



Cite this: *Digital Discovery*, 2025, 4, 3339

Dense Sense: a novel approach utilizing electron density augmented machine learning paradigm to understand the complex odour landscape

Pinaki Saha, ^{*a} Mrityunjay Sharma, ^{bcd} Sarabeshwar Balaji, ^e Aryan Amit Barsainyan, ^f Ritesh Kumar, ^{*bc} Volker Steuber ^a and Michael Schmuker ^a

Olfaction is a complex process where multiple nasal receptors interact to detect specific odorant molecules. Elucidating structure–activity–relationships for odorants and their receptors remains difficult since crystallization of the odor receptors is an extremely difficult process. Therefore, ligand-based approaches that leverage machine learning remain the state of the art for predicting odorant properties for molecules, such as the graph neural network approach used by Lee *et al.* In this paper we explored how information from quantum mechanics (QM) could synergistically improve the results obtained with the graph neural network. Our findings underscore the possibility of this methodology in predicting odor perception directly from QM data, offering a novel approach in the machine learning space to understand olfaction.

Received 23rd May 2025

Accepted 21st September 2025

DOI: 10.1039/d5dd00224a

rsc.li/digitaldiscovery

Introduction

Olfaction is a vital sense for perceiving the world that is crucial to the survival of many animals, *e.g.* in foraging, mating and detecting prey and predators. It also plays an important role in human life, *e.g.* to detect hazards or maintaining hygiene. Although it plays such an important role, the intricacies of the olfaction process are not well understood. In the case of olfaction in humans, olfactory perception involves about 802 genes which encode for the ORs, out of which 388 genes are functional receptors while the remaining 414 are reported to be pseudogenes.⁵

The exploration of these receptors using structural biology tools has been challenging due to their high genetic variability and limited expression in *in vitro* systems. Their instability during the isolation process makes them difficult to crystallize.^{1–3} Additionally, the binding of odorants to the ORs is not very straightforward. Individual ORs can recognize multiple odorants while a single odorant can elicit responses from multiple receptors, resulting in a complex scheme for odorant recognition.⁶ The complexity of odorant-OR binding impedes

structural based understanding of the odorant binding process. Nonetheless, there is some notable research going on concerning the structural biology aspect of olfaction. Utilizing Cryo-EM, researchers have been able to elucidate the structure of a single human olfactory receptor:⁷ OR51E2. In the case of OR51E2 receptor, size selectivity has been observed for carboxylic acid based ligands. Short chain linear carboxylic acids were shown to bind better to OR51E2 receptor compared to their long chain counterparts.³ Recently using the AI based homology modelling tool AlphaFold and molecular dynamics (MD) simulation, researchers have tried to elucidate the structure of different ORs present in the human nasal epithelium.⁸

In contrast to the structure based approach, we have the ligand based approach where we solely focus on features of molecules to predict olfactory properties. Recently Lee *et al.* reported success in utilizing message passing graph neural networks (MPGNNS) for predicting odor labels of various organic molecules.⁴ They achieved an impressive AUROC score of 0.894 using an ensemble of 50 GNNs. This is currently the state of art model whereas for a traditional machine learning approach (random forest) using Morgan fingerprints an AUROC of 0.85 was reported by them and this model was chosen as the baseline model in their paper.

MPGNNS have gained a lot of traction in the field of cheminformatics;⁹ this is due to the fact that they are able to directly input the two-dimensional structure of the molecule, thus enabling featureless learning for molecules and removing the need for calculation of molecular descriptors/features or fingerprints. MPGNNS treat molecules as graphs: the atoms are treated as nodes and the bonds are treated as edges.¹⁰ The message passing step collates all the information in the nodes

^aUniversity of Hertfordshire, UH Biocomputation Group, UK. E-mail: p.saha3@herts.ac.uk

^bCSIR-Central Scientific Instruments Organisation, Sector 30-C, Chandigarh-160030, India

^cAcademy of Scientific and Innovative Research (AcSIR), Ghaziabad-201002, India

^dDepartment of Higher Education, Shimla-171001, Himachal Pradesh, India

^eIndian Institute of Science Education and Research Bhopal (IISERB), Madhya Pradesh-462066, India

^fNational Institute of Technology Karnataka, Surathkal, Karnataka-575025, India



and edges to output an embedding which is fed to a feed forward neural network for either a classification or regression-based task. The embedding obtained post the message passing step can further be fine-tuned using backpropagation during training. This enriches the embedding with information pertinent to the model training. The embeddings thus tend to outperform traditional cheminformatics based molecular fingerprints which have a more generalized scope. Due to these advantages, graph neural networks have been steadily gaining traction in the field of cheminformatics.

Computational chemistry is another powerful tool to elucidate properties and structures of molecules. Previously, researchers have augmented computational chemistry with machine learning to create machine learning-based force fields that are trained on computational chemistry data, one notable example being the Behler–Parrinello neural network (BPNN).¹¹ Here, we augment computational chemistry with graph neural networks using the AIMLDM (atoms in molecule localization and delocalization matrices) approach.¹²

AIMLDM stems from the AIM (atoms in molecule) approach founded by Bader.¹³ This is a partitioning scheme that assigns a 3D region of molecular electron density to each constituent atom of the molecule. Partitioning of the electron density is possible because an atom does not completely lose its identity when it bonds to form a molecule. The electron density of a molecule (ρ) can be written as sum of the electron densities of its atomic constituents (ρ_i):

$$\sum_{i=1}^N \rho_i(\mathbf{r}) = \rho(\mathbf{r}) \quad (1)$$

In the AIM method, the atomic constituents are defined by the following boundary conditions:

$$\nabla \rho \cdot \mathbf{n} = 0 \quad (2)$$

The boundary conditions are given by the dot product of the gradient of electron density ($\nabla \rho$) and the normal vector to the gradient (\mathbf{n}). In case of a covalent bond, there will be an accumulation of electron density at the bonding region and hence $\nabla^2 \rho < 0$, while in case of ionic bonding there will be a depletion

of electron density at the bonding region as ionic bonding is purely an electrostatic interaction, thus $\nabla^2 \rho > 0$. The bonding region thus can be defined by the eqn (2) which in turn corresponds to the atomic boundary regions.

The surface integral of eqn (2) gives the flux for the gradient of electron density:

$$\oint_A dS_A \nabla \rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) d\mathbf{r} = 0 \quad (3)$$

Thus, in the AIM partitioning scheme, the atomic basin is constrained by the zero flux condition. The figure below highlights the treatment of the benzene molecule with the AIM approach (Fig. 2).

AIMLDM utilizes the AIM approach to create localization and delocalization indices for determining the extent of localization and delocalization in a molecular system. To understand localization and delocalization indices we need to understand the Fermi hole.¹⁵ The concept of Fermi hole comes from Pauli's exclusion principle, which states that two electrons of the same spin cannot exist in an orbital. The Fermi hole is thus a region of space surrounding an electron of spin $+\frac{1}{2}$ where an electron of spin $-\frac{1}{2}$ cannot exist. Mathematically, it can be represented in the following way:

$$\rho^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) = \rho^{\alpha}(\mathbf{r}_1) \{ \rho^{\alpha}(\mathbf{r}_1) + h(\mathbf{r}_1, \mathbf{r}_2) \} \quad (4)$$

$\rho_{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2)$ is the pair density of two electrons having the same spin state: α at positions \mathbf{r}_1 and \mathbf{r}_2 , respectively, while $\rho^{\alpha}(\mathbf{r}_1)$ and

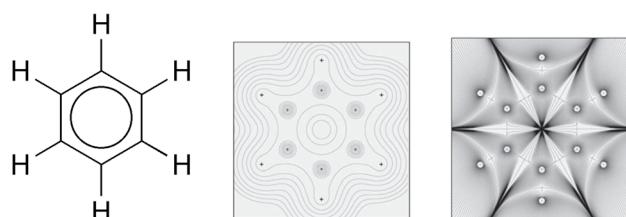


Fig. 2 Electron density contours of benzene and paths of the gradient vector field $\nabla \rho(\mathbf{r})$, highlighting atomic boundaries and atomic basins of benzene.¹⁴

