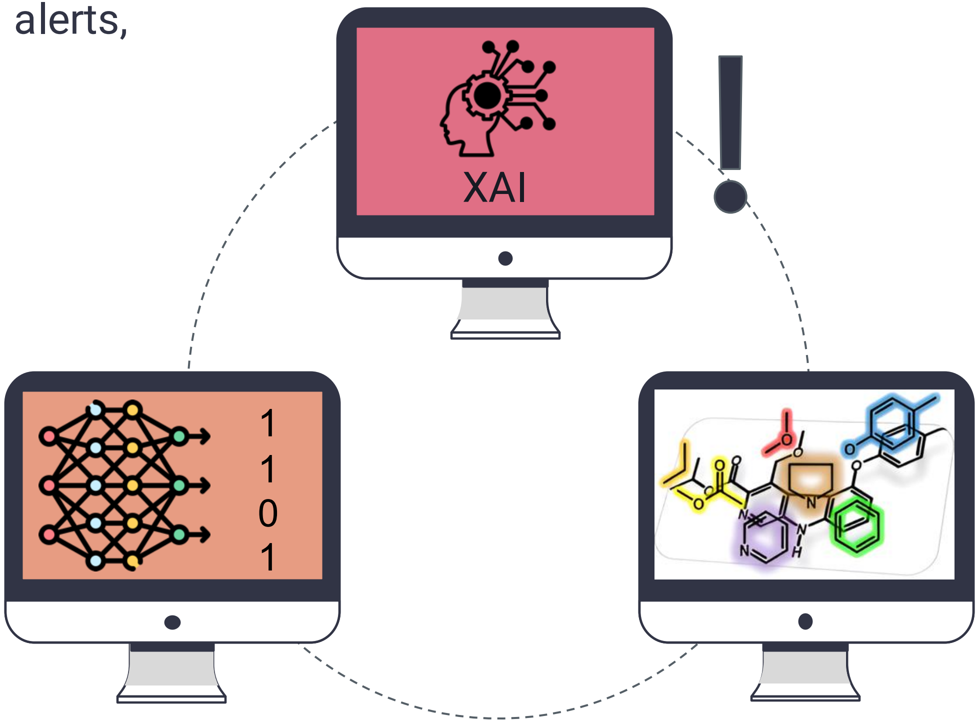
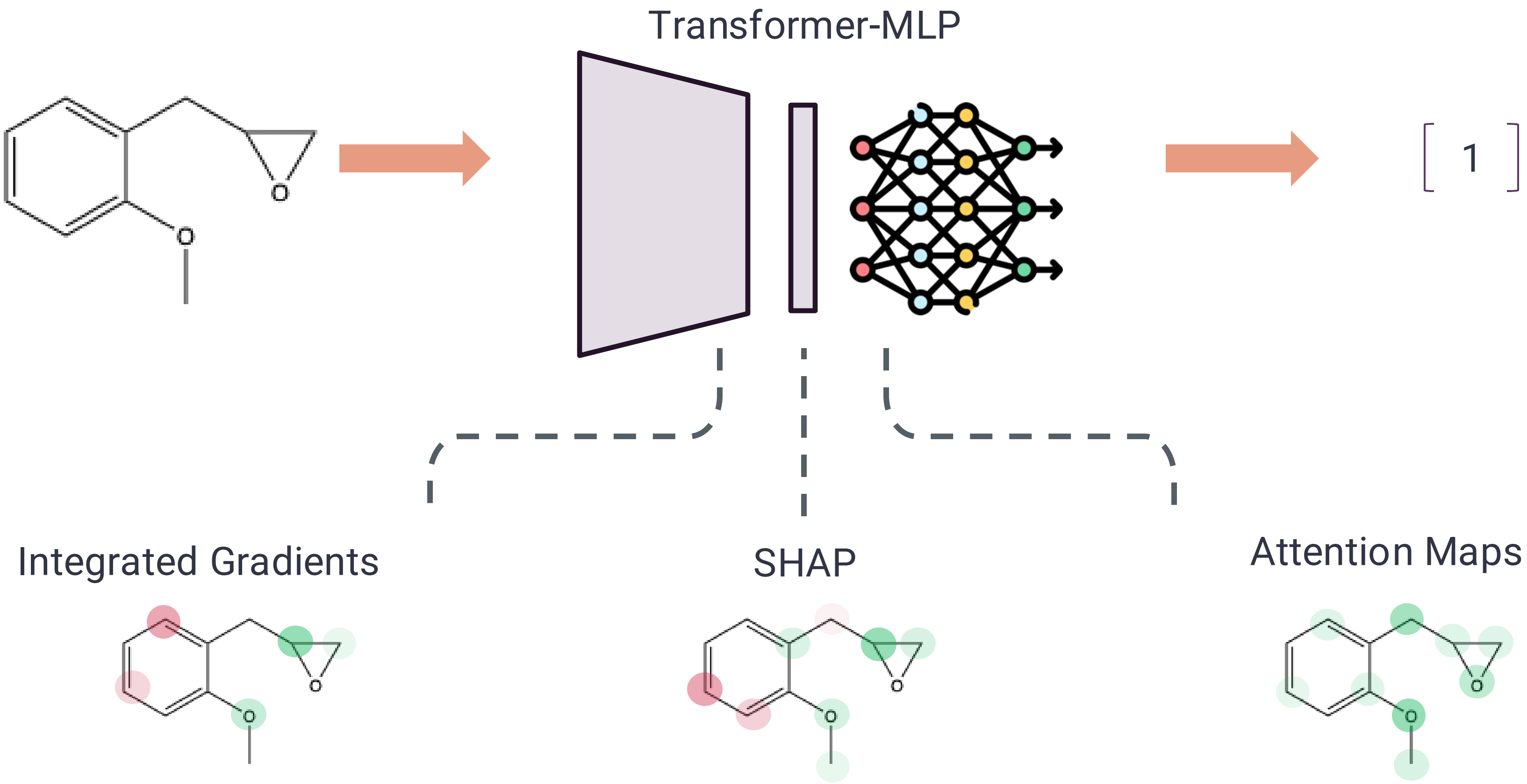
Explainable AI for functional groups

