Exploring Machine Olfaction

An Undergraduate Honors Thesis

Author: Rui Nie
Supervisor: Prof. Ambuj Tewari

May 14, 2023

Abstract

The sense of smell evolved for animals to find food and mates as well as detect dangers. For the human experience, the sense of smell is fundamental to appreciating good food and avoiding eating rotten food. It is also critical in detecting dangers such as natural gas leaks (mercaptan is added to natural gas to give an otherwise odorless gas a characteristic odor). Uncovering the relationship between odor percepts and molecular structures can increase our knowledge of our surrounding environments as well as give us insights into the neuroscience of the human olfactory system at a low cost. However, the mismatch between the structural similarity of molecules and odor similarity complicates the olfaction problem: two molecules with similar structures can smell very different and vice versa. This study primarily exploited two psychophysical datasets in which pre-trained (Dravnieks [7]) and novice human subjects (Keller [15]) reported their subjective ratings towards a selection of odor descriptors respectively. We compared the performance of classical machine learning regression methods on pre-computed physicochemical descriptors with that of Graph Neural Networks (GNN) on raw molecular graphs. We found that GNN demonstrated a greater competency in learning psycho-chemical data obtained from trained panelists, while conventional machine learning techniques are more suitable in learning descriptor applicability generalized from untrained participants. We also implemented transfer learning and found that Dravnieks data that includes higher olfaction experience levels can improve the prediction of odor percepts of ordinary humans with little sensory training. Such capability of GNN to learn complicated general odor percepts by subjects across domain knowledge levels offers a new perspective on more efficient data utilization in the olfaction field.

1 Introduction

The olfactory sense, colloquially the sense of smell, is an indispensable part of human life. It plays an important role in health-related issues, such as affecting human appetite, food hygiene assessment, and identification of safety issues. Beyond survival problems, it is also associated with quality-of-life, by collaborating with taste sensation to produce flavors that enhance the consuming experience of foods and drinks. It also helps us appreciate fragrance and floral smell [39]. With odorous molecules encoded by a large family of odor receptors and synthesized to smell sensation in a "combinatorial" approach, in theory, humans are able to discriminate more than a million odorants even in minuscule quantities [32]. However, the synthetic nature of odor percepts and the fact that odorants are often parts of a mixture in practical life make it difficult to associate odorants with a specific

odor percept. For example, in coffee aroma, 2-furfurylthiol, methanethiol, and 3-mercapto-3-methyl butyl formate are three among the many volatile chemical components of coffee, whose descriptors in isolation are "roasted coffee", "rotten cabbage", and "catty", respectively [1]. Hence, to obtain a more precise picture of odorants' smell, a higher concentration mixed with odorless inert gases is preferred. On the other side, it's noteworthy that some volatile chemicals are toxic to inhale at a relatively high concentration, such as benzene from incomplete hydrocarbon combustion may cause headaches upon the smell [2]. As a consequence, it's inappropriate to neglect the potential harm of directly gathering odor descriptors from odorants. If the odor percepts for a chemical can be learned without needing to survey every existing volatile chemical through real-life experiments, there would be significant advancements in perfumery and human understanding of the surrounding environment.

A molecule is the smallest unit that preserves chemical properties [3]. The excitement mechanism of odor receptors by odor molecules [39] implies a potential relationship between odor and molecular structures. In the field of chemistry, a molecule consists of two or more atoms held together by chemical bonds. The choices of atom types, chemical bonds, and molecular chirality contribute to a massive and diverse set of existing molecules. Though it has been revealed that the presence of some functional groups could indicate the odor, such as ester is often associated with a fruity odor [26], exceptions are not rare in this field. In particular, small variations in the existence and positions of certain atom or function groups could exert a great influence on a molecule's perceived smell. On the other hand, molecules with significantly different molecular compositions or structures may generate a similar odor profile [29]. This further complicates the task of understanding human odor perception of certain molecules.

Anosmia, also known as partial or full loss of smell, has attracted increasing attention as its often reported in conjunction with symptoms of COVID-19 [12]. The surge of publications and citations after the year 2020 (Figure 1) reflects climbing research attention in studying olfaction through quantitative learning methods, where I included "Machine" to take into account that machine learning is one of the widely used and powerful tools for quantitative odor-structure research. The sense of smell can be expressed and communicated in multiple ways. Human-generated odor labels to characterize odor percept help simplify the olfaction prediction problem into a multi-label classification task, often combined with natural language processing [29, 10, 30]. To avoid the potential bias introduced by differences in participants' vocabulary and cultural backgrounds, physiological signals sometimes are considered a more objective representation of odor percepts. For example, calcium imaging recordings that indicate neural activities were used to model participants' odor percepts [27]. Another direct approach is through odor receptor response modeling is investigated in species for which receptor responses are easily accessible, such as Drosophila [8]. However, due to the fact that only around 10% of human odorant receptors have a published matching ligand [20], human odor receptor responses have largely remained unexplored. Other proposed measurements of odor percepts include pairwise perceptual similarity for odorants [34].

In this study, we focus on extending the broad machine learning efforts from the multi-label classification of odor labels to predicting the numerical descriptor applicability of odorants. Two psychophysical odor perception datasets [15, 7] that reflect different olfaction experience levels are targeted. I first modeled descriptor applicability from molecular representation with the help of machine learning regression methods and Graph Neural Networks for two prediction tasks. Then, we infer how informative the Graph Neural Network (GNN)-learned embeddings are by investigating their performance when feeding into classical machine learning models (Support Vector Machine, K-Nearest Neighbors, Random Forest, and Gradient Boosting). In the end, I carried out transfer

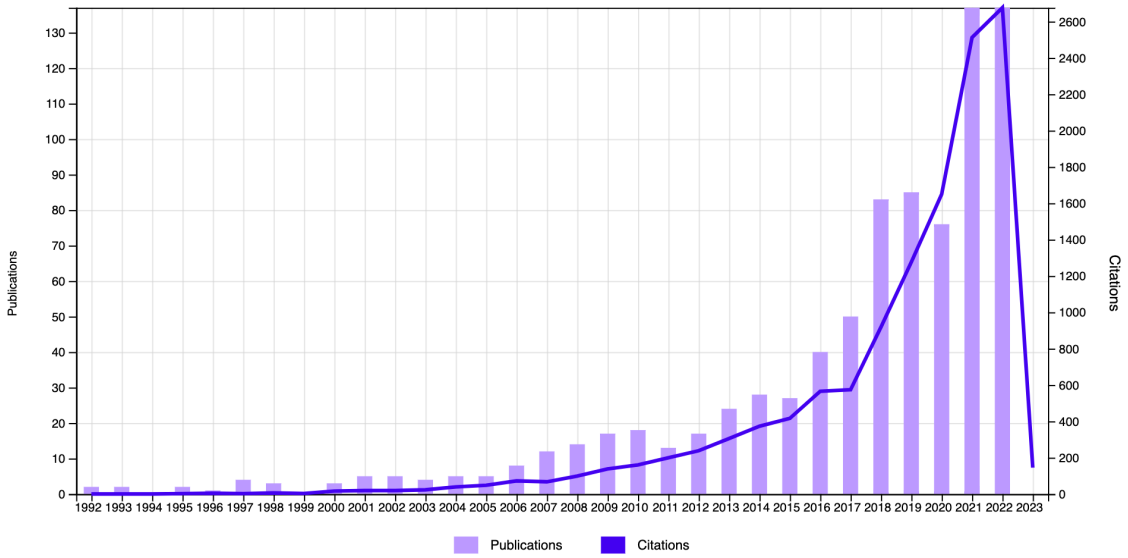

Figure 1: Publication and Citation Trend for "machine AND (olfaction OR smell)" in recent 30 years. Note that the data for the year 2023 is removed since only partial data is available.

learnings across two prediction tasks, in hope of providing a perspective into whether odor percepts data reflects higher olfaction experience levels can improve understanding of layman odor percepts and increase models predicting power.