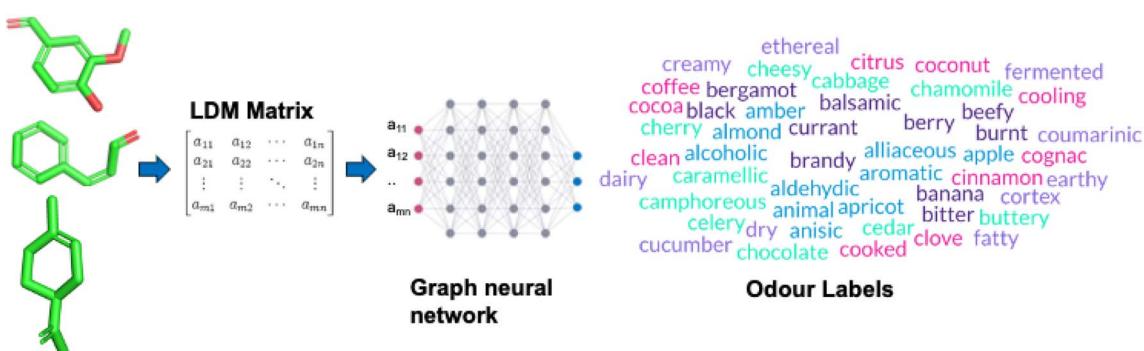


Fig. 1 Representation of QNN framework. The molecules are converted to LDM matrices, which are then fed to the graph neural network. The graph neural network then predicts the odour labels for the molecules.



$\rho^\alpha(\mathbf{r}_2)$ represent the one-electron density of electrons at positions \mathbf{r}_1 and \mathbf{r}_2 , respectively. The quantity $h(\mathbf{r}_1, \mathbf{r}_2)$ represents the Fermi hole. The Fermi hole follows the electron; if the electron is localized, so is the Fermi hole, and if the electron is delocalized the Fermi hole is also delocalized. Thus, the Fermi hole is useful for quantifying localization and delocalization in a molecule. Eqn (4) can be rearranged to the following:

$$\rho^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) = \rho^\alpha(\mathbf{r}_1) \cdot \rho^\alpha(\mathbf{r}_2) + \rho^\alpha(\mathbf{r}_1) \cdot h(\mathbf{r}_1, \mathbf{r}_2) \quad (5)$$

$$\rho^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) = \rho^\alpha(\mathbf{r}_1) \cdot \rho^\alpha(\mathbf{r}_1) + \rho_X(\mathbf{r}_1, \mathbf{r}_2) \quad (6)$$

The quantity $\rho_X(\mathbf{r}_1, \mathbf{r}_2)$ is termed the exchange density. Integrating the exchange density over atomic region A gives the number of same-spin electrons (having spin α) excluded from atomic region:

$$F^\alpha(A, A) = \iint_{\Omega_A} \rho_X(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (7)$$

$F^\beta(A, A)$ pertains to excluded β electron density. The localization index $\lambda(A)$ is defined as:

$$\lambda(A) = |F^\alpha(A, A)| + |F^\beta(A, A)| \quad (8)$$

while the delocalization index $\sigma(A, B)$ is defined as:

$$\sigma(A, B) = F^\alpha(A, A) + F^\beta(A, A) \quad (9)$$

In the AIMLDM approach, integration of the Fermi hole over the atomic regions which are obtained from the AIM partition is performed to get the localization and delocalization indices. We then get the LDM matrix $A_{i,j}$, where the diagonal elements are localization terms while the off diagonal elements are delocalization terms:

$$\mathbf{a}_{i,j=i} = \lambda(i, i) \quad (10)$$

$$\mathbf{a}_{i,j \neq i} = \sigma(i, j) \quad (11)$$

An example of the LDM matrix for the methane molecule (CH_4) is given by Table 1.

The diagonal terms which correspond to localization indices give an idea about distribution of electrons on the atoms while the off-diagonal terms which correspond to the delocalization indices give an idea regarding the bonding present in the molecule. The LDM matrix is conceptually similar to an

Table 1 The LDM matrix for methane is shown, the diagonal elements represent the localization indices while the off-diagonal terms represent the delocalization indices. The summation terms shows the electron count for each atom¹⁶

| | C ₁ | H ₂ | H ₃ | H ₄ | H ₅ | Σ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------|
| C ₁ | 4.04 | 0.492 | 0.492 | 0.492 | 0.492 | 6.007 |
| H ₂ | 0.492 | 0.444 | 0.021 | 0.021 | 0.021 | 0.998 |
| H ₃ | 0.492 | 0.021 | 0.444 | 0.021 | 0.021 | 0.998 |
| H ₄ | 0.492 | 0.021 | 0.021 | 0.444 | 0.021 | 0.998 |
| H ₅ | 0.492 | 0.021 | 0.021 | 0.021 | 0.444 | 0.998 |
| Σ | 6.007 | 0.998 | 0.998 | 0.998 | 0.998 | |

adjacency matrix used in graph representations, but it is not an actual adjacency matrix. This is because in a true adjacency matrix, the diagonal elements must be zero and any off-diagonal elements corresponding to unconnected node pairs must also be zero. In contrast, all elements in the LDM matrix are positive integers.

In MPGNNS the molecules are treated as graph with the atoms as nodes and bonds as edges while in our QNN approach we take the localization indices (diagonal terms) as node features while the delocalization indices (off diagonal terms) are taken as edge features (Fig. 1). This QNN (QM based neural network) model is then applied to a dataset containing 4872 molecules. The molecules were sourced from two datasets of odorants with odor labels: Goodscents and Leffingwell.^{17,18} Previously, an ensemble model¹⁴ consisting of 50 GNNs has been reported of having the highest accuracy score of 0.894 (AUROC) while the baseline performance of an individual ML model was reported to be 0.859. Our model surpasses the baseline performance of their reported individual ML model and is at par with their ensemble model.

Material and methods

The first step of creating AIMLDM matrices was utilizing ORCA¹⁹ to perform a single point energy calculation at B3LYP/6-311++G(d,p) level of theory on all the molecules present in the dataset. Single point energy calculation produces a wavefunction file which is then utilized by AIMall.²⁰ AIMall then performs the AIM partition calculation to produce sum files which contain the information of atomic basins, critical point, electron density and the gradient and the Laplacian of the electron density. These sum files are used for generating LDM matrices (Localization and delocalization matrices) using the AIMLDM software.¹² The matrices are generated as csv files. The matrices are then fed into the graph neural network. Our model architecture is based on Message Passing Neural Networks.²¹ For the message passing layer, we used 2 layers of edge-conditioned matrix multiplication with 63 hidden units and GRU updates, on top of a 45-dimensional atom featurization. In the readout layer, each atom's embedding is folded into its adjacent bond embeddings into a 152-dimensional embedding, and then summed to generate a molecule embedding. This molecule embedding is transformed through a 4-layer fully connected network, with decreasing layer sizes from 1024 to 256 and a final sigmoid function to make label predictions.

QM pipeline

(1) Perform single point energy calculation using ORCA (open source software). This generates binary .gbw files, use ORCA tool: orca_2aim to generate the wavefunction file (.wfx) from the binary files. It took approximately 40 days to finish the calculations for our dataset of 5 K molecules. These calculations were performed on the HPC cluster of the University of Hertfordshire.

(2) Use AIMall suite (commercial software with a free version which allows computing albeit at a slower processing time).

Run the AIMQB suite which is GUI based and browse for wavefunction files and execute them. This will produce the required .sum files. It took 20 days to finish the AIMall based calculations on our dataset with the free version (the paid version will take a smaller amount of time, paid version also supports parallelization). The computation was done on a Macbook Pro (M2 Max chip).