DC2: Khasanova Dina

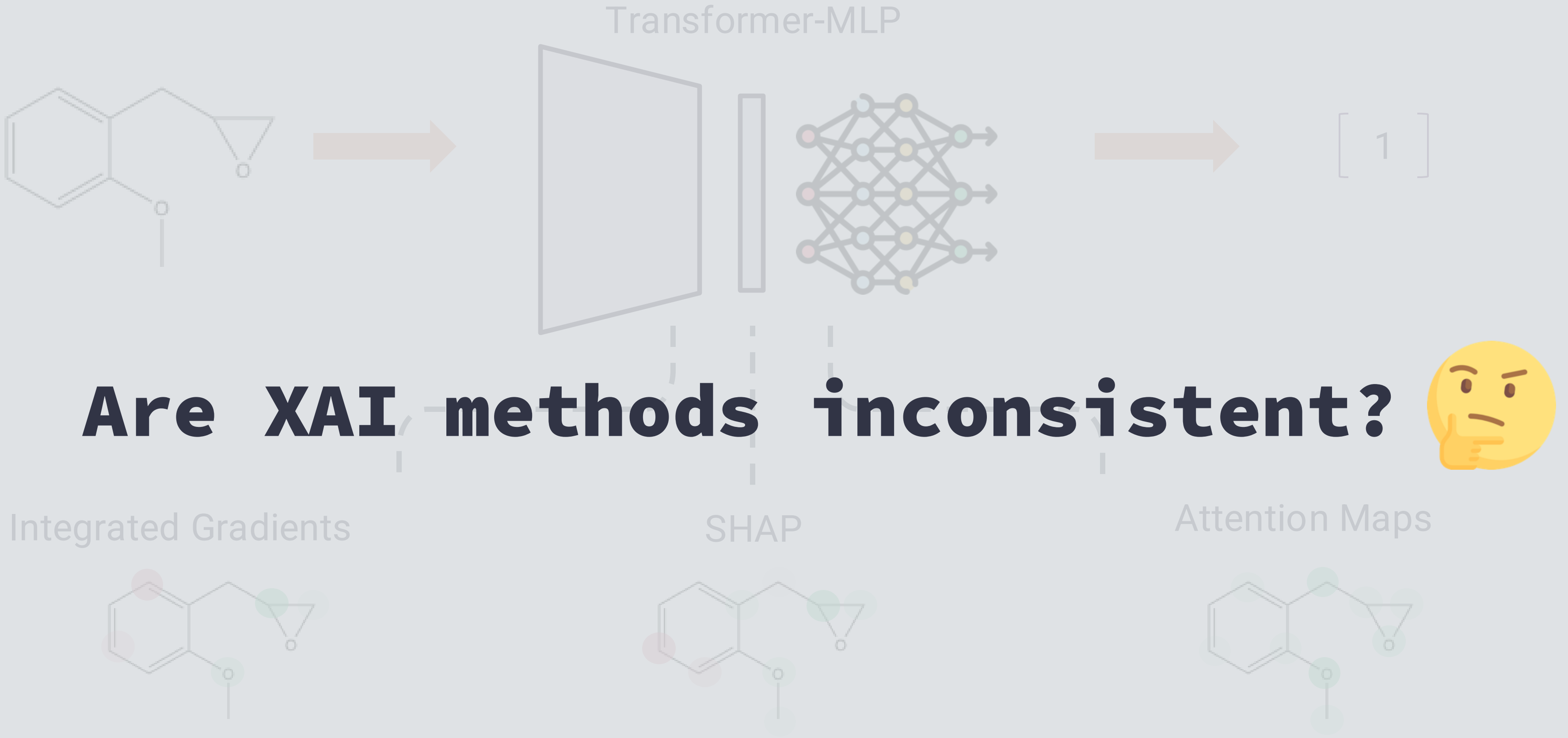
HELMHOLTZ MUNICH
Pfizer
MNAM

The reliability of XAI is important for eventually converting learned rules into structural alerts, CSRML, or SMARTS.





Hartog, P.B.R., Krüger, F., Genheden, S. et al. Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition.
J Cheminform 16, 39 (2024). https://doi.org/10.1186/s13321-024-00824-1



Are XAI methods inconsistent? 🤔

Hartog, P.B.R., Krüger, F., Genheden, S. et al. Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition.
J Cheminform 16, 39 (2024). https://doi.org/10.1186/s13321-024-00824-1

Table 7 AUROC, accuracy, F1, MCC precision and recall scores of MLP models transfer learned on Ames data

	Training	AUROC↑	Accuracy↑	F1↑	MCC↑	Precision↑	Recall↑
No training	Untrained	0.652	0.516	0.143	0.063	0.081	0.619
	Native	0.676	0.634	0.624	0.269	0.607	0.642
Variations	Random split	0.856	0.788	0.788	0.576	0.789	0.787
	Train set	0.873	0.792	0.792	0.584	0.792	0.792
	CNN	0.709	0.650	0.657	0.300	0.671	0.644
	Enumerated	0.810	0.739	0.739	0.478	0.738	0.739
Encoder only	C2C	0.734	0.666	0.666	0.332	0.665	0.666
	R2C	0.738	0.670	0.670	0.339	0.670	0.670
	E2C	0.731	0.665	0.665	0.331	0.665	0.665
	MC2C	0.754	0.682	0.682	0.364	0.683	0.682
	MR2C	0.694	0.653	0.652	0.305	0.651	0.653
	ME2C	0.804	0.719	0.719	0.438	0.719	0.719
Encoder-decoder	C2C	0.716	0.662	0.662	0.324	0.663	0.662
	R2C	0.698	0.634	0.634	0.269	0.634	0.634
	E2C	0.748	0.677	0.678	0.354	0.679	0.676
	MC2C	0.751	0.682	0.682	0.365	0.681	0.683
	MR2C	0.698	0.647	0.647	0.294	0.647	0.647
	ME2C	0.772	0.696	0.697	0.393	0.699	0.695

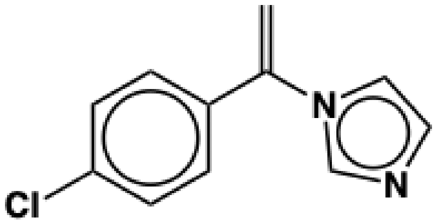
When the model's accuracy is low, each XAI method attempts to identify which features contributed to the incorrect predictions, leading to different features being highlighted and resulting in high cosine distances.

