Teaching Machine Olfaction in an Undergraduate Deep Learning Course: A Pilot Integrating Chemistry, Machine Learning, and Sensory Evaluation

Yikun Han
Department of Statistics,
University of Michigan
and
Michelle Krell Kydd
Independent Researcher,
Flavors and Fragrances
and
Joseph Ward
Department of Chemistry,
University of Michigan
and
Ambuj Tewari
Department of Statistics, and
Department of EECS,
University of Michigan

June 27, 2026

Abstract

We describe a pilot implementation of a classroom activity that introduces machine olfaction, a cutting-edge research topic, into an undergraduate deep learning course, STATS 315, at the University of Michigan. Previously, the course focused on text and image data, but we introduce smell as a new modality to expand students' learning experiences. By incorporating hands-on activities using carefully selected odorants, we engaged students in the data collection process and highlighted the complexities and variability inherent in sensory data. We also exposed students to

graph neural networks and the curated GoodScents-Leffingwell dataset. The module adopted a multidisciplinary approach that combines chemistry, machine learning, and sensory evaluation. This pilot provides initial evidence for the feasibility of incorporating machine olfaction into deep learning instruction in the educational context where we implemented the activity, and it suggests directions for future curricular development. We discuss practical challenges encountered and propose directions for improving and extending this educational activity.

Keywords: Machine Learning Education, Machine Olfaction, Sensory Evaluation, Deep Learning, Data Literacy

1 Introduction

Many universities now offer an introductory course on deep learning at the undergraduate level. At the University of Michigan, STATS 315 (now offered as DATASCI 315), titled Statistics and Artificial Intelligence, is taught in the Department of Statistics and has been introducing deep learning to undergraduate students since 2022. As an introductory course, it is designed to be accessible to students with limited programming and mathematical backgrounds. The formal prerequisites for the course include prior coursework in statistics or data science, programming, and calculus. In practice, students were best prepared if they had completed an introductory statistics or data science course covering basic concepts such as linear regression, had programming experience in Python or R (including data manipulation and basic statistical modeling), and had familiarity with calculus at the level of derivatives and basic optimization concepts. As a result, students typically enter with some exposure to basic data analysis, coding, and mathematical tools used in introductory predictive modeling. The goal of the course is to introduce students to the core ideas of deep learning, including how models are trained from data, how their performance is evaluated, and how they can be applied to prediction tasks across different types of data. The course covers foundational topics such as regression and classification, neural networks, and modern architectures such as convolutional networks and transformers, with an emphasis on hands-on implementation using real datasets.

Given the students' limited knowledge about deep learning, the current design of the course focuses primarily on text and image datasets. These modalities are well-developed within the field of deep learning, with widely available datasets, tutorials, and software tools that make them natural starting points for introductory courses. For example, the datasets used in this course are mainly sourced from easily available online repositories, such as Modified National Institute of Standards and Technology (MNIST) database, movie reviews from the Internet Movie Database (IMDb), and the Boston Housing Price dataset (Chollet 2021). These datasets are pre-collected, well-organized, and labeled, providing a solid foundation for students to effectively learn and apply deep learning techniques.

In contrast to these pre-collected datasets, the activity described in this paper involves students in constructing and analyzing their own dataset. Materials to support replication

of this activity, including code and datasets, are available in an accompanying Open Science Framework (OSF) repository at `https://osf.io/jqn4y/`. This URL is also available as part of our Data Availability Statement (Section 9). The repository includes datasets, preprocessing scripts, and modeling code; see the README file for an overview of the contents and how they relate to the activities described in this paper.

The seven students who volunteered for the data collection activity met the same course prerequisites, although they may have been more motivated or interested in the topic. Not only did these volunteers construct their own dataset, but also the dataset went beyond image and text to include olfaction.

1.1 Motivation for Teaching Machine Olfaction

The inclusion of multimodal data, here referring to the use of multiple types of data such as images, text, and sensory inputs like smell, has become increasingly prevalent over the last decade, both in the fields of machine learning (Baltrušaitis et al. 2018; Barua et al. 2023) and learning analytics (Blikstein 2013; Mu et al. 2020). Educational research suggests that integrating non-visual sensory modalities, such as smell, can enhance memory retention, deepen conceptual understanding, and increase student engagement (Garcia-Ruiz et al. 2021; Marino et al. 2025). Therefore, in addition to text and image data, we introduced smell as a novel modality in our deep learning course.

The application of machine learning (ML) to solve prediction problems in olfaction is called machine olfaction. In comparison with computer vision and natural language processing, machine olfaction is less established. The web makes it easy to create vision data sets with millions of images and curate text corpora with billions of words. However, collecting smell related datasets is inherently challenging. By introducing this new modality, we aim to make students aware of the complexities and difficulties involved in building new and usable datasets.

Furthermore, we seek to increase student engagement by transforming them from passive users of pre-existing datasets into active agents of data generation. Prior work suggests that involving students in the data collection process can increase engagement and provide a more meaningful connection to the data (Kuiper et al. 2025). In our implementation,

the activity had two components: (i) all students in the course were introduced to machine olfaction through lectures and in-class discussion, and (ii) a small group of volunteer students participated in a hands-on data collection activity outside of regular class time, after the full class had been introduced to the topic.

Additionally, the COVID-19 pandemic highlighted the importance of smell in daily life, as anosmia became a common symptom (Olofsson 2025, Chapter 10). Providing general information on how humans perceive, interpret, and articulate smells as a primer for evaluating odorants can make the modality of human olfaction, and its connection to machine olfaction, more tangible.

Finally, introducing students to olfaction in a systematic way can help them better understand the significance of sensory data and its real-world applications, while also increasing awareness of sensory perception's role in health and well-being, as we observed in our implementation.

