Supporting Information:

Graph Neural Network Architectures for Predicting the Electrophilicity Index: Insights from 2D and 3D Molecular Graph Representations

Sandeep Kumar*[1, a)] and Annika Bande[1, 2, b)]

[1)] *Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Hahn-Meitner-Platz 1, 14109 Berlin, Germany*

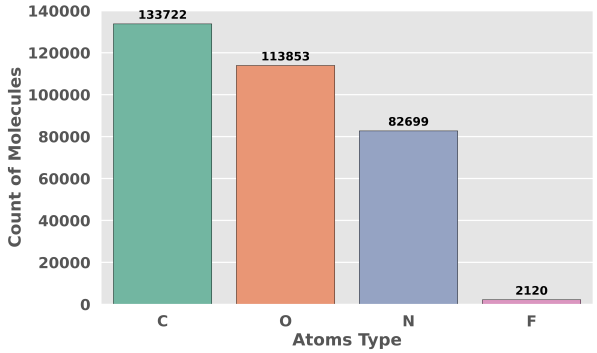
[2)] *Leibniz University Hannover, Institute of Inorganic Chemistry, Callinstr. 9, 30167 Hannover, Germany*

[a)]Electronic mail: sandeep.kumar@helmholtz-berlin.de

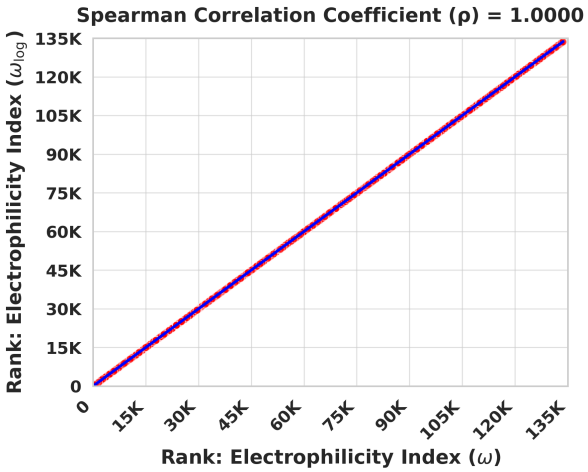
[b)]Electronic mail: annika.bande@helmholtz-berlin.de

TABLE OF CONTENTS

FIG. S1: (a) shows the frequency of atom types across molecules, (b) represents the Spearman rank correlation between the ranked values of the original ω scale and logarithmically transformed ω values.



(a)



(b)

TABLE S1: Hyperparameters of 2D GNN models for ω prediction.

Hyperparameter	**GINE**	**GIN**	**GATv2**	**GraphSAGE**	**AFP**	**GCN**
Epochs	240	300	120	240	300	180
Batch size	32	128	32	32	128	64
Learning rate	1.0×10^{-5}	1.0×10^{-3}	4.4×10^{-5}	4.4×10^{-5}	1.0×10^{-3}	1.0×10^{-3}
Weight decay	1.0×10^{-5}	1.0×10^{-6}	4.86×10^{-6}	5.0×10^{-6}	1.0×10^{-5}	1.0×10^{-6}
Number of layers	–	6	6	4	5	4
Hidden channels	256	192	256	256	256	224
Dropout	0.1	0.1	0.2	0.2	0.0	0.0

TABLE S2: Hyperparameters of 3D GNN models for ω prediction.

Hyperparameter	**GemNet**	**SchNet**	**ALIGNN**
Epochs	163	105	300
Batch size	32	32	8
Learning rate	1.20×10^{-4}	7.13×10^{-4}	1.0×10^{-3}
Weight decay	1.84×10^{-6}	1.57×10^{-6}	1.0×10^{-5}
Number of layers	4	5 (interaction blocks)	4
Hidden channels	256	64	256
Number of filters	–	256	–

TABLE S3: Performance metrics of GNN models for ω prediction (log-transformed ω) using single train/test split and 5-fold cross-validation, ordered best to worst by CV RMSE.

Metric	**GINE**	**AFP**	**GIN**	**GATv2**	**GraphSAGE**	**GCN**
			Train/Test Split			
Epochs	214	70	135	215	300	180
Train MAE	0.0470	0.0578	0.0516	0.0620	0.0635	0.0752
Test MAE	0.0524	0.0595	0.0586	0.0648	0.0662	0.0774
Train RMSE	0.0660	0.0836	0.0713	0.0883	0.0904	0.1049
Test RMSE	0.0774	0.0845	0.0832	0.0909	0.0924	0.1062
Train R^2	0.9893	0.9830	0.9876	0.9810	0.9801	0.9733
Test R^2	0.9857	0.9827	0.9832	0.9800	0.9793	0.9726
Train MSE	0.0044	0.0070	0.0051	0.0078	0.0082	0.0110
Test MSE	0.0060	0.0071	0.0069	0.0083	0.0085	0.0113
			5-Fold Cross-Validation (mean ± std)			
MAE	0.0591±0.0060	0.0620±0.0020	0.0704±0.0038	0.0753±0.0028	0.0773±0.0017	0.0856±0.0050
RMSE	0.0844±0.0062	0.0904±0.0028	0.0993±0.0043	0.1057±0.0041	0.1076±0.0035	0.1178±0.0049
R^2	0.9826±0.0027	0.9801±0.0012	0.9760±0.0020	0.9728±0.0021	0.9718±0.0018	0.9662±0.0029
MSE	0.0072±0.0011	0.0082±0.0005	0.0099±0.0009	0.0112±0.0009	0.0116±0.0008	0.0139±0.0012

TABLE S4: Performance metrics of ALIGNN, GemNet, and SchNet for ω prediction using single train/test split and 5-fold cross-validation (log-transformed ω), ordered best to worst by CV RMSE.