(3) AIMLDM is an open sourced software available at <https://www.cmatta.ca/software/>. Place the .sum files in the folder containing the AIMLDM executable file. Running the executable files produces a folder containing csv files which contains the LDM information. It took 3 days to finish the calculations on a Macbook Pro (M2 Max Chip).

Hyperparameter tuning

To optimize model performance, a systematic hyperparameter tuning strategy using the Optuna framework²² was done, which efficiently explored the search space through a combination of Bayesian optimization and pruning. Key parameters such as learning rate, batch size, weight decay, message passing steps, and feed-forward network configurations were varied during optimization. An early stopping mechanism was incorporated to prevent overfitting and reduce computation by halting training when the validation loss plateaued. Poor-performing runs were pruned dynamically and trials were evaluated using the ROC-AUC metric, and the tuning was parallelized across available CPU cores to accelerate the search process. The details of the hyperparameter values chosen are given in Table 2.

Explainability

Gradients quantify how infinitesimal changes in input features can affect the model predictions. One such attribution technique is Integrated Gradients (IG),²³ which calculates the integral of gradients along a path from a baseline input to the actual input in order to quantify the contribution of each input feature to a model's prediction. Gradients and their variations can therefore be leveraged to efficiently examine how feature changes affect the model's predicted results. IG satisfies key explainability axioms, such as sensitivity and implementation

invariance, making it well-suited for molecular feature attribution. Mathematically, the IG attribution for the i -th input feature is defined as:

$$\text{IntegratedGrads}_i(x) := (x_i - x'_i) \times \int_0^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha \quad (12)$$

where x represents the molecular input which is generally atomic or bond features, x' is the baseline input which is typically a zero vector or a neutral molecule representation, $F(x)$ is the model's output (odor prediction score), and $\frac{\partial F}{\partial x_i}$ is the gradient of the model's output with respect to the i -th feature. The integral is approximated using numerical summation over m steps:

$$\text{IntegratedGrad}_i(x) \approx (x_i - x'_i) \times \sum_{k=1}^m \frac{\partial F(x' + \frac{k}{m}(x - x'))}{\partial x_i} \times \frac{1}{m} \quad (13)$$

In our study, IG is adapted for graph neural networks (GNNs) in which we use the Directed message passing neural network (DMPNN)²⁴ as featurizer to analyze node and edge attributions. Node attributions are computed by perturbing atomic features while maintaining graph connectivity, identifying key atoms that contribute to odor perception.

By altering bond features, edge attributions are obtained, exposing some significant bonding interactions that help in predicting a particular odor. To ensure meaningful comparisons across molecules, attribution values are normalized. This approach enables the identification of functional groups and structural motifs associated with specific odor notes. Moreover, it offers a deeper understanding of how atomic-level features, such as atoms, bonds, and molecular structures, contribute to odor prediction.

Results and discussion

Our main objective is to compare our QNN model to the state of art MPGNN model that has been reported in the literature⁴ but unfortunately the GNN models have not been made available. We thus replicated their models by creating them in-house.²⁵ The GNN research replication work has been termed openPOM (Principal odor map). Principal odor map is a theoretical or computational framework used to visualize and categorize different odors based on their chemical properties and perceived characteristics. Lee *et al.*⁴ have utilized principal component analysis (PCA) of GNN embeddings to generate an odor map where similar odors are clustered together and dissimilar ones are placed further apart. The map is helpful for predicting the odor of a new or uncharacterized molecule on their map position. We recreated the POM reported in the literature and made it open source; the model is deposited in the Deep Chem repository, which is an open source repository

Table 2 Best hyperparameter values chosen for DMPNN + LDM model

| Hyperparameter | Search space | Chosen value |
|-----------------------------|--|-----------------|
| Initial_lr_rate | $[1e^{-4}, 1e^{-2}]$ | 0.00539 |
| Decay rate | $[0.1, 0.9]$ | 0.77710 |
| Decay steps | $[764, 384, 96]$ | 764 |
| Batch size | $[16, 25, 32]$ | 25 |
| Loss aggr type | [sum, mean] | sum |
| Num step message passing | $[2, 5]$ | 5 |
| Message aggregator type | [sum, mean, max] | max |
| ffn_hidden_list | $[[256, 256], [392, 392], [512, 512]]$ | |
| ffn_dropout_p | $[0.0, 0.5]$ | 0.1545 |
| ffn_dropout_at_input_no_act | [True, False] | False |
| Weight decay | $[1e^{-6}, 1e^{-4}]$ | $1.39161e^{-6}$ |



of cheminformatics based tools <https://deepchem.io/tutorials/predict-multi-label-odor-descriptors-using-openpomp/>

We utilized the same 138 odor labels that were used for generating the principal odor map by Lee *et al.*⁴ The odor labels used were: alcoholic, aldehydic, alliaceous, almond, amber, animal, anisic, apple, apricot, aromatic, balsamic, banana, beefy, bergamot, berry, bitter, black currant, brandy, burnt, buttery, cabbage, camphoreous, caramellic, cedar, celery, chamomile, cheesy, cherry, chocolate, cinnamon, citrus, clean, clove, cocoa, coconut, coffee, cognac, cooked, cooling, cortex, coumarinic, creamy, cucumber, dairy, dry, earthy, ethereal, fatty, fermented, fishy, floral, fresh, fruit skin, fruity, garlic, gassy, geranium, grape, grapefruit, grassy, green, hawthorn, hay, hazelnut, herbal, honey, hyacinth, jasmine, juicy, ketonic, lactonic, lavender, leafy, leathery, lemon, lily, malty, meaty, medicinal, melon, metallic, milky, mint, muguet, mushroom, musk, musty, natural, nutty, odorless, oily, onion, orange, orangeflower, orris, ozone, peach, pear, phenolic, pine, pineapple, plum, popcorn, potato, powdery, pungent, radish, raspberry, ripe, roasted, rose, rummy, sandalwood, savory, sharp, smoky, soapy, solvent, sour, spicy, strawberry, sulfurous, sweaty, sweet, tea, terpenic, tobacco, tomato, tropical, vanilla, vegetable, vetiver, violet, warm, waxy, weedy, winey, woody.

The data were sourced from Goodscents and Leffingwell datasets^{17,18} each containing odorant molecules and corresponding odor descriptors. Datasets were carefully curated to match, as closely as possible, the (unpublished) datasets that were used by Lee *et al.*⁴ The full process is documented in the code repository. Post curation we got 4872 odorants. For splitting the dataset into train, valid and test sets, a stratified splitting technique was used to ensure the distribution of chemotypes was the same in the training, validation and test dataset thus removing imbalance in the chemical dataset. The dataset was split as follows:

80% training : 10% validation : 10% test.

To counter imbalance in the label side, we calculated class imbalance ratios so that each odor label's contribution to the loss was weighted by a factor of $\log(1 + \text{class imbalance ratio})$. Class imbalance ratio for a particular odor label is defined as ratio of frequency of that label to the frequency of the most prevalent odor label. This ensures that rarer odor labels are given higher weights. For model training, we set the number of epochs to 30. The loss function utilized for training was softmax based cross entropy and the AUROC metric was used to assess the model. We utilized MPGNN with atom features and bond features which were mentioned by Lee *et al.*⁴ and finally the following features were utilized:

Atom features: valence, bonded neighbours count, hydrogen count, hybridization, formal charge.