1.2 Machine Olfaction

Replicating human olfactory senses with instruments has been a longstanding goal in the field of sensory technology (Guthrie 2017). Electronic noses, or e-noses, are widely used for sensor-based machine olfaction research (Gutierrez-Osuna 2002). Publicly accessible machine olfaction datasets remained rare until 2015, when the DREAM (Dialogue for Reverse Engineering Assessment and Methods) Challenges team organized an olfaction prediction challenge. The 2015 Olfaction Prediction Challenge significantly advanced the application of traditional machine learning approaches to quantitative structure–odor relationship prediction (Keller et al. 2017), using a dataset introduced in earlier work (Keller and Vosshall 2016). Since then, machine olfaction has seen increased exploration and development. This surge in interest is particularly notable in the post-DREAM 2017 and post-COVID period, as shown in Figure 1, which illustrates the growth in publications and citations related to machine olfaction over the last 15 years. Although the COVID-induced surge appears to be over, machine olfaction retains a healthy degree of interest in the community. In fact, the DREAM Olfactory Mixtures Prediction Challenge was organized in 2024 and attempted to do for odorant mixtures what the 2017 challenge did for single-molecule odorants.

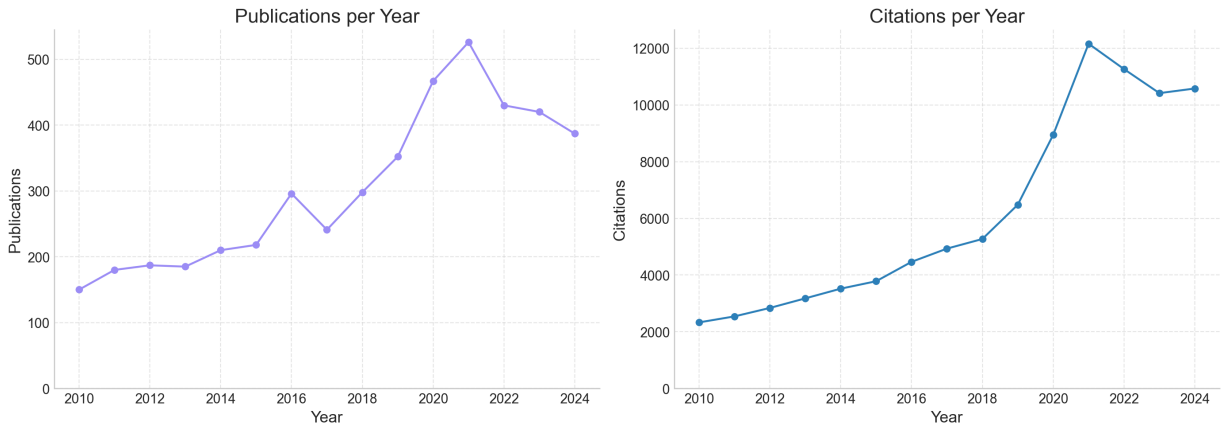

Figure 1: Publication and citation trend for "machine AND (olfaction OR smell)" over the last 15 years.

Graph neural networks (GNNs) are deep learning models designed to operate on graph-structured data, such as molecules represented by atoms as nodes and chemical bonds as edges. In the era of deep learning, GNNs have emerged as the state-of-the-art models for machine olfaction (Sanchez-Lengeling et al. 2019). These models demonstrate human-level odor description performance (Lee et al. 2023) using the curated GoodScents-Leffingwell (GS-LF) dataset, which merges datasets collected by GoodScents (The Good Scents Company 2019) and Leffingwell & Associates (Leffingwell & Associates 2001).

Researchers are also working on building tools, models, and datasets to support odorant-related research. For example, the Pyrfume project (Hamel et al. 2024) provides curated datasets and example usage for data analysis, which is beneficial for beginners in this research community.

Despite these encouraging advances, it remains challenging to integrate machine olfaction into an undergraduate deep learning course for a couple of reasons. First, machine olfaction is a new field and still under active development. Second, this field lies at the intersection of sensory evaluation, chemistry, and data science. Teaching machine olfaction is challenging because students in statistics and data science courses will usually not have much prior background in chemistry or sensory evaluation.

To address these challenges, we designed and implemented a pilot activity integrating sensory evaluation, chemistry, and machine learning within an undergraduate deep learning

course. A summary of the course structure and the placement of the machine olfaction module is provided in Appendix B. In this offering, the activity was introduced toward the end of the course, allowing it to build on students' prior exposure to core deep learning concepts.

We emphasize that while all students in the course were exposed to machine olfaction through lectures and course materials, only a subset of volunteer students participated in the hands-on data collection component.

Contributions to teaching. This paper contributes to deep learning education by (i) demonstrating how to incorporate a new sensory modality into an introductory course, (ii) providing a hands-on, student-centered data collection activity that highlights the challenges of real-world datasets, and (iii) offering practical resources that enable instructors to replicate and adapt the activity in their own classrooms.

2 Machine Learning Background and Smell Activity Preparation

One way to embed machine olfaction in a deep learning course is to present it to the students as a multilabel classification problem. The input is an odorant molecule M, typically encoded as a graph with atoms as nodes and chemical bonds as edges. The output is a binary odor descriptor vector $Y \in \{0,1\}^K$ corresponding to whether or not K fixed odor labels (e.g., "floral", "sweet", "pungent") apply to the input odorant. The deep learning task is to learn a mapping f_{GNN} from molecules to $[0,1]^K$ using a data set consisting of (M, Y) pairs such that for a future molecule M' with correct label Y', the kth entry $f_{GNN,k}(M')$ of the prediction vector is the probability that the kth odor label applied, i.e., $Y'_k = 1$. The most common loss function in a deep learning context is binary cross entropy loss:

$$\mathcal{L}_{\text{BCE}}(f_{GNN}, M, Y) = -\sum_{k=1}^{K} \left[Y_k \log(f_{GNN,k}(M)) + (1 - Y_k) \log(1 - f_{GNN,k}(M)) \right] \ .$$

Figure 2 provides a schematic illustration of how molecules are represented, how odor descriptor labels are assigned, and how the resulting dataset is organized for model training.