2 Methods and Materials

2.1 Data overview

With the advancement in modeling techniques, there increases a demand for reliable benchmark psychophysical data to enhance the predictive ability and robustness of olfaction [9]. The high-quality human olfactory database is hard to establish as a result of the costly and laborious odorant preparation and participant recruitment. The conjecture that human odor perception space possesses more than thousands of dimensions [11] makes the olfaction modeling intimidating. In addition, data integration from multiple sources requires proper coordination among various experimental protocols and procedures, which even further complicates the situation. A current effort by *pyrfume* project [5] maintains an archive for cleaned human and animal olfactory perception data from many recently published research is one among all attempts to put available olfaction data for easier access.

Domain experience makes a difference in human smelling behavior [35], as a higher level of olfaction training tends to lead to a more accurate odor profile characterization. This study aims to understand how domain knowledge level affects modeling behavior towards odor descriptor applicability (section 2.1.1) and how could modeling techniques transfer olfaction knowledge across datasets to aid in understanding lay people's odor percepts. Specifically, there are two datasets of interest: *Atlas of odor profiles* by Dravnieks [7] (Dravniek dataset) from trained panelists, and *the olfactory perception dataset for chemically diverse molecules* collected by Keller's team [15] (Keller dataset)from

healthy untrained subjects. They all adopted a continuous representation of odor percepts for a fixed selection of odor descriptors. A non-zero score suggests the perceived applicability of a descriptor for an odorant, while its numerical value reflects the percentage of applicability, which contains level information regarding how much the perceived applicability is. Such a representation also tries to embody the high-dimension nature of hypothetical olfactory space to the best of current knowledge.

2.1.1 Odor Descriptor Applicability

The study of Dravnieks [7] defined the calculation of odor descriptor applicability (also known as percentage applicability, or PA) as an odor percept measurement that balances across raw individual ratings (Equation 1-3). The PA is specified by a descriptor `d` of interest for a particular chemical `C`. As Keller's research team publicly documented the first-hand subjective ratings for intensity, pleasantness, and a selection of 20 odor descriptors, we calculated descriptors' PA using the same formula for each chemical in the interest of consistency with Dravnieks' data. It is noteworthy that odorants were tested at two concentration levels in Keller's study. In order to differentiate odor profiles of different chemicals, I obtained PA from odor responses collected at a higher concentration.

$$\mathrm{PU}_C^{(d)} = \frac{\sum_{i=1}^{N} \mathbf{1}\{s_C^i(d) > 0\}}{N} \ldots \text{ Percentage of Usage} \tag{1}$$

$$\text{Score level}_C^{(d)} = \frac{\sum_{i=1}^{N} s_C^i(d)}{S \cdot N} \tag{2}$$

$$\mathrm{PA}_C^{(d)} = \sqrt{\mathrm{PU}_C^{(d)} \cdot \text{Score level}_C^{(d)}} \ldots \text{Percentage of Applicability} \tag{3}$$

Note: S is the full potential score of an odor descriptor to describe any chemical C. $s_C^i(d)$ is the score given by individual i to indicate to how much extent the odor descriptor d is applicable to describe the chemical C ($0 \leq s_C^i(d) \leq S$). N is the number of people who participated in describing the odor of C in the experiment.

2.2 Odor Descriptors Choice

The research team of Dravnieks and Keller adopted different sets and different numbers of odor descriptors (Figure 2). With conventional machine learning techniques requiring the agreement in the format and meaning of target values between training and testing data, our purpose of cross-implement trained models calls for the need of matching odor descriptor pairs between two datasets.

Multiple pairs of odor descriptors share a common descriptive word instead of exactly forming a one-to-one mapping relationship. For example, appropriate synthesization rules need to be developed before we can generalize descriptor applicability for `"OAK WOOD, COGNAC"`, `"WOODY, RESINOUS"`, `"CEDARWOOD"` from Dravnieks dataset into a single surrogate for `"WOOD"` used Keller's dataset. Considering that odor perception between trained and naive subjects is correlated to a certain degree [13], we computed the correlation of PA across descriptors used in two studies on 61 overlapped chemicals (Figure S1). I thresholded the overlapping odor descriptor pairs by 0.5 and obtained a highly-correlated one-to-one matching odor descriptor lists, as suggested in Table 1. Note that `BAKERY (FRESHBREAD)-BAKERY` pair was left out as a result of its negligible standard deviation in the Dravnieks dataset. In our final selection of descriptor pairs, the distribution of their *descriptor Applicability*(PA) (Figure S2) varies as a consequence of raters' domain knowledge as well as a different sampling of chemicals.

Figure 2: Wordcloud of odor descriptors used in Keller (left) and Dravnieks (right) tasks

Descriptors (Keller)	Descriptors (Dravnieks)	Correlation
BAKERY	BAKERY (FRESH BREAD)	0.6781
SWEET	SWEET	0.8291
GARLIC	GARLIC, ONION	0.6788
SWEATY	SWEATY	0.7317
DECAYED	PUTRID, FOUL, DECAYED	0.7818
GRASS	HERBAL, GREEN,CUTGRASS	0.5265

Table 1: Matching descriptor pairs with high correlation scores rounded into 4 digits. The final descriptor pairs are marked in the form of italics

2.3 Molecular Representation

Encoding irregular-shaped molecules into a machine-readable format is one of the cores of quantitative property-structure relationships for olfactory percepts [38]. Feature-based and computer-learned representations as two widely used molecular representations that will be adopted in this study.

2.3.1 Mordred descriptors

Mordred [24] is a free descriptor-calculation software application that computes more than 1826 molecular descriptors, including those implemented in RDKit and those describing chemicals' topological/geometrical indices. Since [42] has demonstrated the feasibility of feature-based methods combined with a survey of conventional machine learning classifiers in learning human "musk" odor percept on a limited size of molecules, we explored regression tasks for *descriptor applicability*(PA) through feature-based olfaction learning.

In Mordred molecular descriptors computation, missing value occurs when some molecular descriptors are incompatible with characterizing certain chemicals. We retained molecular descriptors with $\leq 10\%$ NA rate across all participating chemicals, which resulted in a set of 1128 descriptors. With the intention of improving models' predictive performance and avoiding overfit issues, we removed redundant information in Mordred descriptors. We first filtered out those molecular descriptors with exact 0 variances, that is, those features whose values are constant across all molecules within a dataset. Then, by thresholding colinearity to below 0.5, we obtained a final feature set consisting of 54 meaningful and weakly correlated features (Figure 3).

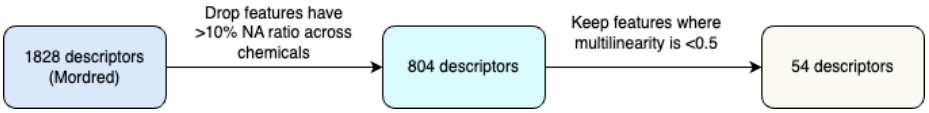

Figure 3: Feature selection process diagram

2.3.2 Computer-learned Representations

Representing chemicals' physicochemical attributes by a set of pre-computed molecular descriptors could potentially be either incomprehensive or biased. Hence, a majority part of this study is to investigate computers' capacity of learning odor percepts from raw molecular graph representations and their implication for addressing the limited data issues in olfaction fields.