Hartog, P.B.R., Krüger, F., Genheden, S. et al. Using test-time augmentation to investigate explainable AI: inconsistencies between method, model and human intuition.
J Cheminform 16, 39 (2024). https://doi.org/10.1186/s13321-024-00824-1

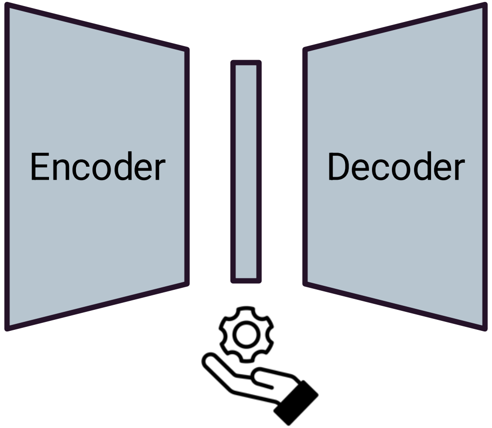
Objective of the Project:

- To build a high-accuracy model for deterministic labels like functional groups, which are well-defined and universally agreed upon, to assess how accurate or inaccurate existing explainability methods are.

non-canonical SMILES

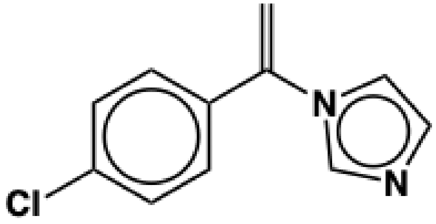


c1(C(=C)n2ccnc2)ccc(cc1)Cl
c1cn(C(c2ccc(Cl)cc2)=C)cn1
c1(C(=C)n2ccnc2)ccc(cc1)Cl
c1ncn(C(c2ccc(cc2)Cl)=C)c1
C=C(n1ccnc1)c1ccc(Cl)cc1
c1(Cl)ccc(C(n2ccnc2)=C)cc1
c1(ccc(cc1)C(n1cncc1)=C)Cl
c1(Cl)ccc(C(n2ccnc2)=C)cc1
Clc1ccc(cc1)C(=C)n1ccnc1
c1(C(n2cncc2)=C)ccc(Cl)cc1
C=C(c1ccc(Cl)cc1)n1ccnc1



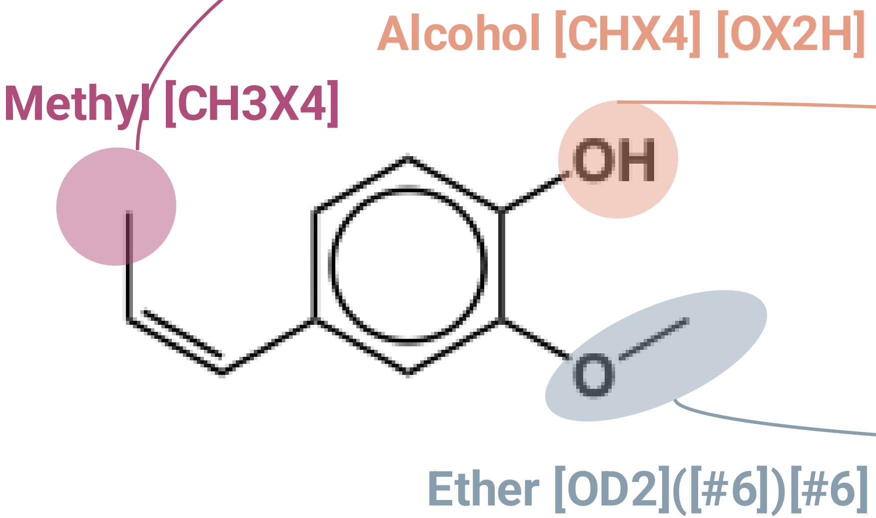
canonical SMILES

C=C(c1ccc(Cl)cc1)n1ccnc1

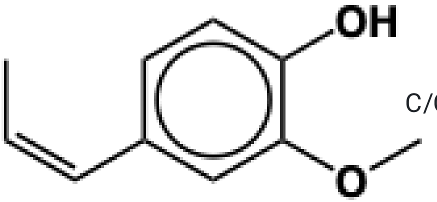


ChEMBL dataset
(≈2M molecules)

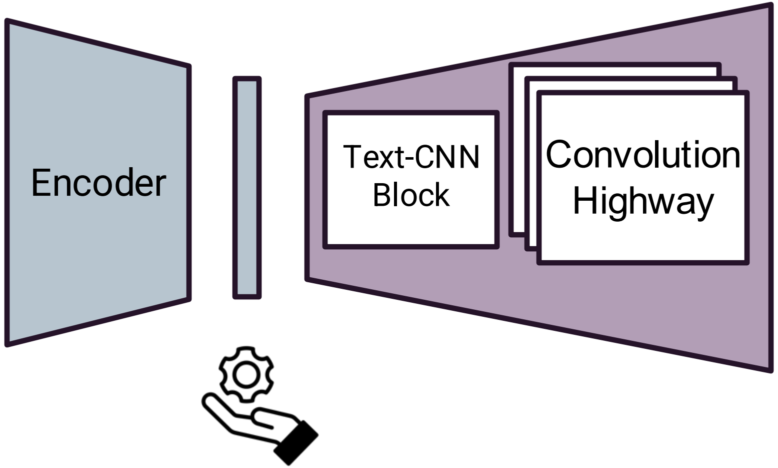
Models



methyl	alcohol	ketone	ether
1	1	0	1



C/C=C\c1ccc(O)c(OC)c1



Functional Groups

Alcohol	...	Methyl
1	...	0
...	...	...
1	...	1

TOX21 dataset
(≈9K molecules)

Models

XAI methods:

Gradient-based

- Integrated Gradients ✔
- SHAP GradientExplainer ✔

Backpropagation-Based

- SHAP DeepExplainer
- DeepLIFT ✔

Perturbation-based

- SHAP KernelExplainer
- LIME
- Occlusion ✔

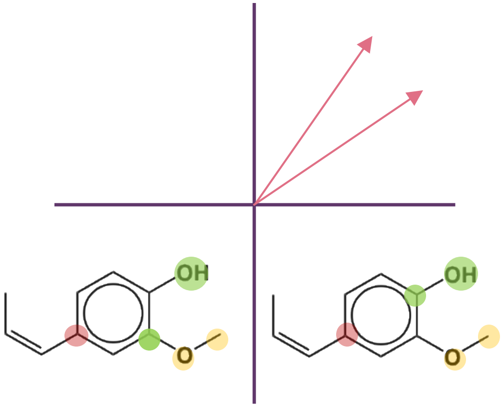
Attention-based

- Grad-CAM ✔
- Attention Maps

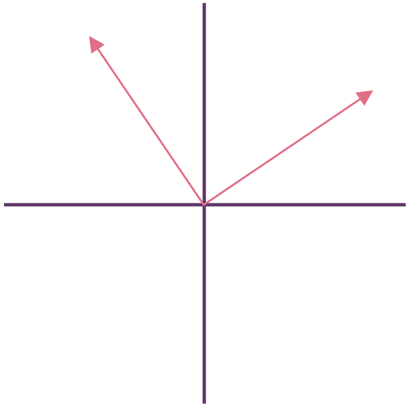
XAI

Cosine similarity

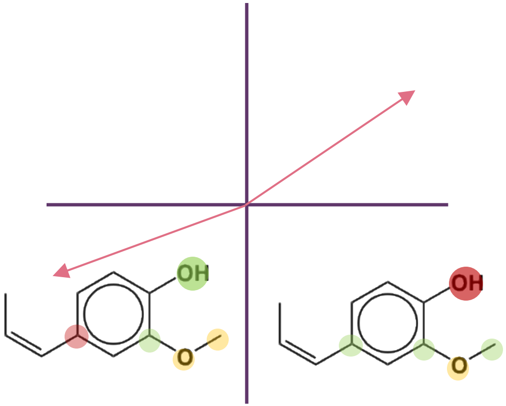
Similar



Unrelated

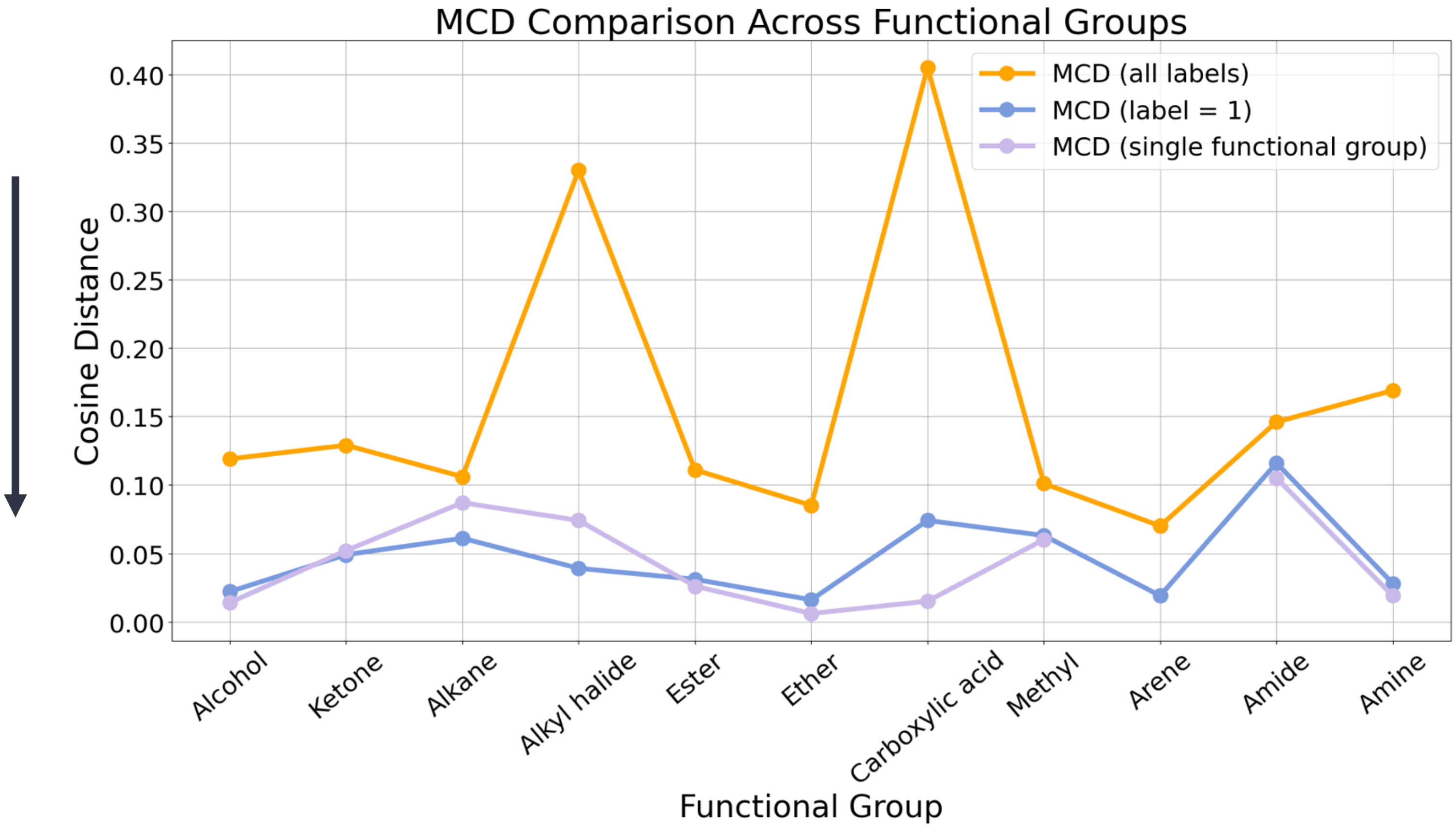


Opposite

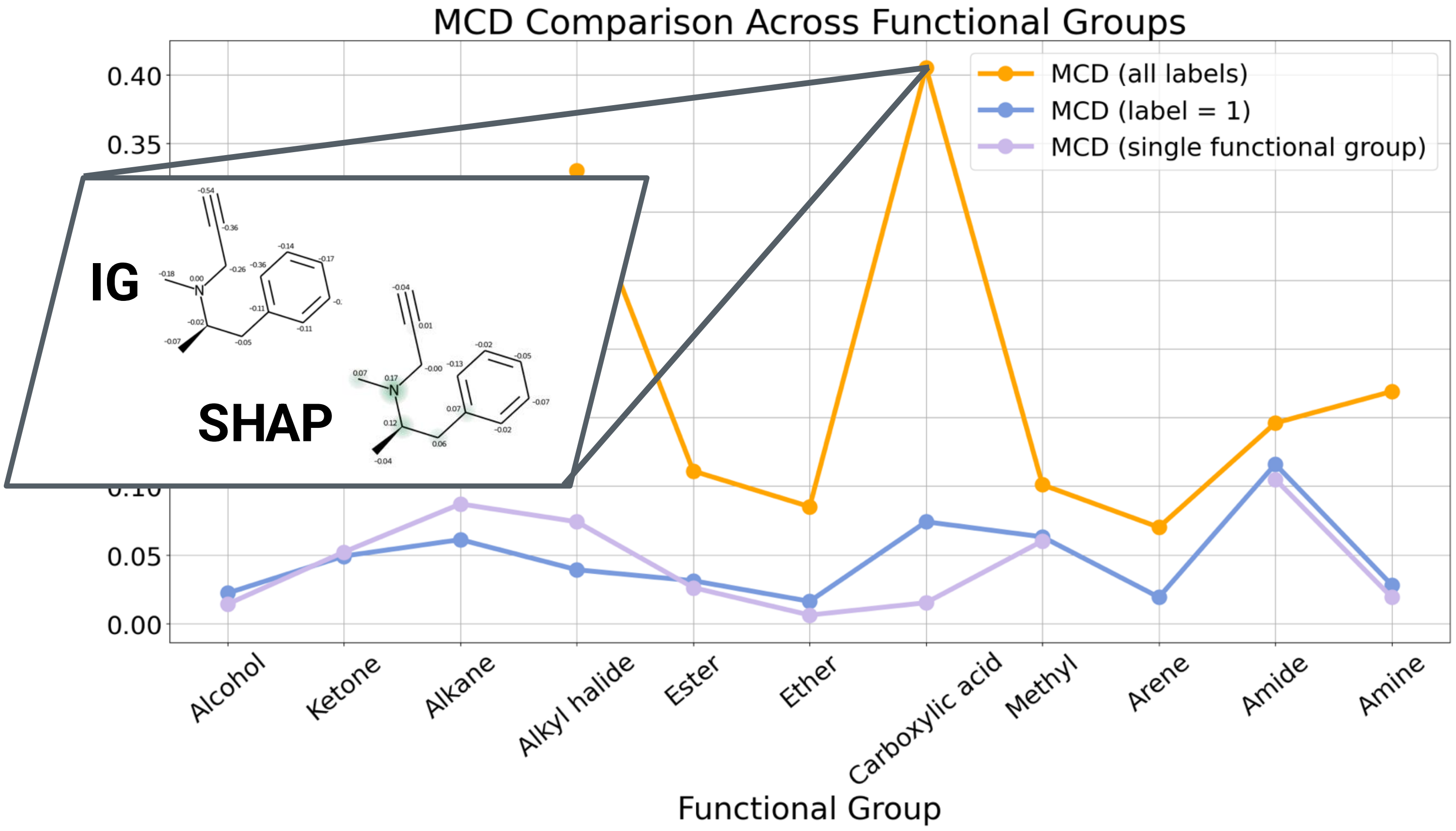


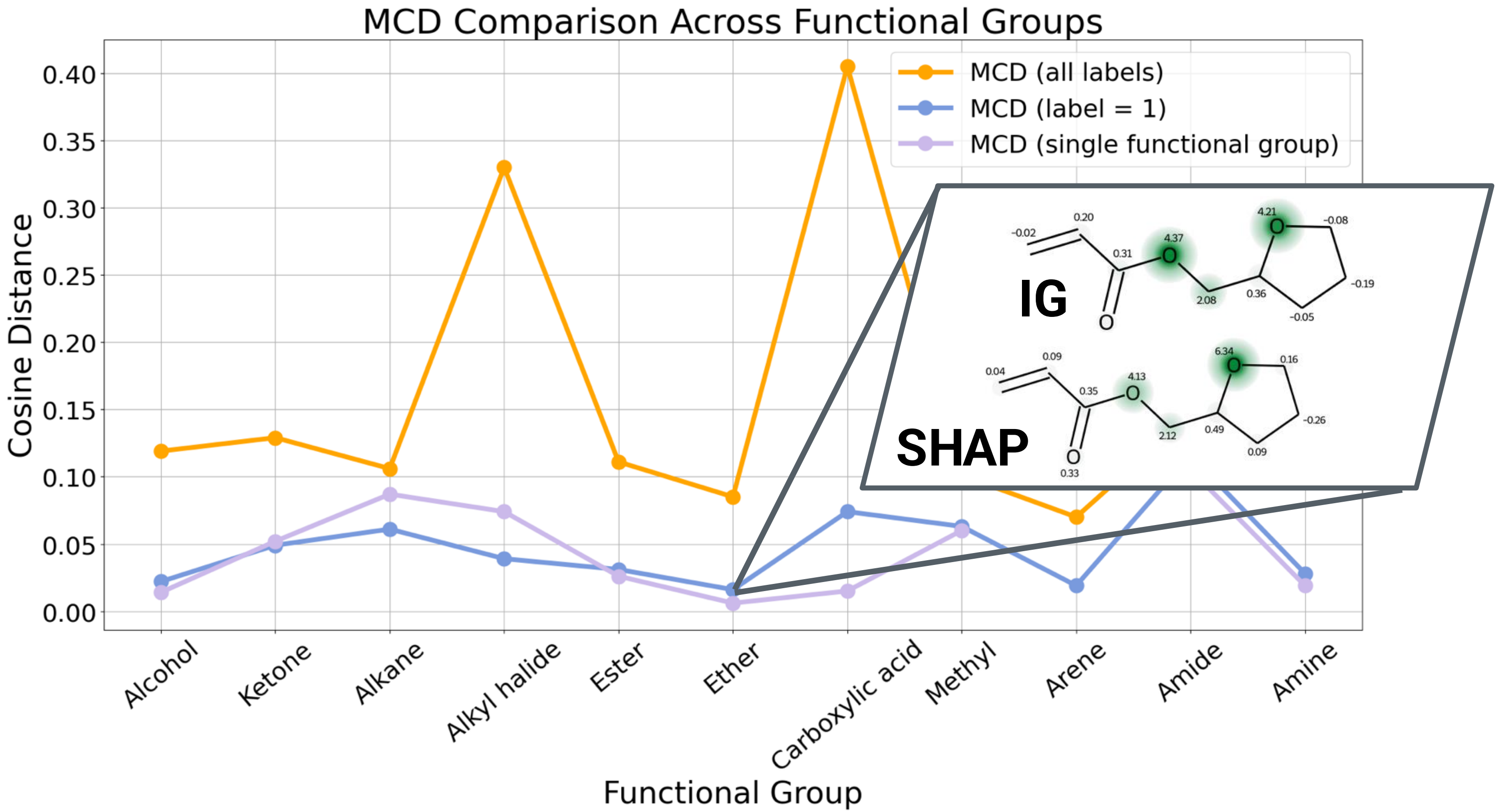
$$A*B = ||A||\ ||B||\ \cos\theta$$