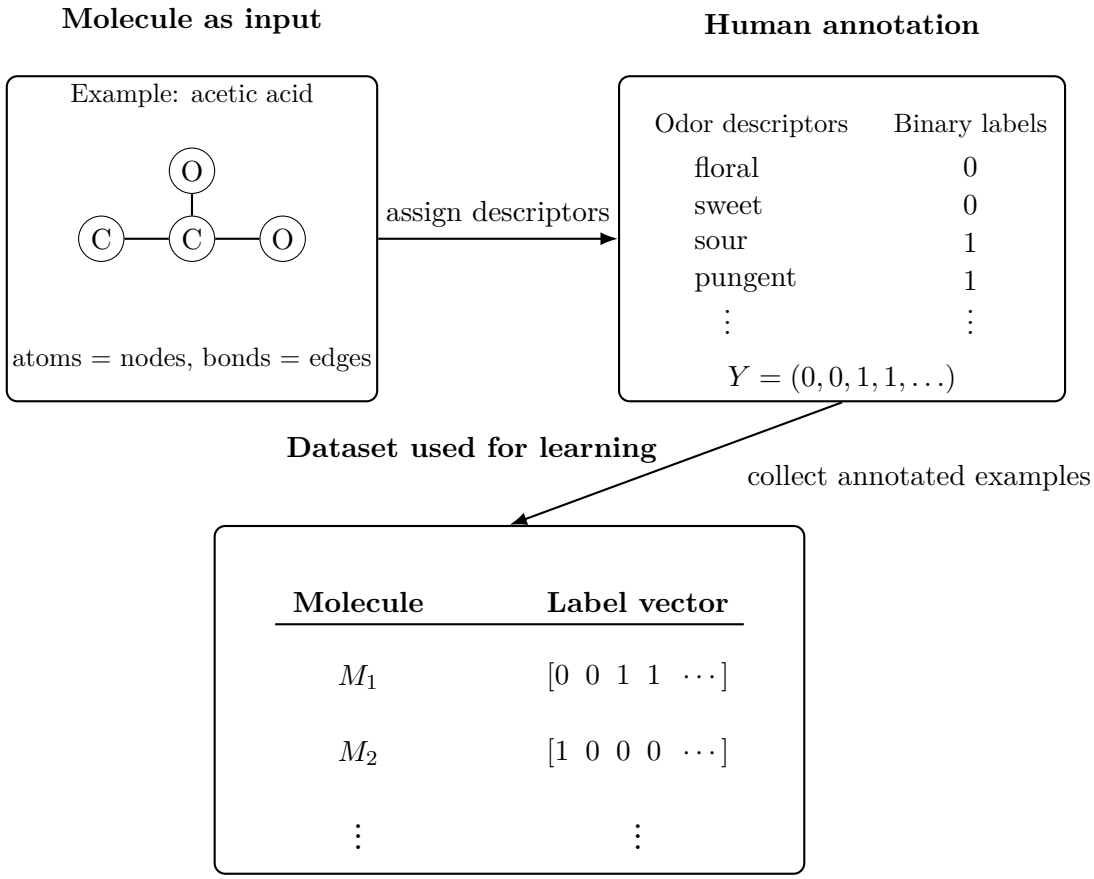

Figure 2: Schematic illustration of the machine olfaction setup. A molecule is represented as a graph with atoms as nodes and chemical bonds as edges. Human annotation assigns odor descriptors to the molecule, producing a binary label vector indicating which descriptors apply. A dataset for machine learning consists of many such molecule–label pairs.

Our ultimate goal was for students to create their own machine olfaction dataset by smelling a collection of single molecular odorants M and recording their own smell perceptions in binary vectors Y. Unlike vision or audition, the typical K-12 education we receive does not train us to describe smells accurately. So we knew we had to train our students in sensory evaluation. But we noticed a tension between our ML data collection goals and sensory evaluation training. In any sensory evaluation training, participants are typically trained with mixtures. This makes sense because most natural odorants consists

of a mixture of a large number of different molecules. For example, coffee contains over 1000 different compounds (Angeloni et al. 2021). The DREAM 2024 challenge (Consortium 2024) might eventually lead to better predictions for olfactory mixtures. However, at the moment, most current machine learning models are unable to handle mixtures for prediction; they can only take digital representations of single molecules as input to predict their odor semantic labels.

To resolve this tension and to prepare students to understand machine olfaction as an ML task, we decided to structure the machine learning module as two guest lectures attended by everyone and an hands-on data collection activity attended only by a small group of volunteers with a high degree of interest in machine olfaction. The first guest lecture taught the basics of sensory evaluation using natural odorants that are mixtures of many different kinds of molecules. The second guest lecture gave an overview of machine olfaction and the role GNNs play in solving the prediction problem. The selection of single molecule odorants to use in the hands-on activity required careful preparation involving our entire interdisciplinary team.

As mentioned above, most people have never been trained for sensory evaluation, let alone sensory evaluation with single-molecule odorants. In real life, people are mostly exposed to mixtures found in nature, and it is rare for them to interact with single molecules, which are unfamiliar to them. If the smell of a single molecule is perceived as atypical, it can shift the perceiver's focus away from describing the odorant's character (what it smells like) to describing its hedonic tone (whether it is likable or not).