A graph-based representation of molecules regards atoms as vertices and chemical bonds as edges that connect vertices. Atom-level features, edge indices, and edge attributes are encoded that help reconstruct the graph structure and store necessary structural information, including atom composition, aromatic, and hybridization information. An example of how Thioglycolic Acid's graph representation and structural formula are connected is illustrated in Figure 4 and 5. Once the training of neural network based on graphs is onset, GNN layers will learn embeddings from nodes and edges through optimization and message passing mechanism between parts of the graph. The intermediate embeddings are extracted as computer-learned molecular representations and serve as inputs to prediction networks. Further details about the GNN architecture will be discussed in section 2.4.2.

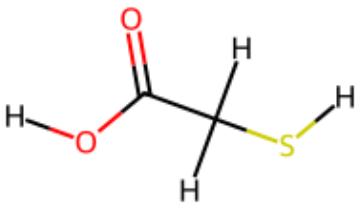

Figure 4: The structural formula of Thioglycolic Acid

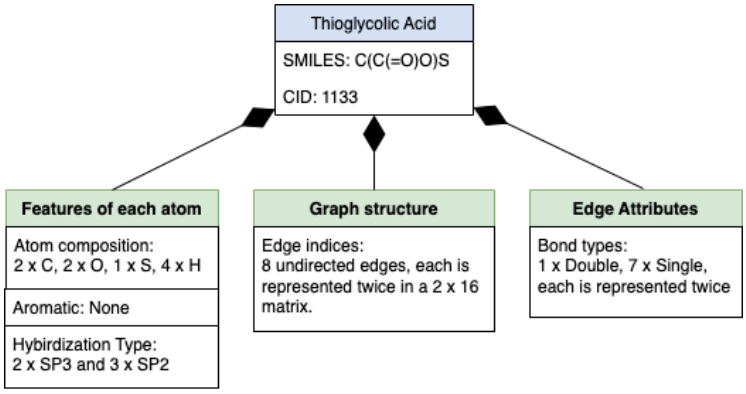

Figure 5: The computer-based representation of Thioglycolic Acid

2.4 Models

2.4.1 Machine Learning Regression Methods

Built upon the preprocessed Mordred features, we employed simple linear regression, support vector machine (SVM), random forest regression method (RF), gradient boosting (XGB) regression method, and k-neighbors for regression (KNN). Hyperparameters are determined by grid-search in a five-

fold cross-validation(Supplemental materials). Each selected learning method represents a specific category of statistical learning method.

2.4.2 Graph Neural Networks

The GNN has proven to handle graph data well in fields including quantitative structure-odor relationship [33], especially for odor label classification tasks. I proposed using a GNN (Table S2, Figure 6) that adopted Plain Generalized Graph Networks (GEN) [18] for graph convolution that updates using features both from nodes and edges at each layer. Messages of neighboring nodes and edges are aggregated in a permutation invariant manner, achieved by a differentiable and learnable `Softmax` message aggregation method. Furthermore, the aggregation effect is enhanced and improved through the implementation of a message normalization method, which combines messages with other features during the normalization. All learned representations from graph convolution will be downsampled in global-add-pooling and the reduced-size embeddings are used for classification or regression tasks through a final 2-layer fully connected neural network.

2.4.3 Adjust hyperparameters to prevent overfit

Both two aforementioned datasets maintain only hundreds of molecules, which is not traditionally considered a sufficient amount for model fitting. We tried to avoid the potential overfitting problem by adjusting machine learning regression methods' hyper-parameters, such as scaling down the number of trees trained in Random Forest and Gradient Boosting models. And for perceivably complex neural networks, I increase the dropout rate at each layer in prediction neural networks to 0.47 in an attempt to prevent overfitting. Additionally, previous work [42] on classifying "musk" label and ligand-receptor matching demonstrated the feasibility of machine learning models to learn on a relatively small sample (47 samples) on molecular fingerprints while still achieving satisfactory accuracy ($\geq 98\%$) on a larger size of unseen test data (224 odorants). This empirical evidence points to a promising potential for implementing machine learning for the human sense of smell even with very limited psychophysical data and becomes one of the important starting points of this project.

2.5 Measurement of fit

As a measurement for model performance, we are mostly interested in `explained_variance`, abbreviated as EV, complemented by `root mean squared error` (RMSE) and `correlation` with ground-truth value. EV calculates the ratio of explained variance by the model to the total variance in the target values. It has a similar formula as the coefficients of determination (R^2 score), while EV forces the score to take finite numbers below 1. An EV closer to 1 implies a better fit of the model, and smaller EVs are worse. RMSE estimates the closeness of predictions to their real value and correlation implies to what extent the relationship between predictions and true values can be generalized as a linear relationship.

2.6 Transfer Learning

Transfer learning brings solutions to a particular problem to solve a different but related problem [36]. In ideal scenarios, knowledge transfer occurs during the transfer learning, which consequently improves machine learning processes or grows the learning more robust. The transfer learning process and requirement for the task are dependent on the model architecture choice. As briefly mentioned in section 2.2, simple machine learning models take the same feature inputs and regress towards the same hard-coded target, that is, our matched odor descriptor pairs. For GNNs, however, they can

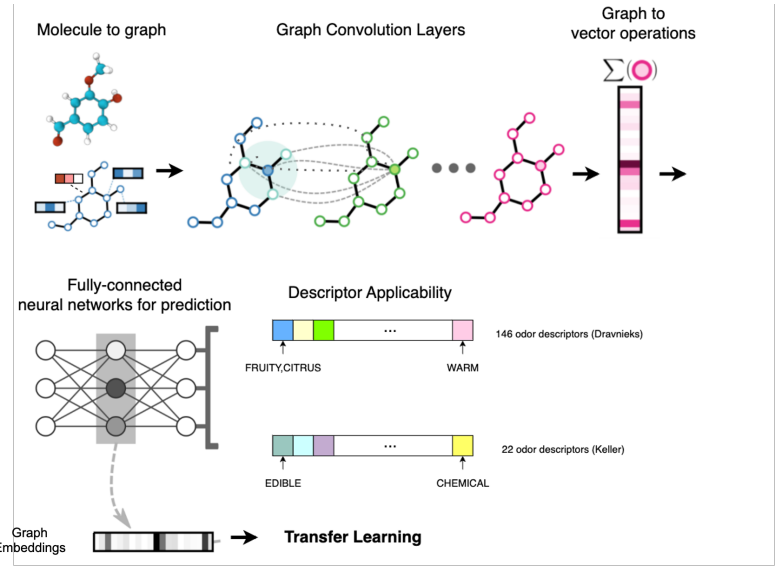

Figure 6: Graph Neural Network's schematic [30]. Each molecule is first transformed into a graph representation. The convolution layer extracts features from vertices and edges to form a graph vector, which is then used for regression through fully-connected neural networks. Embeddings were extracted before the implementation of the prediction network, as learned meaningful molecular representation for olfaction studies

be fine-tuned to regress towards entirely different target sets by operating on the molecular graph of the same representation structures.

3 Related Work

Deciding how algorithms see and process molecular information is the first step for statistical studies that use molecules as inputs. Physicochemical descriptors have been widely employed to represent molecules in machine learning studies. They are generated by computational software such as Dragon or Mordred, which encompass a wide range of topics, such as topological, geometrical, charges, edge adjacency indices, and so on [14, 8, 31, 29, 6]. Existing efforts from data scientists have demonstrated its power in odor percepts predicting tasks with an appropriate choice of algorithms [14]. Fingerprints calculated through RDKit python library are also treated as a valuable source for gathering molecular information for statistical learning [42, 43, 33, 29]. It is widely used as the baseline feature extraction method in olfaction research in comparison to computer-learned features, [30, 21, 41] and it collaborates well with neural networks [33]. On the other hand, given the extraordinary capabilities of deep neural networks for extracting features from massive and noisy data, increasingly more researchers started considering learning molecular features from raw molecular graphs [19, 28]. Other molecular representation methods within the machine have also been attempted, such as 2D images of odor molecules generated from structure diagrams [4, 33] and sensor signals from electronic nose [40, 37, 25].