Cosine Distance = 1 - Cosine Similarity

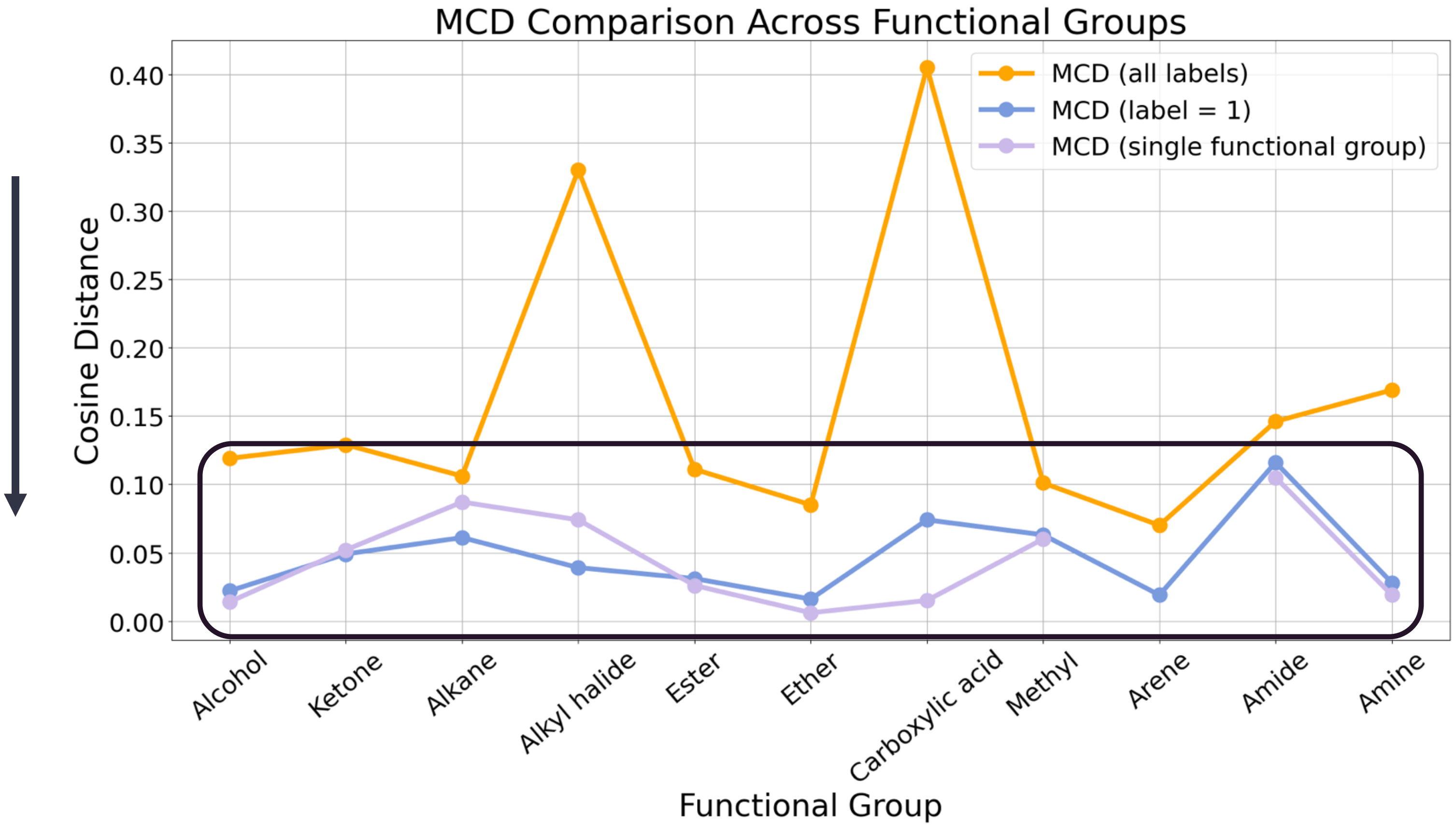


MCD Comparison Across Functional Groups
MCD (all labels)
MCD (label = 1)
MCD (single functional group)
Cosine Distance
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Ether
Carboxylic acid
Methyl
Arene
Amide
Amine
Functional Group