Given these challenges, we chose to select single molecules, drawing on the experience of previous researchers in similar experiments. We initially considered 338 single molecules from the training set used in previous studies (Keller et al. 2017). Using their compound identifiers (CIDs), we fetched data from PubChem to ensure these molecules were not acutely toxic, environmentally hazardous, corrosive, flammable, or health hazards. We then grouped them by their primary odor (the odor label which had the highest value for a given single molecule).

Given the lecture duration of 80 minutes, we initially selected 24 molecules (six different types of primary odor, with four single molecules for each type) and ensured that the least

pleasant one had a pleasantness score above 40 (on a scale of 100). This allocation allowed for approximately three minutes for each evaluation. However, due to the large minimum purchase quantities, we found the Flinn Scientific Organic Smell Identification Kit[1] to be a convenient and economical alternative, providing 20 different molecules, which we quickly determined to be pleasant and acceptable. The 20 molecules were: Ethyl methylphenylglycidate, Isoamyl acetate, 2-Phenylethanol, (S)-Carvone, Vanillin, (R)-Limonene, Undecalactone, Geraniol, (R)-Carvone, Methyl salicylate, 1-Octen-3-ol, Octyl acetate, Trans-cinnamaldehyde, Benzyl acetate, Menthol, Methyl ionone, 2-Phenylpropanol, Octalactone gamma, Acetic Acid, and Anethole.

The details about these molecules, including their unique Chemical Abstracts Service (CAS) Registry Numbers, which serve as standardized identifiers for chemical substances, and their SMILES (Simplified Molecular Input Line Entry System) strings, which provide a text-based representation of molecular structure for computational use, are provided in Appendix A, along with information on how the molecules were prepared for use in small bottles.

Blotters were used to dip into the small bottles and could be smelled after the ethanol evaporated. Two rounds of evaluation by three of the authors were conducted before presenting them to the students and the results indicated that most of the chemicals were pleasant.

3 Implementation

We implemented three activities: two guest lectures to familiarize all students with sensory evaluation and machine olfaction, respectively, and one hands-on activity in which a subset of volunteer students recorded their own olfactory perceptions when smelling odorants listed in Table 1. These volunteer activities were conducted outside of regular class time after the full class had been introduced to the topic. With slight abuse of terminology, we will refer to them as Flinn kit odorants.

[1] https://www.flinnsci.com/organic-smell-identification—super-value-kit/ap1566/

3.1 Guest Lectures

We invited two guest lecturers, Michelle Krell Kydd and Alex Wiltschko, who each spoke to an audience of over 100 students. Both lectures were delivered as Zoom presentations with volunteers present in-person to facilitate engagement between the speaker and the students.

3.1.1 Smell & Tell

Given that most students are more familiar with mixtures, we began with a Smell & Tell session titled "Stats Smellathon: An Introduction to Sensory Evaluation," designed and delivered by Michelle Krell Kydd. The session started with a 45-minute sensory introduction to provide students with a general understanding of olfaction, sensory evaluation, and sensor technology.

After the introduction, we included a scent flight consisting of four anonymized plant-based odorants in solution on perfume blotters for smelling: rose geranium leaf, lavender, balsam fir, and chocolate. The mixtures selected for the Smell & Tell session were selected for their hedonic quality (positive). We used materials that are used in perfumes and flavors: rose geranium leaf oil is categorically floral and contains geraniol (which is also in the Flinn kit); lavender oil is categorically rustic/camphorated and contains linalyl acetate; balsam fir is categorically woody-coniferous-balsamic; chocolate absolute is categorically gourmand.

Students were instructed to write down sensory evaluations for an odorant using a free-form text field, describing the perceived smell in their own words. This was followed by listening to three student volunteers who agreed to share their sensory impressions with the class. Three volunteers initially shared their impressions, but as the wireless microphone was passed around by the GSIs, additional students volunteered, and the number of participants grew to four, five, or more in a class of over 100 students.

Furthermore, students were asked to take notes on evaluations shared by their peers and compare them to their own for the purpose of identifying patterns in emotions, memories, and descriptions of scents. Though the word "collaboration" was not explicitly articulated as the purpose or goal of the sensory evaluation exercise, the activity, which is modeled on professional perfumery training, is designed to foster a form of subjective consensus.

For instructors who want to include a session like this in their course but do not have access to a sensory evaluation expert, we want to point out that Michelle Krell Kydd joined virtually via Zoom. Instructors therefore do not have to restrict their search to sensory evaluation experts who are local to their area. Moreover, basics of sensory evaluation such as the use of blotters are covered well in several videos available online. For instructors who do not have access to a sensory evaluation expert, introductory video resources can provide a practical substitute. For example, publicly available demonstrations of fragrance blotter use, such as instructional videos on YouTube[2] on how to evaluate scents using blotter strips, show how to apply odorants, control exposure, and assess scent characteristics in a structured way. Such resources can be used to prepare students before conducting a simplified version of the activity described here. We also note that a standard textbook reference for sensory evaluation is *Sensory Evaluation Techniques* by Meilgaard et al. (2016).

3.1.2 Mapping Scent

After students were introduced to sensory evaluation, we aimed to expand their knowledge by introducing them to machine olfaction, a cutting-edge area in machine learning. We were fortunate to have Alex Wiltschko, founder and CEO of Osmo and a 2009 University of Michigan alumnus, give a virtual lecture on "Mapping Scent".

Dr. Wiltschko explained how researchers convert molecules into formats that machine learning models can process. In particular he explained how molecules can be represented as graphs with atoms as nodes and chemical bonds as edges. He then introduced the concept of message passing for graph neural networks (Gilmer et al. 2017), in which each atom updates its representation by aggregating information from its neighboring atoms along chemical bonds (see Figure 2). He highlighted the unique challenges of machine olfaction: unlike vision, where color can be easily represented by the RGB (Red, Green, Blue) model, odor receptor interactions are complex and not straightforward to model.