What underlies the learning method choice is the pre-assumption of quantitative relationships be-

tween molecular structure and odor properties (QSOR) and the level of complexity for this task. Though differing in purposes, machine-learning approaches have been extensively experimented with to understand odor percepts from pre-computed physicochemical descriptors. SVM [31, 16, 23] and tree-based methods are two main widely-explored approaches in classifying human-generated odor descriptors [17, 29, 16], not only because of their desirable model performance but also for their proper level complexity.

Deep learning has been recently put forward as a powerful technique to decode QSOR. [33] proposed two pipelines using deep neural networks from molecular fingerprints and convolution neural networks based on images of chemicals for classifying human-generated odor labels. A similar task for classification was attempted in [30] by adopting Graph Neural Networks. With a focus on proposing novel architecture, it also tested the feasibility of transfer learning on a dataset for DREAM olfaction data challenge[14], where the winning model for individual odor percept prediction was fitted using random forest on Dragon physicochemical descriptors [22]. Other deep learning approaches obtain molecular representation through feeding odorous gases to stimulate electronic sensors [40, 37].

4 Results and Discussion

4.1 Descriptor applicability modeling

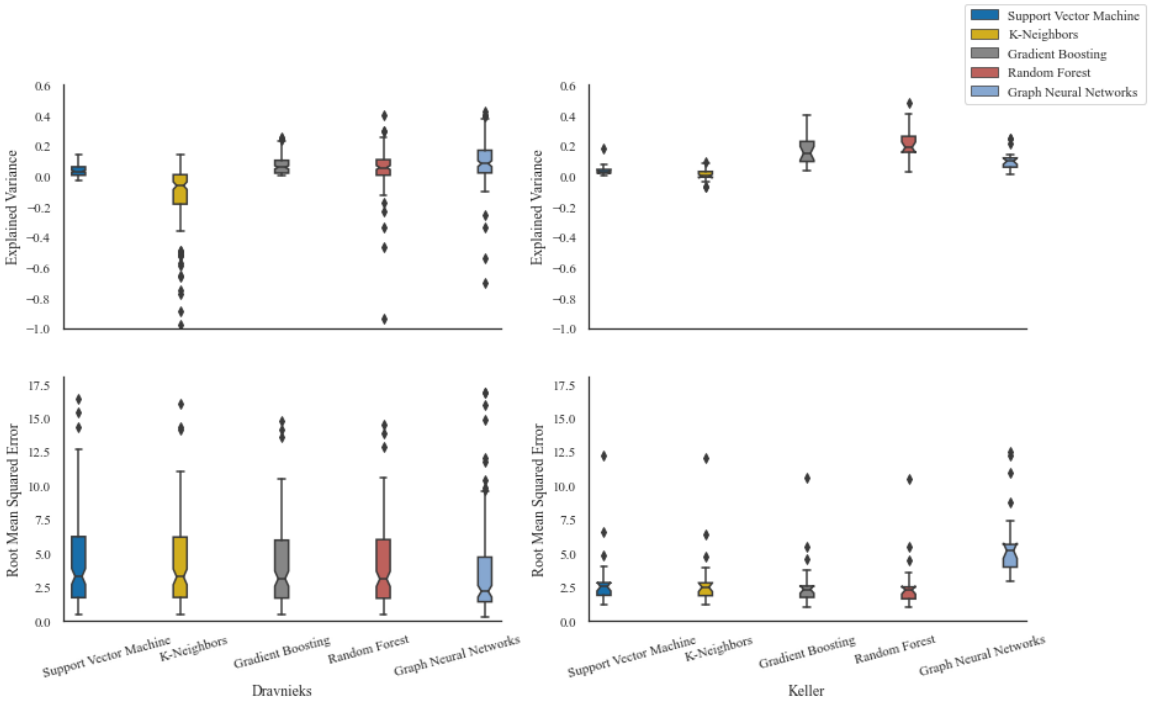

Figure 7: Model performance for each odor descriptor, as measured by explained variance and root mean squared error on the test set

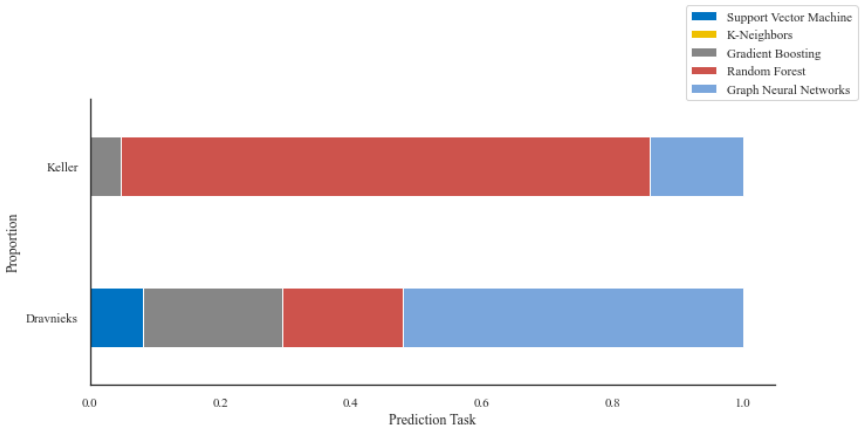

Figure 8: Proportion of odor descriptor-wise best-performing models measured by EV

Conventional machine learning on physicochemical descriptors has exhibited remarkable learning ability in modeling individuals' and population odor responses in DREAM Olfaction challenge [14] that uses Keller dataset [15]. We compared learning *descriptors applicability* (PA) directly from molecular graphs using GNN with conventional approaches based on Mordred descriptors. Figure 7 and 8 provide an overview of individual model behaviors on two prediction tasks among which the main difference is the subjects' prior olfactory training. Note that linear regression as the most preliminary method was attempted, but the all-negative testing explained variance indicated that a simple linear relationship has poor generalizability in this scenario. Therefore, it is omitted in all the following visualization. The criteria for hyper-parameter search for all models are based on the weighted explained variance (EV) as criteria across all odor descriptors. Best-performing GNN for two prediction tasks was selected to achieve high but comparable weighted EV for comparison of model architecture in different prediction tasks.

With limited samples and unprocessed molecular structure information, GNN achieved an overall better performance both predicting values and explaining variance in Dravnieks predicting task (Figure 7-8). However, machine learning approaches especially random forest and gradient boosting model predominantly outperformed all other models in Keller prediction tasks. This indicates that the complex neural network structure potentially has advantages in decoding quantitative odor-structure relationship (QSOR) for trained people, while pre-computed physicochemical features team better with simple machine learning with appropriate complexity level in the quantifying lay people's sense of smell.

4.2 Embeddings as meaningful feature inputs for machine learning

The graph convolution layers of GNN generalize a representation associated with human population odor percept for each molecule. We extract the output of the pooling layers as learned embeddings, which are sequentially used as input for prediction neural networks in the whole GNN architecture. Considering Mordred descriptors have pre-fixed descriptor lists and sometimes suffer from missing values and colinearity among features, I hypothesize that GNN-learned embeddings can be used as more informative and comprehensive molecular features for simple machine learning techniques to learn on.