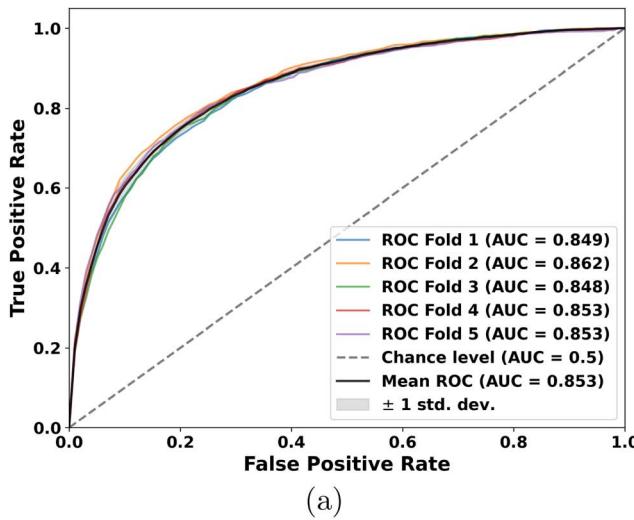
Bond features: aromaticity, cyclicity, degree of bonding.

Utilizing these features we trained the MPGNN and got the results as shown in Table 4.

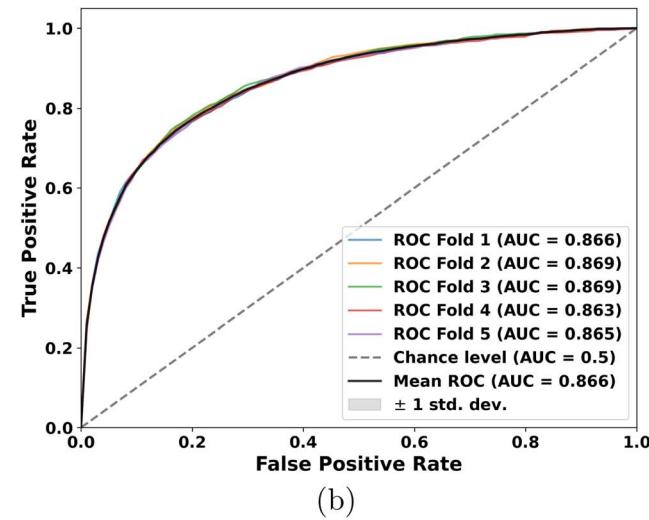
The training was completed within 2 hours and 32 minutes on a system that had hardware specification as mentioned in Table 3. We applied five-fold cross-validation, consistent with the approach used in the principal odor map study⁴ by Lee *et al.* In this paper they have not reported the performance of the

Table 3 Specifications of Hardware Resources

| Component | Specification |
|------------------|---|
| Operating System | Ubuntu 18.04.6 LTS (x86_64) |
| Kernel | 5.4.0-150-generic |
| Host | DS400TG-48R Intel |
| CPU | Intel Xeon Silver 4214 (48 cores, 3.20 GHz) |
| GPU | NVIDIA Tesla V100 PCIe 32 GB |
| Memory | 10193 MiB/128568 MiB RAM |



(a)



(b)

Fig. 3 ROC values for both DMPNN + LDM and openPOM 5 folds. The figures show ROC values separately for each fold and also the mean value. (a) ROC value for the DMPNN + LDM model. The coloured line shows the mean ROC value. (b) ROC value for the openPOM model. The coloured line shows the mean ROC value.

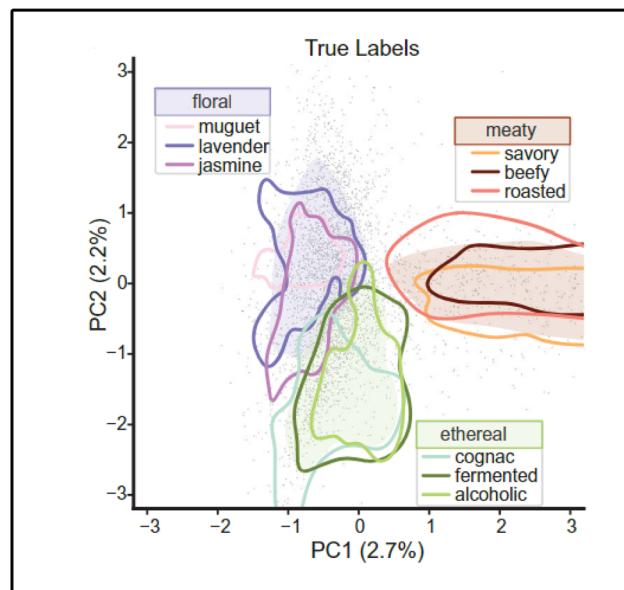


Table 4 Results of the MPGNN training process

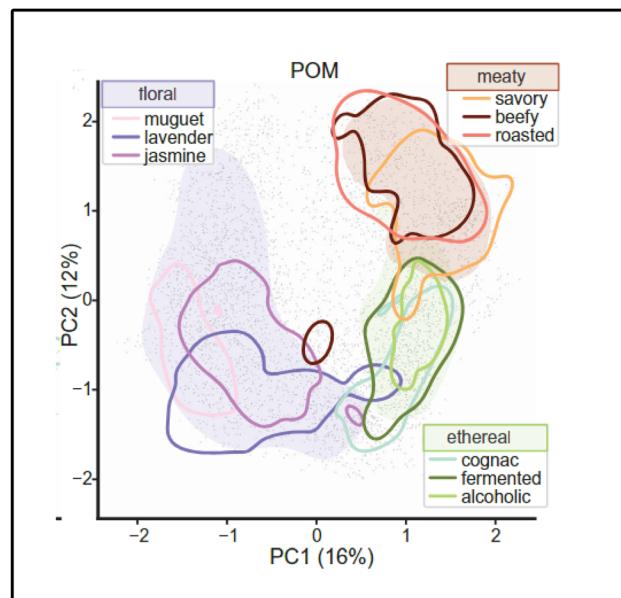
| Dataset | AUROC |
|------------------------|---------------------|
| Validation (5-fold CV) | 0.8661 ± 0.0015 |
| Validation | 0.864 |
| Test | 0.871 |

individual MPGNN model but nonetheless it is safe to assume that the accuracy obtained for single MPGNN will be between the accuracy (5 fold CV) reported for the baseline traditional ML model and the accuracy (5 fold CV) reported for the ensemble model:

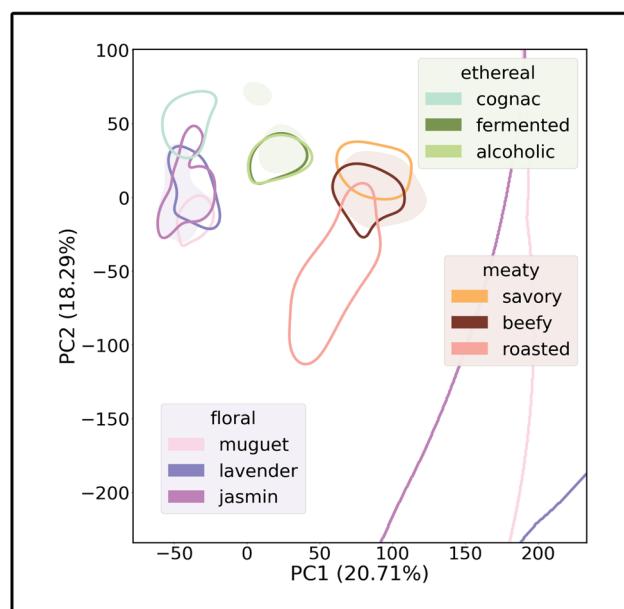
$$0.894 \text{ (ensemble)} > \text{single MPNN accuracy} > 0.859 \text{ (baseline)}$$