He also emphasized two challenges unique to machine olfaction. First, the connection between molecular structure and its odor perception is subtle. Structurally similar molecules can have very different smells. On the other hand, it is also possible for two

[2]For example, see `https://youtu.be/VL2cw7Mro2g`.

odorants with almost identical smells to have significantly different structures. Second, olfaction data is much more scarce than other types of data, and the methods used to pre-train models for text or images do not work as effectively for olfaction. He also explained the importance of curating the GoodScents-Leffingwell data to ensure it was large enough to train a neural network. Finally, Dr. Wiltschko introduced the "principal odor map" (Lee et al. 2023), a model that captures the relationships among odors and correlates strongly with natural metabolic compounds. This map could potentially identify new molecules to address global health challenges, underscoring the societal impact of digitizing the sense of smell.

Instructors who do not have access to a suitable speaker can easily replace the "mapping scent" guest lecture with videos that students can watch offline and discuss in class with the guidance of the instructor. One resource we particularly recommend is Lecture 10 (ML for Scent) from the 2020 edition of MIT's Deep Learning course 6.S191[3]. This specific lecture provides an accessible overview of how molecular representations are used in machine learning models and can serve as a ready-to-use teaching resource for introducing machine olfaction concepts.

3.2 Hands-On Data Collection Activity

The purpose of the hands-on sensory evaluation activity was to directly support the machine learning component of the course by generating a dataset in the molecule-odor label vector format described in Section 2 and illustrated in Figure 2.

After the guest lectures, we offered the entire class a chance to participate in the hands-on data collection activity. However, as the activity was scheduled outside of class, involved a significant time commitment, and did not come with any extra course credits, only seven students who had a strong interest in machine olfaction volunteered to participate. Unlike the natural mixtures used in the Smell & Tell guest lecture, data collection in this session focused exclusively on 20 single molecule Flinn kit odorants. Participants were allowed to proceed at their own pace, with approximately four minutes allocated to each molecule's

[3]As of March 2026, both the lecture slides as well as the lecture video were available at `https://introtodeeplearning.com/2020/index.html`.

evaluation.

Before providing the students with odor semantic descriptor labels, we narrowed down the descriptor range through several steps. The curated dataset of the state-of-the-art model that we used features as many as 138 labels, which would impose a high cognitive load on participants. To prepare a manageable set of labels for the student activity, we constructed a reduced descriptor vocabulary from a larger pool used in machine learning models using the following approach:

- A sensory evaluation expert (Michelle Krell Kydd) conducted a blind assessment of all 20 molecules and selected applicable odor descriptors for each odorant.
- An open-source machine learning model (Barsainyan et al. 2023; Lee et al. 2023), based on graph neural networks that take molecular structure as input and predict odor descriptors, was used to generate predicted labels for the same set of molecules.
- The two sets of descriptors were combined and restricted to the original pool of 138 labels, resulting in a final set of 36 descriptors used in the activity.

The resulting set of 36 descriptors was then used to construct the questionnaire presented to students during the data collection activity.

Figure 3 presents the descriptor-selection frequency heatmap, showing, for each of the 20 single-molecule odorants, how many of the seven student participants selected each of the 36 odor descriptors. A value of 0 indicates that none selected that descriptor (suggesting consensus that it did not apply), whereas values 1–7 indicate increasing levels of agreement among students. Higher values (e.g., 6 or 7) correspond to clear-cut agreements, such as "vanilla" and "sweet" for vanillin, or "cinnamon" for transcinnamaldehyde. We note that the GNN model used in this activity was pre-trained on a larger dataset (the GoodScents-Leffingwell dataset) and was not trained specifically on the 20 molecules used in the activity. The value of the pretrained GNN model is that its single deterministic set of descriptors for each molecule provide a stark contrast to the variability observed across student responses.

Our Open Science Framework (OSF) repository: `https://osf.io/jqn4y/`, provides materials to support replication and extension of the activity, including scripts for constructing safe odorant datasets from public chemical databases, a preprocessed olfaction

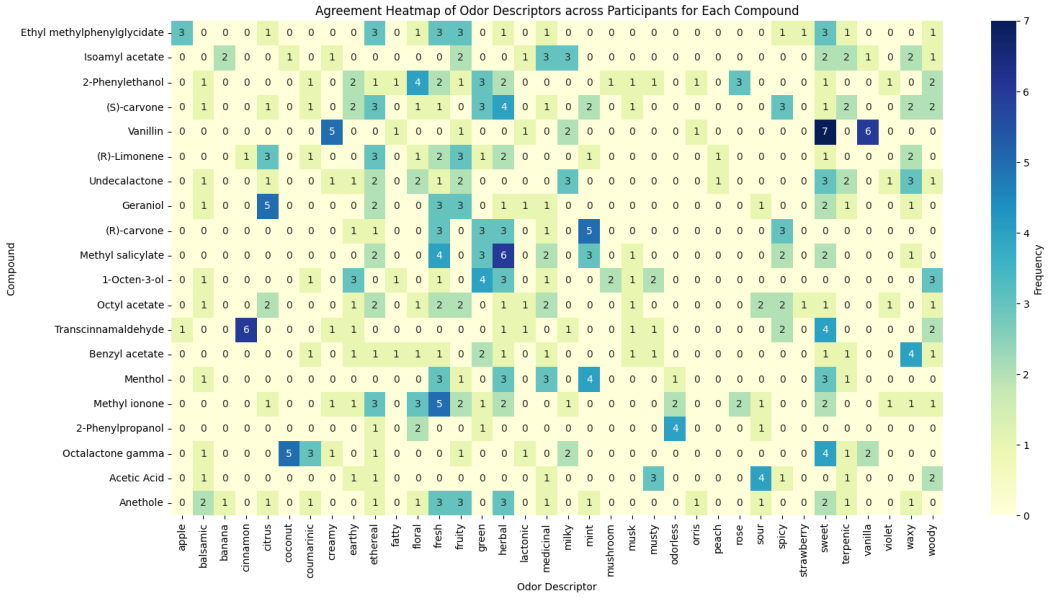

Figure 3: Descriptor-selection frequency across seven participants for each compound. Each cell $[r, c]$ shows how many students selected descriptor c for compound r (0–7). The figure highlights both agreement (high-frequency cells) and variability (dispersed low-frequency selections) in student-reported odor descriptors across compounds.