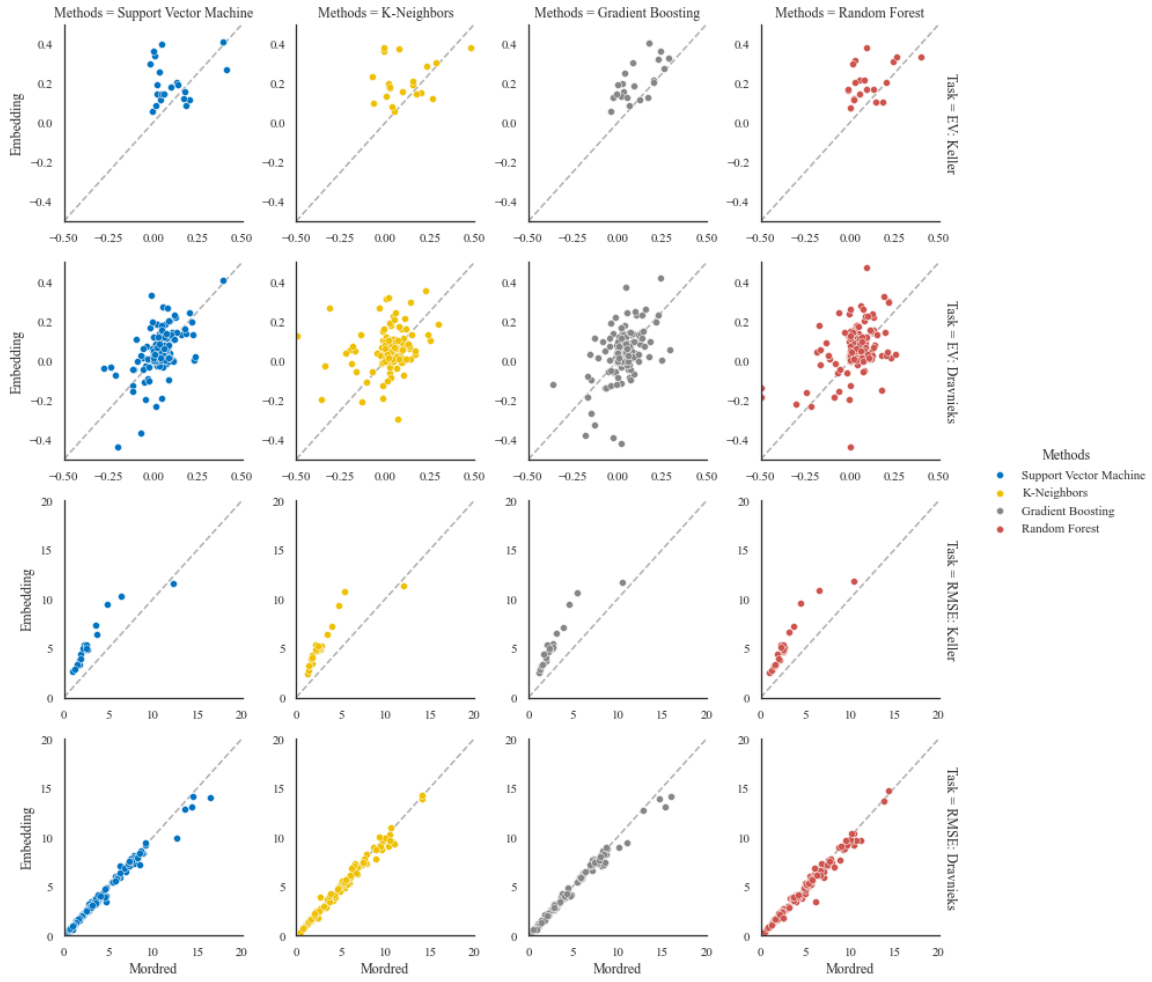

Figure 9: EV and RMSE on the same test set. Every point in the plot represents the metric-of-interest of an odor label for the prediction task (row name) under the model (column name). On each subplot, the x-coordinate is the odor descriptor-specific test metric of the model learned from Mordred features, and the y-coordinate is the test metric of the model fitted on the learned embeddings. On the diagonal dash line, odor descriptor-specific machine learning models perform equally well for a certain metric-of-interest.

For Keller's prediction task, there's a noticeable improvement in all machine learning regression models' ability to explain variances in *descriptor applicability* when used learned embeddings as input (Table 2, Figure 9). This implies the enhancement of models' capability of capturing the complexity of underlying relationships. However, as suggested by a consistent increase in RMSE for all odor descriptors, the shifting to learned embeddings leads to an overall worsened performance in predicting exact *descriptor applicability* value. Hence, embeddings may complement the model complexity by incorporating information outside the pre-fixed Mordred descriptor list, while extra

Methods	EV	RMSE
SVM	71.23%	73.29%
KNN	62.33%	69.86%
XGB	36.99%	36.99%
RF	55.48%	48.63%

Methods	EV	RMSE
SVM	100%	4.76%
KNN	100%	4.76%
XGB	80.95%	0%
RF	52.38%	0%

Table 2: The percentage of all odor descriptors towards which machine learning regression models from GNN-embeddings outperformed the ones on Mordred descriptors, as measured by each metric of interest in the column. The comparison is made specific to each prediction task (left: Dravnieks, right: Keller)

information can risk introducing additional noise in predicting the exact value for lay-human's odor percepts.

On the other hand, although the statistics in Table 2 appear to support the superiority of using XGB with Mordred descriptors, the approximately even layout of odor descriptor-wise metrics in Figure 9 suggest no observable difference between two feature representation when they collaborate with machine learning regression model in Dravnieks prediction task. By referring back to the GNN performance discussed in Section 4.1, one could infer that the complexity added by the two-layer fully-connected prediction networks is crucial to understanding and predicting odor percepts from pre-trained people.

To sum up, when it comes to learning lay-people's olfaction patterns reflected by *descriptor applicability*, one may benefit from replacing Mordred descriptors with graph convolution layer-learned embeddings to increase machine learning regression models' capability. Nevertheless, such a replacement can suffer from deviation from accurate predicting values. In addition, it exhibited little impact when experimented on olfaction responses gathered from trained participants.

4.3 Transfer learning across olfaction experience levels

Despite the emerging efforts in the olfaction field [5], the choice of odor descriptors used, experimental procedure, and participants' olfaction knowledge makes data integration difficult and laborious. On the other hand, by nature of human learning trajectory, we hypothesize that olfactory responses from experienced panelists can boost understanding of lay people's odor percepts from molecular structures. To validate our conjecture, we conducted transfer learning for all previously discussed models originally trained on the Dravnieks dataset to capture QSOR in the Keller prediction task. Although the machine learning models require us to focus on odor descriptor matching pairs for meaningful comparison, we are also able to compare all transfer learning results by GNN with the original GNN trained for the Keller prediction task.

Among five matching pairs, `"SWEET"` and `"GARLIC"` were consistently well-understood as all transfer learning models, with Gradient Boosting, Random Forest, original GNN, and transfer learning GNN model from Dravnieks task achieving competing explained variance (Figure 10). For the rest three descriptor pairs (`"SWEATY"`, `"DECAYED"`, and `"GRASS"`) toward which all simple machine learning failed to capture an instrumental QSOR, both transfer learning and original GNN uncovered complex QSORs as demonstrated by a strictly positive explained variance.