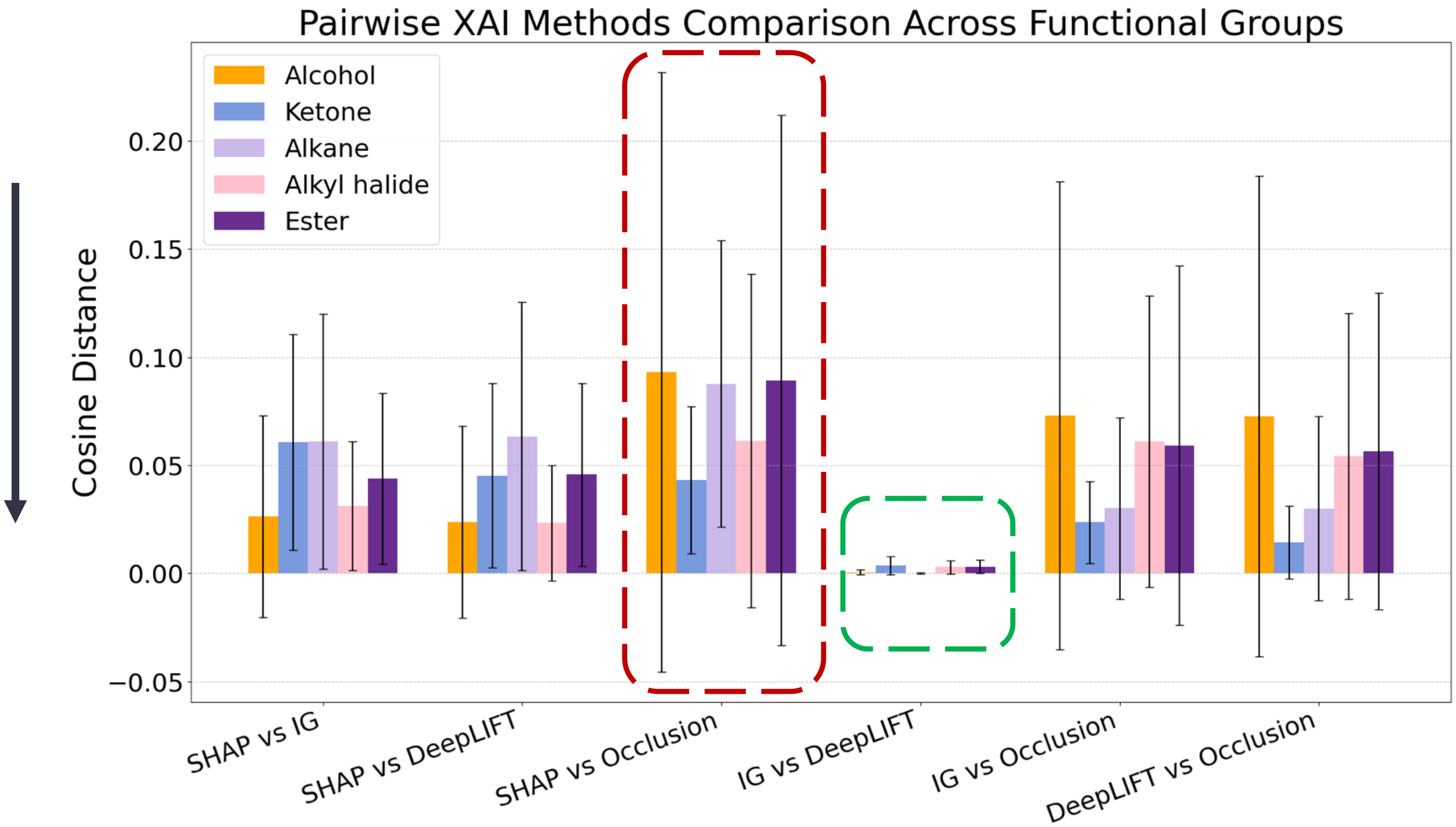

MCD Comparison Across Functional Groups
0.40
0.35
0.05
0.00
MCD (all labels)
MCD (label = 1)
MCD (single functional group)
IG
SHAP
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Ether
Carboxylic acid
Methyl
Arene
Amide
Amine
Functional Group


MCD Comparison Across Functional Groups
MCD (all labels)
MCD (label = 1)
MCD (single functional group)
Cosine Distance
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Ether
Carboxylic acid
Methyl
Arene
Amide
Amine
Functional Group
IG
SHAP