dataset, and code for training and using a machine learning model to generate odor descriptor predictions. It also includes the code used for the activities and analysis in this paper, along with the Google Form used to collect student responses, which presented checkboxes for each of the 36 labels and an open text field for instances where none of the labels applied.

4 Qualitative Reflections on Student Learning

The observations in this section are based on a combination of sources, including video recordings and transcripts of the guest lectures, as well as responses collected through a Google Form during the hands-on activity.

The students developed an appreciation for the variation in human olfactory perception. Additionally, students brought their cultural backgrounds into the sensory perception process. For instance, when evaluating the methyl salicylate odorant, some students referenced Fengyou Essence, a type of traditional Chinese medicine, in their free text responses.

The opportunity for hands-on data collection provided the students with a unique learning experience. Students reported that the activity evoked personal memories and made the concepts of sensory data and machine learning more tangible. This aligns with longstanding recommendations in statistics and data science education literature, which emphasize the benefits of student-designed or collected data in fostering deeper engagement and understanding (Carver et al. 2016; Finzer 2013; Cobb 2007) as well as more recent work in JSDSE that incorporates student-generated data into classroom activities (Kuiper et al. 2025). Notably, one undergraduate student who participated in the activity is now engaged in ongoing research on machine olfaction as part of the University of Michigan CASI team, which recently performed well in the DREAM Olfactory Mixtures Prediction Challenge (Consortium 2024).

We note a limitation of our implementation that can be easily overcome in the future in our own course as well as courses elsewhere. Due to the limited machine learning background of students and only one overview lecture on GNNs, we did not attempt to have the students themselves learn how to use a GNN model to predict the odor labels. Instead, we ran the model on the same sets of molecules as they smelled and discussed

the results with them. They still got a direct exposure to multilabel prediction problems and to GNNs as a tool to predict molecular properties. But a second course on machine learning that introduces GNNs early on can easily have the students use their own data to further explore GNN predictive models. Such exploration can involve further fine-tuning of the existing model to better predict the students' responses. Alternatively, the students can study the relationship of the existing model's prediction with the students' responses. The model weights we provide as part of the data repository accompanying this paper can help with such additional exploration and further enhance student learning.

5 Future Directions

In the future, we plan to implement several improvements to the machine olfaction educational content. Before allowing motivated students to participate in hands-on activities, it would be beneficial to formally introduce graph neural networks and how they are used to solve tasks in chemistry. We will focus on the message passing neural network as a foundational example. Additionally, we will incorporate the GoodScents-Leffingwell dataset specific to machine olfaction, along with established datasets from the Open Graph Benchmark (Hu et al. 2020, 2021). A key component will be demonstrating how molecules can be represented as graphs for these models.

Furthermore, we aim to design enhanced sensory evaluation training. Providing more background information could lead to more precise and higher-quality data collection. We also plan to offer detailed training sessions focusing on both mixtures and single molecules before formal evaluations to better prepare the students.

We scheduled our sensory evaluation session at 1 PM, shortly after lunch, which may have introduced confounding food-related odors and affected students' olfactory responses. In addition, olfactory sensitivity is known to vary with circadian timing, following a daily rhythm in which sensitivity fluctuates throughout the day (Herz et al. 2018). Future implementations of the activity will take these factors into account when selecting a time for data collection, rather than defaulting to early afternoon scheduling.

An additional direction is to explore how the hands-on data collection component could be more broadly integrated into the course. In our current implementation, this component

was conducted by a small group of volunteer students, which allowed for flexibility but limited participation. One possibility is to incorporate data collection as a required activity for the entire class, although this would require careful consideration of logistical constraints such as time, materials, and coordination. Another option is to have the volunteer students share their experiences, including the data collection process and challenges encountered, so that all students can benefit from the hands-on component even if they do not directly participate.

If there is more time available to integrate such an activity with a deep learning course, students could be encouraged to download a publicly available machine olfaction model or even build their own using publicly available olfaction datasets. They could then test these models on their own sensory evaluations to gain a better understanding of the strengths and limitations of machine olfaction models.

Finally, future iterations of this activity will incorporate more systematic collection and analysis of student feedback to better understand the learning process and refine the design of the module.

6 Conclusion

Image and text data analysis have become standard components of data science education. Well-organized datasets are readily available through various Python packages making them convenient pedagogical tools. We were motivated to walk on less trodden paths and explore the integration of new sensory data modalities into our teaching activities. Exposing students to new data types can enhance their understanding of real-world applications and broaden their skills in diverse domains.

Engaging students in the process of collecting their own data offers a unique experience. This is particularly true in an era where we are increasingly becoming passive users of machine learning technology powered by data that was labeled by other people often via crowdsourcing. Through this active learning approach, students gain firsthand insight into the variability and noise inherent in dataset creation. Such an experience can deepen their understanding of the general machine learning workflow and the challenges associated with building robust models.

Additionally, engaging with multiple disciplines can be highly beneficial for students. Integrating cutting-edge applications like machine olfaction not only broadens their horizons but also fosters a mindset that values multidisplinary collaboration, preparing them for future research endeavors.