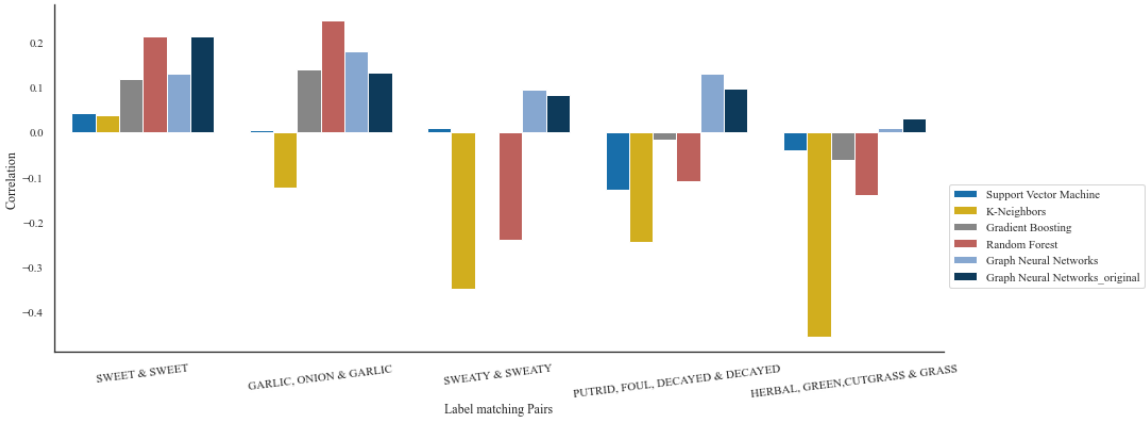

Figure 10: Explained Variance of transfer learning on the test set for odor descriptor matching pairs. The transfer learning result is also compared to the original GNN fully trained and tested on the Keller prediction task.

Neural networks architecture enables the transfer learning with varying targets through modifying and tuning final prediction networks. We studied the prediction correlation with ground-truth values and models' capacity to explain targeting variance for all odor descriptors adopted by the Keller dataset. Figure 11 shows no remarkable change in the interquartile range of explained variance and correlation. For the majority of odor descriptors used in Keller's task, transfer learning did not exert an impactful influence on improving model performances. However, the shortening of the whisker range in boxplots for transfer learning suggested significant improvement in both EV and correlation for the worst predicted odor descriptors by the original GNN. This complements our discovery discussed in Figure 10 that the model performance enhancement is dependent on the odor descriptor of interest. It points to a promising potential of incorporating psychophysical olfaction data that reflect domain knowledge to advance deep learning-based models in understanding ordinary people's odor percept.

5 Conclusion

In this study, I focus on modeling the odor descriptor applicability gathered from participants at two olfactory experience levels: the Dravnieks dataset from trained panelists, and the Keller dataset from untrained volunteers. Conventional machine learning regression models (Gradient Boosting, Random Forest, K-Neighbors, and Support Vector Machine) based on Mordred physicochemical molecular descriptors and Graph Neural Networks based on raw molecular graphs are explored on two prediction tasks. It is unclear and difficult to establish to what extent professionally trained people's olfaction experience agrees with ordinary people in the same smelling task. Taking the advantage of various machine learning techniques in the data science field, I explored their behavior on the olfactory experience of people with different domain knowledge training. First, I discovered that complex neural networks-based models demonstrated better capabilities of learning *descriptor applicability* from trained people, while less complex model structures such as random forest and gradient boosting generalize lay people's odor percepts better. Furthermore, I compared the informativeness of Mordred molecular descriptors with GNN-learned embeddings from raw molecular

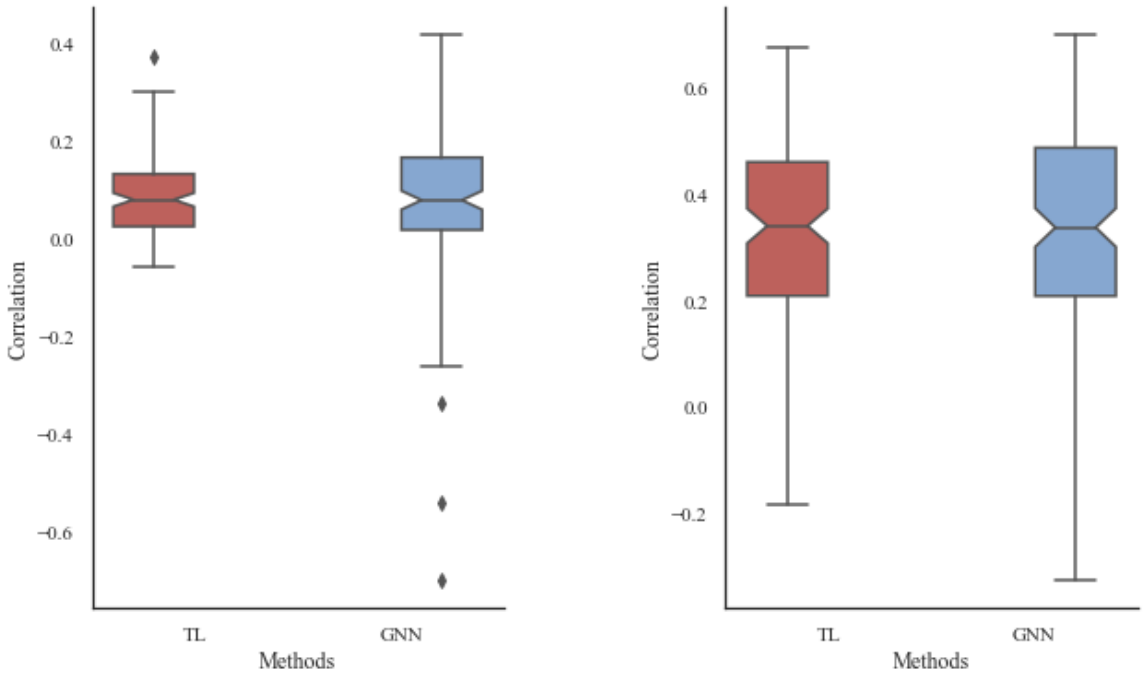

Figure 11: Performance distribution of explained variance (left) and correlation (right) by Transfer Learning (TL) model and Graph Neural Networks (GNN) model in Keller prediction task

graphs as feature input for machine-learning models. Compared to descriptor representation, predictors trained on learned embeddings explain a larger portion of variance in the target values from Keller's task but had greater difficulties predicting exact *descriptors applicability*. In the end, the whole transfer learning process validated our hypothesis that the incorporation of psychophysical olfactory data from professionally trained subjects could help build a more predictive deep learning model for ordinary people's sense of smell.

The research efforts discussed in this thesis complement current machine learning research for olfaction by integrating human subjects' domain knowledge level differences to better inform ordinary people's sense of smell. In future studies, various algorithm mechanisms could be experimented with to accommodate QSOR, such as attention mechanism or graph transformer. Moreover, progress data integration research could generally advance olfaction research, as it paves ways to establish benchmark datasets with satisfactory sample sizes in pursuit of robustness and predictiveness of models.

6 Acknowledgement

First and foremost I would like to express my heartfelt gratitude to my advisor, Professor Ambuj Tewari, for invaluable advice, guidance, and continuous support that throughout my research journey. His passion and dedication to the interdisciplinary studies of statistics, computer science, and chemistry inspired me to courageously challenge the limit of natural science using a quantitative approach. I would like to express my gratitude to my graduate mentors, Ziteng Pang and Jacob Trauger, for their intellectual mentorship and generous commitment. I extend my sincere thanks to all the members of CASI Lab for their generous sharing of resources and their outstanding achievements that motivated and inspired me throughout my studies.

I sincerely appreciate the unconditional support for all my decisions from my parents and little puppy, Seven. Without their tremendous understanding and encouragement, it would be impossible for me to complete my study on the other side of the globe and devote myself to a subject that I'm passionate about. I would like to thank my friends whom I met in Changsha, Beijing, and Ann Arbor. They are the most important part of my life and have given me various forms of caring and love during my growth. Special thanks go to my partner, Yuxuan Zhu, for his unwavering support, love, and cherished moments we have spent together. Lastly, I would like to thank all the faculties and staff at the statistics department for the studentship that allowed me to conduct this thesis.

