Technology,* 2022.

[31] A. Goel, K. Gajula, R. Gupta and B. Rai, "In-silico screening of database for finding potential sweet molecules: A combined data and structure based modeling approach," *Food Chemistry,* vol. 343, 2021.

[32] "Identification of novel umami molecules via QSAR models and molecular docking," *Food and Function,* vol. 13, pp. 7529-7539, 2022.

[33] W. P. Walters and R. Barzilay, "Applications of Deep Learning in Molecule Generation and Molecular Property Prediction," *Accounts of Chemical Research,* vol. 54, pp. 263-270, 2021.

[34] O. Wieder, S. Kohlbacher, M. Kuenemann, A. Garon, P. Ducrot, T. Seidel and T. Langer, "A compact review of molecular property prediction with graph neural networks," *Drug Discovery Today: Technologies,* vol. 37, pp. 1-12, 2020.

[35] D. Jiang, Z. Wu, C.-Y. Hsieh, G. Chen, B. Liao, Z. Wang, C. Shen, D. Cao, J. Wu and H. Tingjun, "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models," *Journal of Cheminformatics,* vol. 13, 2021.

[36] C. Rojas, D. Ballabio, K. P. Sarmiento, E. P. Jaramillo, M. Mendoza and F. García, "ChemTastesDB: A curated database of molecular tastants," *Food Chemistry: Molecular Sciences,* vol. 4, 2022.

[37] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik and R. P. Adams, "Convolutional Networks on Graphs for Learning Molecular Fingerprints," in *Advances in Neural Information Processing Systems*, 2015.

[38] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research,* vol. 16, pp. 321-357, 2002.

[39] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing and V. Pande, "MoleculeNet: a benchmark for molecular machine learning," *Chemical Science,* vol. 9, pp. 513-530, 2018.

[40] A. Di Pizio, Y. B. Shoshan-Galeczki, J. E. Hayes and M. Y. Niv, "Bitter and sweet tasting molecules: It's complicated," *Neuroscience Letters,* vol. 700, pp. 56-63, 2019.

[41] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research,* vol. 9, pp. 2579-2605, 2008.

[42] S. Damodaran and K. L. Parkin, Fennema's Food Chemistry, CRC Press, 2017.

[43] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017.