7 Limitations

This study is based on a pilot implementation of a machine olfaction activity in an undergraduate deep learning course, and several limitations should be noted. First, the hands-on data collection component was conducted by a small subset of students: 7 out of approximately 110 students who were exposed to the guest lectures volunteered to participate. As a result, the observations related to the data collection activity reflect the experiences of this self-selected group, who may have been more motivated or interested in the topic than the broader class.

Second, the evidence presented in this work is primarily qualitative and based on the context of a single course. While student feedback and observations suggest that the activity can enhance engagement and provide insight into the challenges of data collection, the study was not designed to provide causal evidence of effectiveness. Our results may not generalize to other institutions, student populations, or course structures.

Finally, the activity combines multiple components, including guest lectures, sensory evaluation, and optional hands-on data collection. As a result, it is difficult to isolate the individual contribution of each component to the observed outcomes. Future work could explore these elements separately and in different instructional contexts.

8 Acknowledgments

We gratefully acknowledge the support of a New Instruction/New Initiative (NINI) grant from the College of Literature, Science, and the Arts at the University of Michigan. We also thank Dr. Alexander Wiltschko, founder and CEO of Osmo, for his machine olfaction guest lecture in STATS 315. M. K. K. served as a sensory evaluation consultant for the pilot design and participant training and was compensated for her time. She contributed

intellectually to the conception and design of the pilot and to writing the manuscript and is listed as a co-author.

9 Data Availability Statement

The Python code used for training the model on the GoodScents-Leffingwell dataset, making predictions on a sample dataset, and visualizing data, along with the trained model weights and a PDF version of the Google Form used for data collection, is available at `https://osf.io/jqn4y/`.

10 AI Use Statement

No generative artificial intelligence tools were used in the design, analysis, or substantive writing of this manuscript. Limited use of ChatGPT Instant 5.3 was made for minor language editing, polishing, and figure layout assistance.

References

Angeloni, S., A. M. Mustafa, D. Abouelenein, L. Alessandroni, L. Acquaticci, F. K. Nzekoue, R. Petrelli, G. Sagratini, S. Vittori, E. Torregiani, et al. (2021). Characterization of the aroma profile and main key odorants of espresso coffee. *Molecules 26*(13), 3856. DOI:10.3390/molecules26133856.

Baltrušaitis, T., C. Ahuja, and L.-P. Morency (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence 41*(2), 423–443. DOI:10.1109/TPAMI.2018.2798607.

Barsainyan, A. A., R. Kumar, P. Saha, and M. Schmuker (2023). Openpom - open principal odor map. `https://github.com/BioMachineLearning/openpom`.

Barua, A., M. U. Ahmed, and S. Begum (2023). A systematic literature review on multimodal machine learning: Applications, challenges, gaps and future directions. *IEEE Access 11*, 14804–14831. DOI:10.1109/ACCESS.2023.3243854.

Blikstein, P. (2013). Multimodal learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge*, pp. 102–106. DOI:10.1145/2460296.2460316.

Carver, R., M. Everson, J. Gabrosek, N. Horton, R. Lock, M. Mocko, A. Rossman, G. H. Roswell, P. Velleman, J. Witmer, et al. (2016). Guidelines for assessment and instruction in statistics education (gaise) college report 2016. `https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-(gaise)-reports`.

Chollet, F. (2021). *Deep learning with Python*. Manning. `https://www.manning.com/books/deep-learning-with-python`.

Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology innovations in statistics education 1*(1). DOI:10.5070/T511000028.

Consortium, D. C. (2024). Dream olfactory mixtures prediction challenge. `https://www.synapse.org/Synapse:syn53470621/wiki/626022`.

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education 7*(2). DOI:10.5070/T572013891.

Garcia-Ruiz, M. A., B. Kapralos, and G. Rebolledo-Mendez (2021). An overview of olfactory displays in education and training. *Multimodal Technologies and Interaction 5*(10), 64. DOI:10.3390/mti5100064.

Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl (2017). Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1263–1272. DOI:10.5555/3305381.3305512.

Guthrie, B. (2017). Machine olfaction. *Springer Handbook of Odor*, 55–56. `https://www.springerprofessional.de/en/machine-olfaction/12098570`.

Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: A review. *IEEE Sensors journal 2*(3), 189–202. DOI:10.1109/JSEN.2002.800688.

Hamel, E. A., 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 (2024). Pyrfume: A window to the world's olfactory data. *Scientific Data 11*(1), 1220. DOI:10.1038/s41597-024-04051-z.

Herz, R. S., E. Van Reen, D. H. Barker, C. J. Hilditch, A. L. Miller, R. Seifer, and M. A. Carskadon (2018). The influence of circadian timing on olfactory sensitivity. *Chemical Senses 43*(1), 45–51.

Hu, W., M. Fey, H. Ren, M. Nakata, Y. Dong, and J. Leskovec (2021). Ogb-lsc: A large-scale challenge for machine learning on graphs. In J. Vanschoren and S. Yeung (Eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, Volume 1. `https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/db8e1af0cb3aca1ae2d0018624204529-Abstract-round2.html`.

Hu, W., M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec (2020). Open graph benchmark: Datasets for machine learning on graphs. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 22118–22133. Curran Associates, Inc. DOI:10.5555/3495724.3497579.

Keller, A., R. C. Gerkin, Y. Guan, A. Dhurandhar, G. Turu, B. Szalai, J. D. Mainland, Y. Ihara, C. W. Yu, R. Wolfinger, et al. (2017). Predicting human olfactory perception from chemical features of odor molecules. *Science 355*(6327), 820–826. DOI:10.1126/science.aal2014.