A Supplemental Materials

A.1 Tables and Figures

AATSC0i	ATSC3c	NaaS	PEOE_VSA4	FCSP3
C1SP2	EState_VSA4	ATSC7d	NssS	SaaNH
nAHRing	JGI3	nCl	Xch-7dv	AATSC1v
RPCG	SMR_VSA6	SaaO	AATS1m	ATSC5se
JGI8	NdssC	JGI6	nBase	ATSC3m
nFaHRing	n8FHRing	IC2	n9FARing	NdsCH
n8FRing	JGI7	EState_VSA3	PEOE_VSA10	JGI4
C2SP1	ATSC2dv	PEOE_VSA2	n10FAHRing	PEOE_VSA3
ATSC6se	NssNH	n8FARing	SdsN	NdCH2
SsssN	JGI5	EState_VSA6	IC0	BalabanJ
NtN	SlogP_VSA11	PEOE_VSA9	n8FAHRing	

Table S1: Final Mordred descriptor sets as feature inputs for conventional machine learning regression methods, using colinearity criteria

	Keller	Dravnieks
Deep GCN Layer	Plain Generalized Graph Networks: [15, 20, 27] LayerNorm normalization ReLU activation Dropout rate = 0.47	[15, 20, 27, 36] LayerNorm normalization ReLU activation 0.47
Pooling Layer	Global add pool with dimension 175	175
Regressor Networks	Two fully-connected layers at dimension [96,36,N] N= 21: number of predicting columns	[96,36,N] N= 146

Table S2: Hyper-parameter of best-performing GNN model trained for individual prediction task

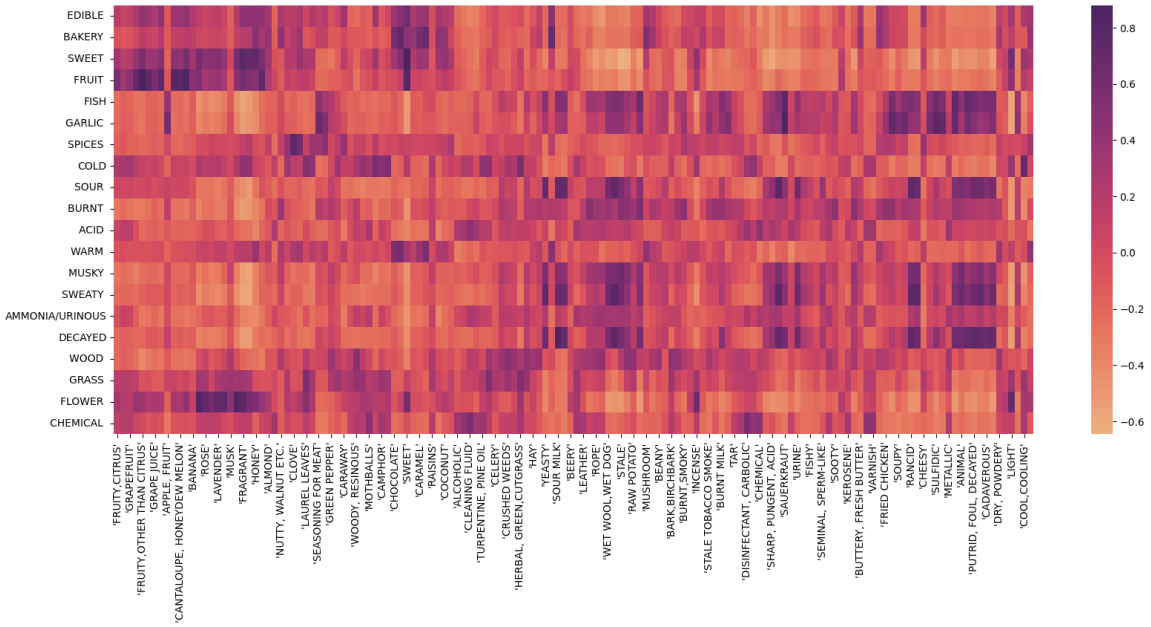

Figure S1: Correlation Heatmap for *Descriptor Applicability* on overlapped 61 chemicals in Dravnieks' dataset and Keller's dataset

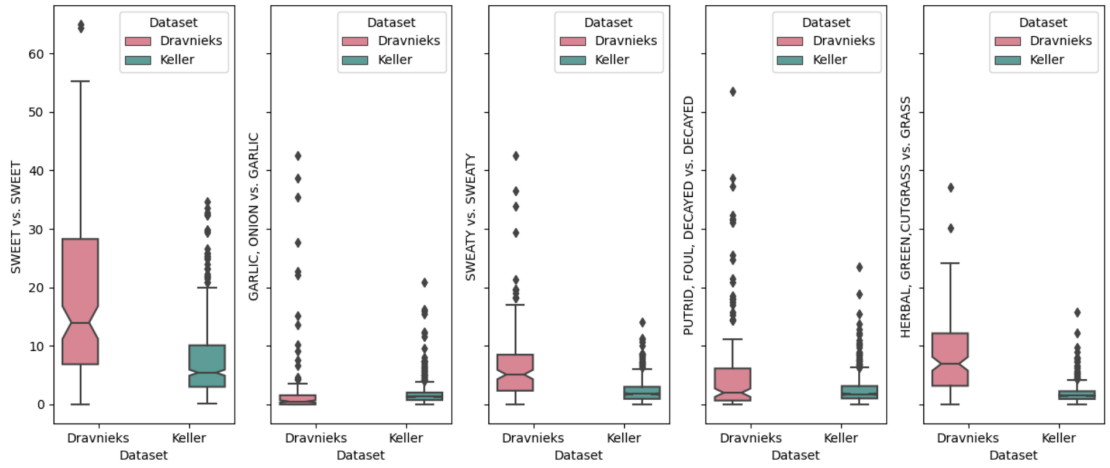

Figure S2: Boxplot for matched odor descriptor pairs.

A.2 Data And Code Availability

1. **Data Availability**: The raw dataset for Dravnieks and Keller's prediction tasks are archived by pyrfume project at https://github.com/pyrfume/pyrfume-data.

2. **Code Availability**: The code used to process data, the experimental code, and visual-

ization results are provided at https://github.com/rrrrn/ExploringMachineOlfaction. Hyperparameter selected for each odor descriptor and models' test performance can also be found under the 'result' directory.

References

[1] *The chemical compounds behind the aroma of coffee*, Feb 2015.

[2] *Facts about benzene*, Apr 2018.

[3] *molecule*, Aug 2022.

[4] W. Bo, Y. Yu, R. He, D. Qin, X. Zheng, Y. Wang, B. Ding, and G. Liang, *Insight into the structure–odor relationship of molecules: A computational study based on deep learning*, Foods, 11 (2022), p. 2033.

[5] J. B. Castro, T. J. Gould, R. Pellegrino, Z. Liang, L. A. Coleman, F. Patel, D. S. Wallace, T. Bhatnagar, J. D. Mainland, and R. C. Gerkin, *Pyrfume: A window to the world's olfactory data*, bioRxiv, (2022).

[6] R. Chacko, D. Jain, M. Patwardhan, A. Puri, S. Karande, and B. Rai, *Data based predictive models for odor perception*, Scientific reports, 10 (2020), pp. 1–13.

[7] A. Dravnieks and A. Dravnieks, *Atlas of odor character profiles*, ASTM Philadelphia, 1985.