Results

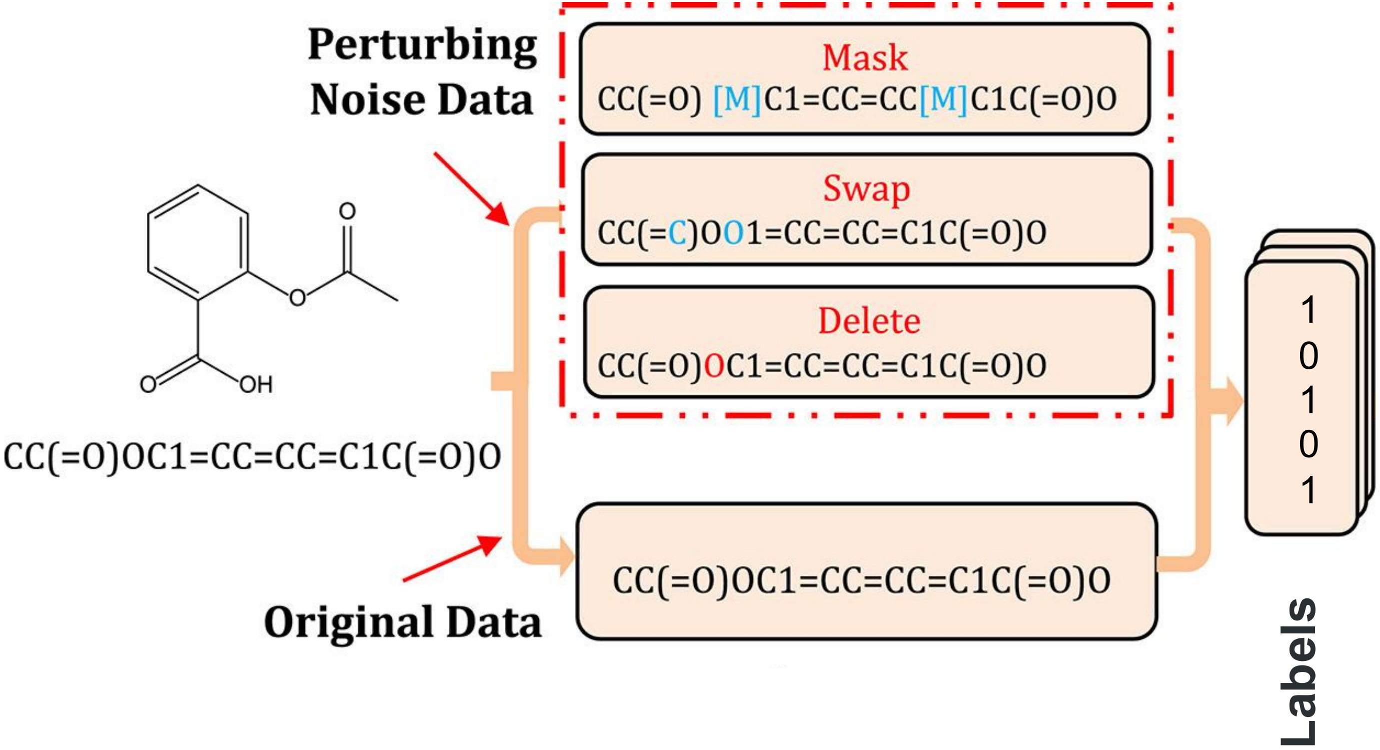


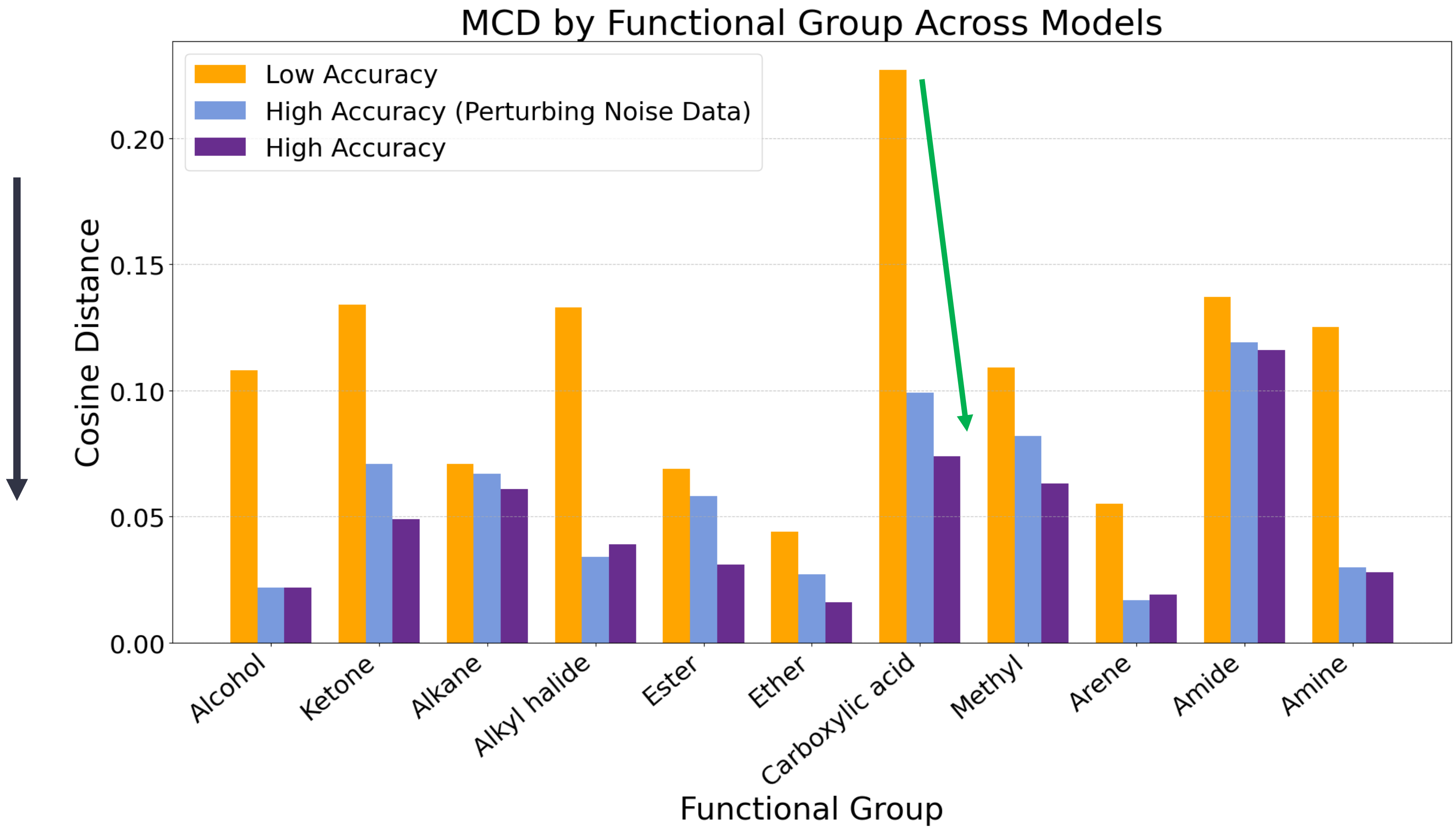
MCD Comparison Across Functional Groups
MCD (all labels)
MCD (label = 1)
MCD (single functional group)
Cosine Distance
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Ether
Carboxylic acid
Methyl
Arene
Amide
Amine
Functional Group



Pairwise XAI Methods Comparison Across Functional Groups
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Cosine Distance
0.20
0.15
0.10
0.05
0.00
−0.05
SHAP vs IG
SHAP vs DeepLIFT
SHAP vs Occlusion
IG vs DeepLIFT
IG vs Occlusion
DeepLIFT vs Occlusion

Results


Perturbing Noise Data
Mask
CC(=O) [M]C1=CC=CC[M]C1C(=O)O
Swap
CC(=C)OO1=CC=CC=C1C(=O)O
Delete
CC(=O)OC1=CC=CC=C1C(=O)O
O
O
OH
CC(=O)OC1=CC=CC=C1C(=O)O
CC(=O)OC1=CC=CC=C1C(=O)O
Original Data
1
0
1
0
1
Labels


MCD by Functional Group Across Models
Low Accuracy
High Accuracy (Perturbing Noise Data)
High Accuracy
Cosine Distance
0.00
0.05
0.10
0.15
0.20
Alcohol
Ketone
Alkane
Alkyl halide
Ester
Ether
Carboxylic acid
Methyl
Arene
Amide
Amine
Functional Group

Thank you
for your attention!