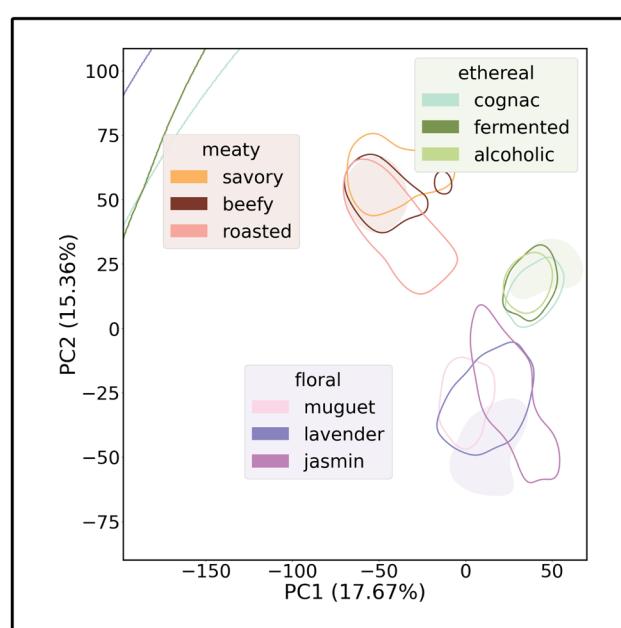
(a)



(b)



(c)



(d)

Fig. 4 Areas dense with molecules that have the broad category labels floral, meaty, or alcoholic are shaded; areas dense with narrow category labels are outlined. The principal odour map (a) recapitulates the true perceptual map⁴ (b) POM coordinates (256 dimensions)⁴ (c) Principal odour map replicated using our DMPNN + LDM (d) Principal odour map replicated using our DMPNN + Morgan fingerprint; note that only relative (not absolute) coordinates matter. Figure (a) and (b) are reproduced with permission from Science (<https://doi.org/10.1126/science.ade4401>).



The accuracy of our single MPGNN model (0.864, 5-fold CV) falls well within this range (Fig. 3). Next, we attempted to replicate the Principal Odor Map reported by Lee *et al.* by plotting the principal components of their GNN embeddings. Fig. 4c presents the principal odor map we reconstructed using embeddings from our trained graph neural network. The AUROC scores obtained on DMPNN + LDM and the replication of POM points to the fact that we have successfully recreated MPGNN utilized by Lee *et al.*⁴ We then utilized it for benchmark analysis of our QNN model.

The dataset of approximately 5000 molecules, which was used for training the openPOM, was also utilized for training the QNN model. The first step involved the calculation of LDM matrices for these molecules. This was achieved by performing a single-point energy calculation on these molecules and generating a wavefunction file containing their wavefunction information. The single-point energy calculation was performed at the B3LYP/6-311++G(d,p) level of theory.²⁶ This choice is justified because our dataset consists solely of organic molecules, and the B3LYP functional has been extensively used for studying the structures of organic molecules.²⁷ 6-311 is triple zeta basis set which offers higher flexibility for describing valence electrons. Polarization functions (d, p) are essential for capturing the anisotropic electron distribution. Diffuse functions (++) are important for accurately modeling electron-rich regions and long-range interactions. There do exist other functionals which also would have been suitable for our dataset like range-separated functionals with 100% asymptotic exchange.^{28,29} These functionals are designed to improve upon standard DFT approximations by more accurately modeling the long-range behavior of electron exchange interactions which could potentially occur during odorant-receptor binding. We nonetheless utilized the B3LYP/6-311++G(d,p) basis set due to the dual advantage of accuracy and computational speed. The basis set is suitable for our dataset size of 5000 molecules (which is the largest available open source dataset for labelled odorants) but it would not be feasible for a dataset containing millions of compounds. We can reduce the basis set to increase the computational speed but then we tradeoff the accuracy. We hypothesized a solution where promolecule density³⁰ can be used to tackle the scalability issue, a promolecule density can be calculated on the fly for molecules as it is generated by combining precalculated atomic densities. This promolecule density then can be utilized for olfaction based machine learning.

The wavefunction files so obtained from single point calculations were then utilized for the creation of LDM matrices using the software codes AIMall and AIMLDM.exe. The dataset of LDM matrices was then used for training the graph neural network. The graph neural network was used to predict the odor label for the molecules. We utilized the same 138 odor labels that were used for openPOM. The dataset was split in the following way: 80 : 10 : 10 (training : validation : test). To counter imbalance in the label side, we calculated class imbalance ratios to get each odor label's contribution to the loss being weighted by a factor of $\log(1 + \text{class imbalance ratio})$. Class imbalance ratio for a particular odor label is defined as ratio of frequency

of that label to the frequency of the most prevalent odor label. This ensures that rarer odor labels are given a higher weighting.

The optimization process utilized a cross-entropy loss function, and the model performance was evaluated using the AUROC metric. We utilized various architecture of graph neural networks: message passing neural networks (MPNN), directed message passing neural net (DMPNN) and graph convolutional neural networks (GCN). Using solely the LDM indices for training the various GNNs we got the following result as shown in Table 5. The QNN training performance using molecular fingerprints as input features (40 epochs) is shown in Fig. 6.

The results of the QNN model are impressive for this challenging odor label classification task given the fact that only electron localization and delocalization data were provided to the QNN model. The DMPNN + LDM model gives us the best results amongst all the GNNs by achieving a validation score of 0.871, which is better than the openPOM result. We also created a perceptual map for the DMPNN + LDM model. Fig. 4 demonstrates that the replicated Dense-Sense model preserves the perceptual organization of odor space established by the Principal Odor Map (POM). Similar to the true perceptual map (a) and the original POM embedding (b), the replicated model (c) produces well-separated clusters for broad odor categories (floral, meaty, alcoholic) and their associated subcategories, indicating that the model effectively captures human odor similarity relationships from molecular structure.

The baseline model reported by Lee *et al.* shows an accuracy of 0.859. Their baseline model is a random forest model that uses Morgan fingerprints. We also deployed molecular fingerprints and descriptors in the graph neural network framework to see the comparison of their performance with our QM model. We tried both DMPNN and MPNN architectures with the MACCS/Morgan fingerprints and RDkit descriptors. The test and validation AUROC score for the fingerprints and descriptors based GNN models were similar to the score obtained for DMPNN + LDM model (Table 6) and the perceptual map obtained for the Morgan fingerprint had similar contours as the one obtained for the DMPNN + LDM model (Fig. 4). The POM model is a GNN framework which incorporates atomic and bond features; inclusion of fingerprints and descriptors did not lead to a drastic increase in the AUROC score. This could be due to the fact that the descriptors and fingerprints do not encode any additional olfaction specific information which would aid the model. To further enhance the classification performance in this multilabel setting, we optimized per-label thresholds for

Table 5 Results of the QNN training process. DMPNN + LDM neural network provides best results amongst all the GNN as per the validation score and is competitive with openPOM as per test score

| GNN architectures (40 epochs) | Validation (AUROC) | Test (AUROC) |
|-------------------------------|--------------------|--------------|
| MPNN + LDM | 0.8383 | 0.8499 |
| GCN + LDM | 0.849 | 0.85 |
| DMPNN + LDM | 0.871 | 0.86 |
| openPOM | 0.864 | 0.871 |



converting predicted probabilities into binary outputs (Table 7). With these thresholds applied to the validation predictions, the model achieved a macro-averaged precision of 0.4244, recall of 0.4615, and F1 score of 0.3853. The OpenPOM model achieved similar results: a macro-averaged precision of 0.4387, recall of 0.5039 and F1 Score of 0.4156. These metrics are reasonable for a high label, imbalanced and noisy dataset. We then employed an ensemble approach to combine graph neural networks for improved performance. We first explored combining DMPNN models, as they demonstrated the best results. We tested ensembles of 10 and 30 DMPNNs and aggregated their result by averaging out their predictions. The AUROC metric was used to evaluate the model performance. We tested two cases where we varied random seeds and one without random seed variation. Another approach involved creating ensembles of different graph-based neural network architectures, with odor label predictions aggregated using the mean. DMPNN + LDM model, openPOM model (MPNN) and GCN + LDM were utilized for this approach. The results are shown in Table 8.