[8] S. Gabler, J. Soelter, T. Hussain, S. Sachse, and M. Schmuker, *Physicochemical vs. vibrational descriptors for prediction of odor receptor responses*, Molecular Informatics, 32 (2013), pp. 855–865.

[9] R. C. Gerkin, *Parsing sage and rosemary in time: the machine learning race to crack olfactory perception*, Chemical Senses, 46 (2021).

[10] E. D. Gutiérrez, A. Dhurandhar, A. Keller, P. Meyer, and G. A. Cecchi, *Predicting natural language descriptions of mono-molecular odorants*, Nature communications, 9 (2018), pp. 1–12.

[11] R. Haddad, R. Khan, Y. K. Takahashi, K. Mori, D. Harel, and N. Sobel, *A metric for odorant comparison*, Nature methods, 5 (2008), pp. 425–429.

[12] C. Hopkins, P. Surda, and N. Kumar, *Presentation of new onset anosmia during the covid-19 pandemic*, Rhinology, 58 (2020), pp. 295–298.

[13] M. Jeltema and E. Southwick, *Evaluation and applications of odor profiling*, Journal of Sensory Studies, 1 (1986), pp. 123–136.

[14] A. Keller, R. C. Gerkin, Y. Guan, A. Dhurandhar, G. Turu, B. Szalai, J. D. Mainland, Y. Ihara, C. W. Yu, R. Wolfinger, et al., *Predicting human olfactory perception from chemical features of odor molecules*, Science, 355 (2017), pp. 820–826.

[15] A. Keller and L. B. Vosshall, *Olfactory perception of chemically diverse molecules*, BMC neuroscience, 17 (2016), pp. 1–17.

[16] J. Kowalewski, B. Huynh, and A. Ray, *A system-wide understanding of the human olfactory percept chemical space*, Chemical senses, 46 (2021).

[17] D.-H. Lee, S.-E. Woo, M.-W. Jung, and T.-Y. Heo, *Evaluation of odor prediction model performance and variable importance according to various missing imputation methods*, Applied Sciences, 12 (2022), p. 2826.

[18] G. Li, C. Xiong, A. Thabet, and B. Ghanem, *Deepergcn: All you need to train deeper gcns*, arXiv preprint arXiv:2006.07739, (2020).

[19] Q. Liu, D. Luo, T. Wen, H. GholamHosseini, X. Qiu, and J. Li, *Poi-3dgcn: Predicting odor intensity of monomer flavors based on three-dimensionally embedded graph convolutional network*, Expert Systems with Applications, 199 (2022), p. 116997.

[20] J. D. Mainland, Y. R. Li, T. Zhou, W. L. L. Liu, and H. Matsunami, *Human olfactory receptor responses to odorants*, Scientific data, 2 (2015), pp. 1–9.

[21] M. Matveieva and P. Polishchuk, *Benchmarks for interpretation of qsar models*, Journal of cheminformatics, 13 (2021), p. 41.

[22] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, *Dragon software: An easy approach to molecular descriptor calculations*, Match, 56 (2006), pp. 237–248.

[23] E. J. Mayhew, C. J. Arayata, R. C. Gerkin, B. K. Lee, J. M. Magill, L. L. Snyder, K. A. Little, C. W. Yu, and J. D. Mainland, *Transport features predict if a molecule is odorous*, Proceedings of the National Academy of Sciences, 119 (2022), p. e2116576119.

[24] H. Moriwaki, Y.-S. Tian, N. Kawashita, and T. Takagi, *Mordred: a molecular descriptor calculator*, Journal of cheminformatics, 10 (2018), pp. 1–14.

[25] Y. Nozaki and T. Nakamoto, *An olfactory sensor array for predicting chemical odor characteristics from mass spectra with deep learning*, Biomimetic Sensing: Methods and Protocols, (2019), pp. 29–47.

[26] R. J. Ouellette and J. D. Rawn, *21 - carboxylic acid derivatives*, in Organic Chemistry, R. J. Ouellette and J. D. Rawn, eds., Elsevier, Boston, 2014, pp. 699–745.

[27] E. Poivet, N. Tahirova, Z. Peterlin, L. Xu, D.-J. Zou, T. Acree, and S. Firestein, *Functional odor classification through a medicinal chemistry approach*, Science advances, 4 (2018), p. eaao6086.

[28] X. Qiu, Y. Cheng, D. Luo, C. He, K. Angus, and Q. Liu, *Odor-gcn: Graph convolutional network for predicting odor impressions based on molecular structures*, (2022).

[29] K. Saini and V. Ramanathan, *Predicting odor from molecular structure: a multi-label classification approach*, Scientific reports, 12 (2022), pp. 1–11.

[30] B. Sanchez-Lengeling, J. N. Wei, B. K. Lee, R. C. Gerkin, A. Aspuru-Guzik, and A. B. Wiltschko, *Machine learning for scent: Learning generalizable perceptual representations of small molecules*, arXiv preprint arXiv:1910.10685, (2019).

[31] L. Shang, C. Liu, Y. Tomiura, and K. Hayashi, *Machine-learning-based olfactometer: prediction of odor perception from physicochemical features of odorant molecules*, Analytical chemistry, 89 (2017), pp. 11999–12005.

[32] A. Sharma, R. Kumar, I. Aier, R. Semwal, P. Tyagi, and P. Varadwaj, *Sense of smell: structural, functional, mechanistic advancements and challenges in human olfactory research*, Current neuropharmacology, 17 (2019), pp. 891–911.

[33] A. Sharma, R. Kumar, S. Ranjta, and P. K. Varadwaj, *Smiles to smell: decoding the structure–odor relationship of chemical compounds using the deep neural network approach*, Journal of Chemical Information and Modeling, 61 (2021), pp. 676–688.

[34] K. Snitz, A. Yablonka, T. Weiss, I. Frumin, R. M. Khan, and N. Sobel, *Predicting odor perceptual similarity from odor structure*, PLoS computational biology, 9 (2013), p. e1003184.

[35] S. Tempere, M.-L. Hamtat, G. de Revel, and G. Sicard, *Comparison of the ability of wine experts and novices to identify odorant signals: A new insight in wine expertise*, Australian Journal of Grape and Wine Research, 22 (2016), pp. 190–196.

[36] L. Torrey and J. Shavlik, *Transfer learning*, in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, IGI global, 2010, pp. 242–264.

[37] T. Wen, Z. Mo, J. Li, Q. Liu, L. Wu, and D. Luo, *An odor labeling convolutional encoder–decoder for odor sensing in machine olfaction*, Sensors, 21 (2021), p. 388.

[38] D. S. Wigh, J. M. Goodman, and A. A. Lapkin, *A review of molecular representation in the age of machine learning*, Wiley Interdisciplinary Reviews: Computational Molecular Science, 12 (2022), p. e1603.

[39] D. A. Wilson, A. Donald, R. J. Stevenson, R. J. Stevenson, and R. J. Stevenson, *Learning to smell: olfactory perception from neurobiology to behavior*, JHU Press, 2006.

[40] D. Wu, D. Luo, K.-Y. Wong, and K. Hung, *Pop-cnn: Predicting odor pleasantness with convolutional neural network*, IEEE Sensors Journal, 19 (2019), pp. 11337–11345.

[41] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, et al., *Analyzing learned molecular representations for property prediction*, Journal of chemical information and modeling, 59 (2019), pp. 3370–3388.

[42] K. Yoshikawa, J. Deguchi, J. Hu, H.-Y. Lu, and H. Matsunami, *An odorant receptor that senses four classes of musk compounds*, Current Biology, 32 (2022), pp. 5172–5179.

[43] A. Yoshimori, *Prediction of molecular properties using molecular topographic map*, Molecules, 26 (2021), p. 4475.