Keller, A. and L. B. Vosshall (2016). Olfactory perception of chemically diverse molecules. *BMC neuroscience 17*, 1–17. DOI:10.1186/s12868-016-0287-2.

Kuiper, S., A. Chakraborty, T. George, L. W. Kay, L. M. Lesser, G. Rowell, D. Pearl, S. Crawford, and A. Olsen (2025). The greenhouse effect: Using student-generated agricultural data to warm up students for data-based decision making. *Journal of Statistics and Data Science Education 33*(2), 143–151.

Lee, B. K., E. J. Mayhew, B. Sanchez-Lengeling, J. N. Wei, W. W. Qian, K. A. Little, M. Andres, B. B. Nguyen, T. Moloy, J. Yasonik, et al. (2023). A principal odor map unifies diverse tasks in olfactory perception. *Science 381*(6661), 999–1006. DOI:10.1126/science.ade4401.

Leffingwell & Associates (2001). Pmp 2001 – database of perfumery materials and performance. `http://www.leffingwell.com/bacispmp.htm`.

Marino, M. A., F. Wisnewski, A. Alexis, and C. Reilly (2025). Multisensory learning in the undergraduate chemistry lab: The case for smell. *Journal of Chemical Education*. DOI:10.1021/acs.jchemed.4c01545.

Meilgaard, M. C., G. V. Civille, and B. T. Carr (2016). *Sensory Evaluation Techniques* (5th ed.). CRC Press.

Mu, S., M. Cui, and X. Huang (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors 20*(23), 6856. DOI:10.3390/s20236856.

Olofsson, J. (2025). *The Forgotten Sense: The New Science of Smell—and the Extraordinary Power of the Nose*. New York: HarperCollins. `https://www.harpercollins.com/products/the-forgotten-sense-jonas-olofsson`.

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

The Good Scents Company (2019). The good scents company information system. `http://www.thegoodscentscompany.com/`.

Appendix A Details about 20 Odorants

We quantified the purity and composition of the Flinn kit using Gas Chromatography – Mass Spectrometry and found all materials exceeded 90% purity, except amyl acetate

which contained a significant quantity of isoamyl acetate. To eliminate confusion, the amyl acetate was replaced with (R)-Limonene purchased from Sigma-Aldrich. Since we needed numerous kits for the number of students, we made our own using the chemicals, of analogous purity, purchased from Sigma-Aldrich. Approximately 100 mg of each chemical (approximately the same amount provided in the Flinn kit) was placed in a 4mL vial and diluted with 3mL of 190-proof biotech-grade alcohol, which does not contain any denaturing agents that could taint the odor.

Odorant	**CAS Registry Number**	**Non-stereo SMILES**
Ethyl methylphenylglycidate	7705-31-5	CCOC(=O)C1C(O1)(C)C2=CC=CC=C2
Isoamyl acetate	123-92-2	CC(C)CCOC(=O)C
2-Phenylethanol	60-12-8	OCCc1ccccc1
(S)-Carvone	648-41-5	CC(=C)C1CC=C(C)C(=O)C1
Vanillin	121-33-3	COC1=CC(C=O)=CC=C1O
(R)-Limonene	5989-27-5	CC1=CCC(CC1)C(=C)C
Undecalactone	10028-17-4	O=C1CCCCCCCCCCO1
Geraniol	106-24-1	CC(=CCCC(=CCO)C)C
(R)-Carvone	549-92-7	CC(=C)C1CC=C(C)C(=O)C1
Methyl salicylate	119-36-8	COC(=O)c1ccccc1O
1-Octen-3-ol	138-77-7	CCCCCC(C=C)O
Octyl acetate	112-14-1	CCCCCCCCOC(C)=O
Transcinnamaldehyde	100-43-0	O=C/C=C/C1=CC=CC=C1
Benzyl acetate	140-11-4	CC(=O)OCc1ccccc1
Menthol	89-78-1	CC(C)C1CCC(C)CC1O
Methyl ionone	127-51-5	CC(=O)C(C)=CC1C(C)=CCCC1(C)C
2-Phenylpropanol	122-60-1	CC(CO)c1ccccc1
Octalactone gamma	100-29-8	CCCCC1CCC(=O)O1
Acetic Acid	64-19-7	C(=O)O
(Trans) Anethole	4180-23-8	COc1ccc(C=CC)cc1

Table 1: The 20 selected odorants with their CAS registry numbers and corresponding non-stereo SMILES (without stereochemical descriptors).

Table 1 provides the standardized identifiers for the odorants used in this study. This selection ensured a controlled yet diverse range of odor profiles for student evaluation and data collection. Both CAS Registry Numbers and SMILES strings are included for each

odorant: the CAS numbers facilitate practical tasks such as ordering compounds from chemical vendors, while the SMILES strings support cheminformatics applications and molecular modeling in machine learning contexts.

Appendix B Course Structure & Module Placement

This appendix summarizes the structure of the undergraduate deep learning course in which the machine olfaction activity was implemented.

The course is organized into the following modules:

- **Foundations:** Linear regression, classification, loss functions, and gradient-based optimization.
- **Neural Networks:** Fully connected (dense) neural networks and model evaluation.
- **Convolutional Neural Networks:** Architectures and applications to image data.
- **Transformers:** Sequence modeling, attention-based architectures and applications to text data.
- **Machine Olfaction Module:** Guest lectures on sensory evaluation and machine olfaction, followed by a hands-on data collection activity.

The machine olfaction module is introduced in the final weeks of the course, after students have been exposed to core deep learning concepts. This placement allows the activity to build on students' prior understanding while illustrating how these methods extend to less traditional data modalities such as smell. A copy of the syllabus of the course is available at `https://osf.io/jqn4y/files/cq62b`.