The best ensemble training result was achieved with an ensemble of ten DMPNNs as shown in Table 9. After applying per-label thresholds to the validation predictions, the model attained a macro-averaged precision of 0.4806, recall of 0.4862, and F1 score of 0.4247. Although our ensemble result does not surpass the result for the ensemble model (AUROC: 0.894) reported by Lee *et al.*,⁴ our ensemble is at par with their ensemble approach (AUROC: 0.88). Recently, Burns *et al.*³¹ reported the application of graph neural networks for the prediction of olfaction labels. They utilized the coordinates of three dimensional structures of molecules as input and added additional features to their graph neural network: Hirshfeld charge and Hirshfeld volume. The model showed promising results but at significantly higher epochs as compared to our models or the POM model reported by Lee *et al.*⁴ Burns *et al.* reported AUROC

Table 6 Results of QNN training using fingerprint-based features over 40 epochs

| GNN architectures (40 epochs) | Validation (AUROC) | Test (AUROC) |
|--|--------------------|--------------|
| DMPNN + MACCS | 0.872 | 0.862 |
| DMPNN + Morgan (radius = 2, size = 1024)) | 0.872 | 0.866 |
| DMPNN + RDkit | 0.873 | 0.863 |
| MPNN + MACCS | 0.876 | 0.870 |
| MPNN + RDkit | 0.873 | 0.867 |

Table 7 Results of the QNN 5 fold results of the training process. DMPNN + LDM neural network provides best results among all the GNN as per the validation score and is competitive with openPOM as per test score

| GNN architectures (40 epochs) | Validation (5-fold CV) |
|-------------------------------|------------------------|
| MPNN + LDM | 0.8405 ± 0.0018 |
| GCN + LDM | 0.8485 ± 0.0041 |
| DMPNN + LDM | 0.853 ± 0.0034 |

Table 8 Results of the ensemble training process of GNNs having different architectures

| Ensemble | Random seed | Validation (AUROC) | Test (AUROC) |
|--------------------|-------------|--------------------|--------------|
| DMPNN + MPNN + GCN | 1 | 0.869 | 0.865 |
| DMPNN + MPNN + GCN | 42 | 0.867 | 0.859 |

Table 9 Results of the ensemble training process of DMPNNs. The best Validation and test score was achieved by ensemble of ten DMPNNs

| Ensemble | Validation (AUROC) | Test (AUROC) |
|-------------------------|--------------------|--------------|
| 10 DMPNNs (random seed) | 0.857 | 0.872 |
| 10 DMPNNs | 0.88 | 0.875 |
| 30 DMPNNs (random seed) | 0.857 | 0.869 |
| 30 DMPNNs | 0.88 | 0.875 |

of 0.874 and 0.865 for their validation and test set respectively, but at 400 epochs. At 40 epochs the AUROC for validation is shown to be approximately around 0.65, which is significantly less than the results reported by us or Lee *et al.*⁴ at 40 epochs. Also, a fair comparison cannot be warranted between their and our models as their dataset is smaller (~3500 molecules) compared to our dataset of ~5000 molecules.

Explainability

The best way to validate explainability was by using an explainability analysis for compounds with functional group-based odor labels, such as ketonic, phenolic, *etc.* These labels are ideal for explainability analysis because their structural basis is well understood. The explainability analysis must then highlight the functional group as the odorganic region of the molecule.

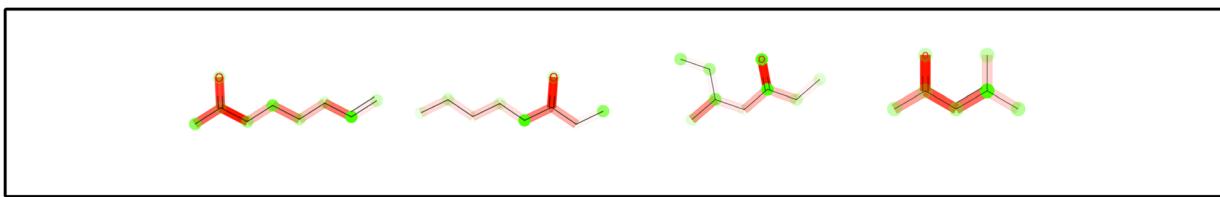
Our explainability analysis confirmed this. We took the functional group odor labels that our model was able to successfully predict (AUROC per label score >0.8). Thus, we obtained ketonic, phenolic, sulfurous, and lactonic-based odor labels. The results of the explainability analysis are as follows:

Ketonic odor label (AUC per label score: 0.94). In case of compounds with ketonic labels, the explainability analysis shown in Fig. 5a correctly predicts the carbonyl group as the main contributor regarding olfaction. In case of the first molecule in Fig. 5a. The ketone region is correctly highlighted by the explainability analysis.

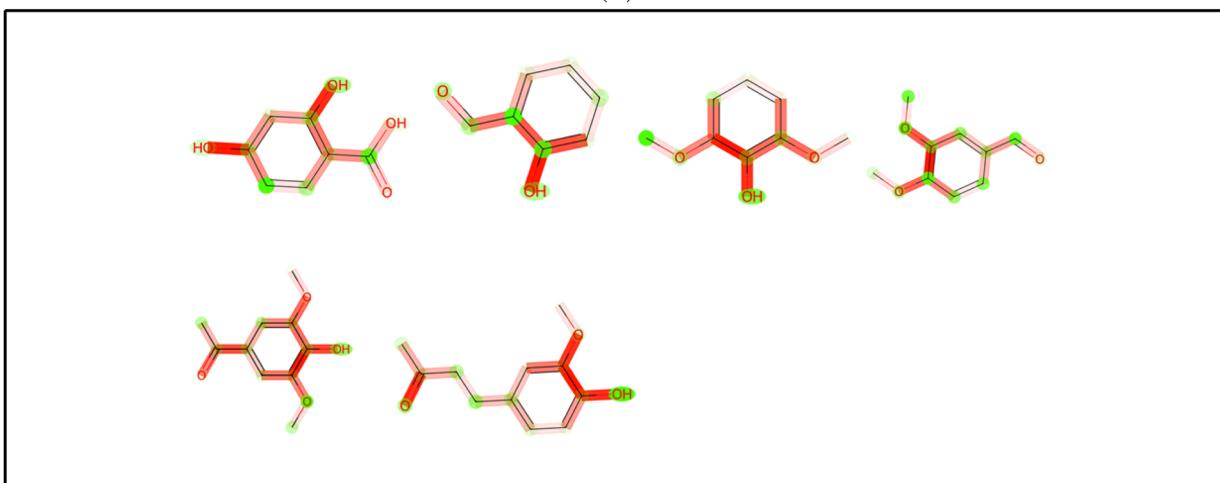
Phenolic odor label (AUC per label score: 0.89). In case of compounds with phenolic labels, the explainability analysis correctly predicts the hydroxy group of phenol as the main contributor regarding olfaction. In some molecules it also highlights the carbonyl group attached to the aromatic ring; this may be due to the fact that the enolic tautomer of such a carbonyl will be electronically similar to phenolic OH as the



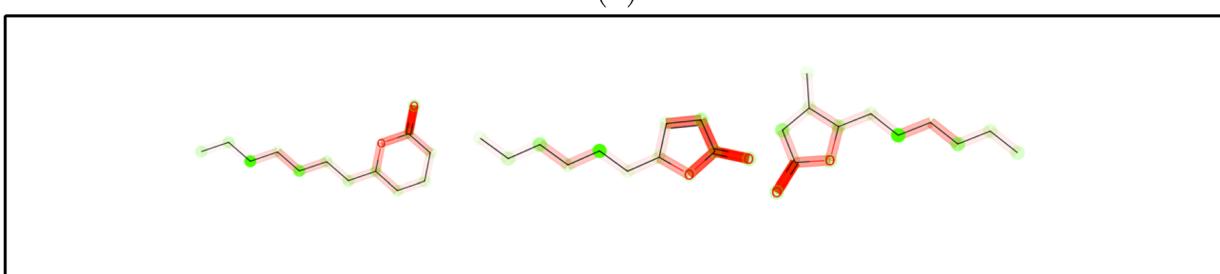
Open Access Article. Published on 08 October 2025. Downloaded on 1/23/2026 10:44:39 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

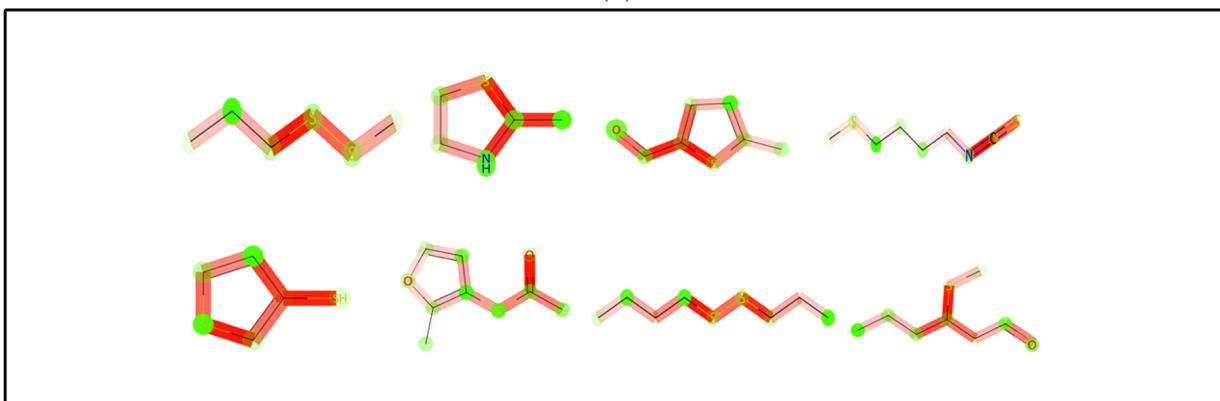
(a)



(b)



(c)



(d)

Fig. 5 Explainability analysis of compounds with different odour labels: (a) ketonic, (b) phenolic, (c) lactonic, and (d) sulfurous. The importance of bonds towards olfaction is highlighted using color gradation, where red indicates high importance and pink indicates low importance. In the case of atoms, dark green atoms have the highest importance, while light green have the lowest.

enolic form in this case will be a hydroxy group conjugated with an aromatic system.

Lactonic odor label (AUC per label score: 0.92). In case of the compounds with lactonic labels, the explainability analysis in Fig. 5c correctly predicts the lactone group (aliphatic cyclic ester) as the main contributor to olfaction.

Sulfurous odor label (AUC per label score: 0.97). In case of the compounds with sulfurous labels, the explainability

analysis as shown in Fig. 5d correctly predicts the thio group/dithia linkage as the main contributor to olfaction.

Post this validation exercise we looked at odor labels which had a more complex relationship with the structure of odorant:

Vanilla odor label. The major chemical compound responsible for vanilla odor is the vanillin molecule. The structure of the vanillin molecule is given in Fig. 6a. The explainability analysis in this case highlights the importance of phenol and *para*-carbonyl group and *ortho* hydroxy groups. The vanillin

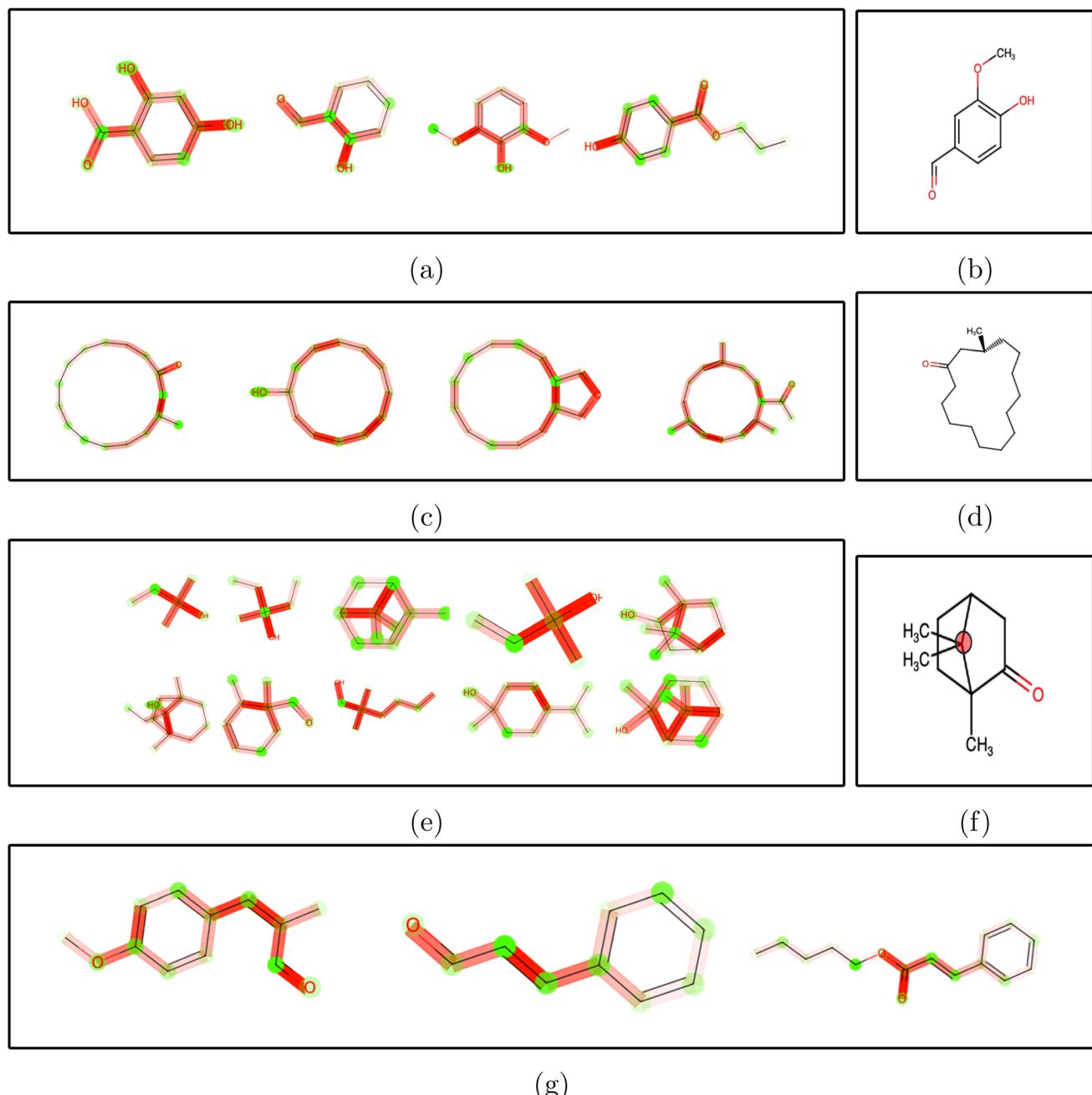


Fig. 6 Explainability analysis of compounds with different odour labels: (a) vanillin (b) Structure of vanillin, (c) musk, (d) Structure of muscone molecule, and (e) camphoreous (f) structure of camphor molecule. Red circle highlights the quaternary bridgehead carbon, and (g) cinnamon. The importance of bonds towards olfaction is highlighted using color gradation, where red indicates high importance and pink indicates low importance. In the case of atoms, dark green atoms have the highest importance, while light green have the lowest.



molecule contains both these motifs. We further performed our analysis on other odor labels (results given in SI). We also utilize explainability analysis where a complex relationship exist between the odor and the odorant molecule:

Musk odor label. The explainability analysis in this case highlights the importance of the macrocyclic ring for the musk odor label. The macrocyclic ketone and heterocyclic oxygen group are also deemed significant. Natural musk odor is primarily due to the constituent called muscone which is a macrocyclic ketone as shown in Fig. 6d. As per the literature³² OR5AN1 is a human musk-recognizing receptor, which binds with muscone *via* hydrogen-bond formation from the tyrosine-260 residue along with hydrophobic interactions with surrounding aromatic residues on the receptor. The keto group and the macrocyclic ring thus play an important role in imparting musk odor; this is corroborated *via* the explainability analysis.

Camphoreous odor label. In case of camphoreous molecules the explainability analysis highlights the importance of the quaternary carbon, especially the quaternary carbon bridgehead for the camphoreous smell. The bridgehead carbon is a prominent feature of the camphor molecule as shown in Fig. 6f.

Cinnamon odor label. The explainability analysis in this case highlights the importance of conjugation of the olefin bond with a carbonyl group. Cinnamaldehyde (second molecule in Fig. 6g) is the molecule which imparts aroma to the cinnamon spice.

Explainability gives a special edge to our model in designing a new fragrance/odorant. We can specifically design novel molecules by conserving the odorogenic regions given by the explainability analysis; this can be a useful utility for researchers working in the generative AI field in olfaction.

Conclusion

Lee *et al.*'s seminal work on the principal odor map is a big step towards the digitization of smell. We have replicated their principal odor map and created a MPGNN model with an accuracy of 0.86 (AUROC). We have made the model and our codes open source to help olfaction researchers benchmark their work. We utilized this model to benchmark our novel quantum mechanics (QM) data-based neural networks (DENSE SENSE). We have successfully created a model which works solely on QM based features and shows high accuracy on the odor dataset. The QM data is obtained from LDM matrices which contain localization and delocalization indices of a molecule. These have been used by GNN to accurately predict odor labels for our dataset (best model AUROC: 0.87). Our QM based GNN model performs better than the conventional GNN model. In case of ensemble modelling, our QM based GNN model (best model AUROC: 0.88) is at par with the conventional GNN model (AUROC: 0.894). Our explainability analysis performed in this paper gives unique insight regarding the contribution of molecular features to olfaction. This is a novel machine learning approach of utilizing QM based data. Our work paves the way for a novel machine learning paradigm that can be utilized for various prediction tasks.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

All data and code supporting this study are openly available in public repositories. The machine learning codes and analysis scripts for DenseSense are available at Figshare. <https://doi.org/10.6084/m9.figshare.30237730>. The machine learning training data and results for OpenPOM are available at Kaggle. <https://doi.org/10.34740/kaggle/dsv/6447845>.

Supplementary information is available. See DOI: <https://doi.org/10.1039/d5dd00224a>.

References

- Y. Waltenspühl, J. Ehrenmann, C. Klenk and A. Plückthun, Engineering of challenging G protein-coupled receptors for structure determination and biophysical studies, *Molecules*, 2021, **26**, 1465.
- D. M. Thal, Z. Vuckovic, C. J. Draper-Joyce, Y.-L. Liang, A. Glukhova, A. Christopoulos and P. M. Sexton, Recent advances in the determination of G protein-coupled receptor structures, *Curr. Opin. Struct. Biol.*, 2018, **51**, 28–34.
- Q. Zhao and B.-l. Wu, Ice breaking in GPCR structural biology, *Acta Pharmacol. Sin.*, 2012, **33**, 324–334.
- B. K. Lee, 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.*, A principal odor map unifies diverse tasks in olfactory perception, *Science*, 2023, **381**, 999–1006.
- Y. Niimura and M. Nei, Evolution of olfactory receptor genes in the human genome, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12235–12240.
- M. Zarzo, The sense of smell: molecular basis of odorant recognition, *Biol. Rev.*, 2007, **82**, 455–479.
- C. B. Billesbølle, C. A. de March, W. J. van der Velden, N. Ma, J. Tewari, C. L. Del Torrent, L. Li, B. Faust, N. Vaidehi, H. Matsunami, *et al.*, Structural basis of odorant recognition by a human odorant receptor, *Nature*, 2023, **615**, 742–749.
- B. Berwal, P. Saha and R. Kumar, Improving Olfactory Receptor Structure Modeling *via* Hybrid Methods, *bioRxiv*, 2024, pp. , pp. 2024–05.
- P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, *et al.*, Graph neural networks for materials science and chemistry, *Commun. Mater.*, 2022, **3**, 93.
- J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, *Machine learning meets quantum physics*, Springer, 2020; pp. pp. 199–214.
- J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.*, 2007, **98**, 146401.
- I. Sumar, R. Cook, P. W. Ayers and C. F. Matta, AIMLDL: a program to generate and analyze electron localization-



delocalization matrices (LDMs), *Comput. Theor. Chem.*, 2015, **1070**, 55–67.

13 R. F. Bader, Atoms in molecules, *Acc. Chem. Res.*, 1985, **18**, 9–15.

14 N. Sukumar, *A Matter of Density: Exploring the Electron Density Concept in the Chemical, Biological, and Materials Sciences*, John Wiley & Sons, 2012.

15 P. Saha, *Application of Electron Density Based Analysis in the Study of Nanoclusters and Biomolecular Interactions*, Shiv Nadar University Thesis, 2018.

16 I. Sumar, *Chemical Applications of Electron Localization-Delocalization Matrices (LDMs) with an Emphasis on Predicting Molecular Properties*, 2016.

17 W. Luebke, *The Good Scents Company Information System*, 2019, <http://www.thegoodsentscompany.com>.

18 J. C. Leffingwell, *Leffingwell & Associates*, 2005, <http://www.leffingwell.com>.

19 F. Neese and *et al.*, Orca. An *ab initio*, density functional and semiempirical program package version, 2009, vol. 2.

20 T. A. Keith and *et al.*, AIMAll, version, 2017, vol. 19, p. 12.

21 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, Neural message passing for quantum chemistry, *International conference on machine learning*, 2017, pp. 1263–1272.

22 T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

23 M. Sundararajan, A. Taly and Q. Yan, Axiomatic attribution for deep networks, *International conference on machine learning*, 2017, pp. 3319–3328.

24 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, *J. Chem. Inf. Model.*, 2019, **59**, 3370–3388.

25 A. A. Barsainyan, R. Kumar, P. Saha and M. Schmuker, *Predict Multi-Label Odor Descriptors using OpenPOM*, OpenPOM, 2023.

26 P. Stephens, F. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **1994**, 98.

27 J. Tirado-Rives and W. L. Jorgensen, Performance of B3LYP density functional methods for a large set of organic molecules, *J. Chem. Theor. Comput.*, 2008, **4**, 297–306.

28 V. Tognetti and L. Joubert, On Atoms-in-Molecules Energies from Kohn-Sham Calculations, *ChemPhysChem*, 2017, **18**, 2675–2687.

29 L. N. Anderson, M. B. Oviedo and B. M. Wong, Accurate electron affinities and orbital energies of anions from a nonempirically tuned range-separated density functional theory approach, *J. Chem. Theory Comput.*, 2017, **13**, 1656–1666.

30 P. Saha and M. T. Nguyen, Electron density mapping of boron clusters *via* convolutional neural networks to augment structure prediction algorithms, *RSC Adv.*, 2023, **13**, 30743–30752.

31 J. W. Burns and D. M. Rogers, QuantumScents: Quantum-mechanical properties for 3.5 k olfactory molecules, *J. Chem. Inf. Model.*, 2023, **63**, 7330–7337.

32 L. Ahmed, Y. Zhang, E. Block, M. Buehl, M. J. Corr, R. A. Cormanich, S. Gundala, H. Matsunami, D. O'Hagan, M. Ozbil, *et al.*, Molecular mechanism of activation of human musk receptors OR5AN1 and OR1A1 by (R)-musccone and diverse other musk-smelling compounds, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E3950–E3958.