Metric	**ALIGNN**	**GemNet**	**SchNet**
Train/Test Split (log-transformed ω)			
Train MAE	0.0222	0.0265	0.0329
Test MAE	0.0286	0.0360	0.0374
Train RMSE	0.0292	0.0352	0.0433
Test RMSE	0.0412	0.0531	0.0532
Train R^2	0.9979	0.9970	0.9954
Test R^2	0.9963	0.9932	0.9931
5-Fold Cross-Validation (mean ± std, log-transformed ω))			
MAE	0.0313±0.0051	0.0403±0.0016	0.0493±0.0070
RMSE	0.0457±0.0035	0.0595±0.0019	0.0670±0.0073
R^2	0.9949±0.0007	0.9914±0.0005	0.9890±0.0024

FIG. S2: Predicted vs. actual and residuals vs. actual of ω (log-transformed ω) for GINE, AFP, and GIN models (top to bottom).

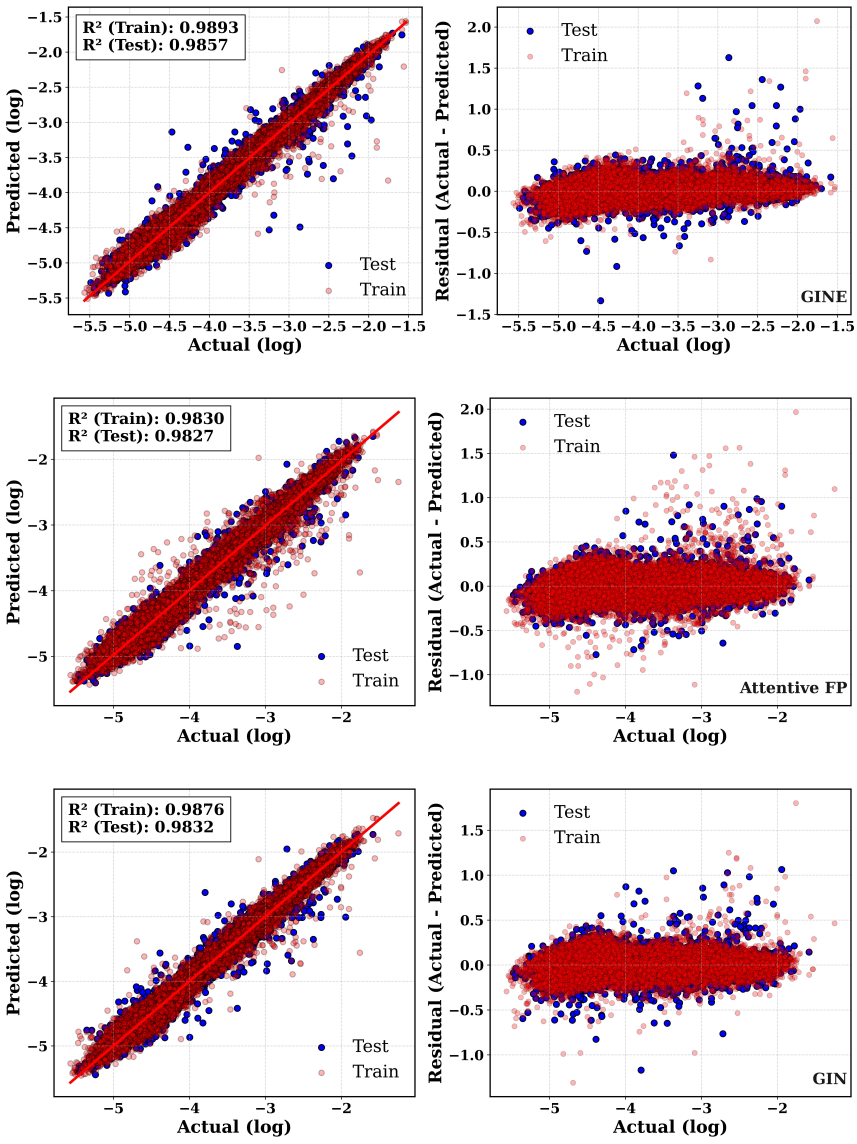


FIG. S3: Predicted vs. actual and residuals vs. actual of ω (log-transformed ω) for GATv2, GraphSAGE, and GCN models (top to bottom).

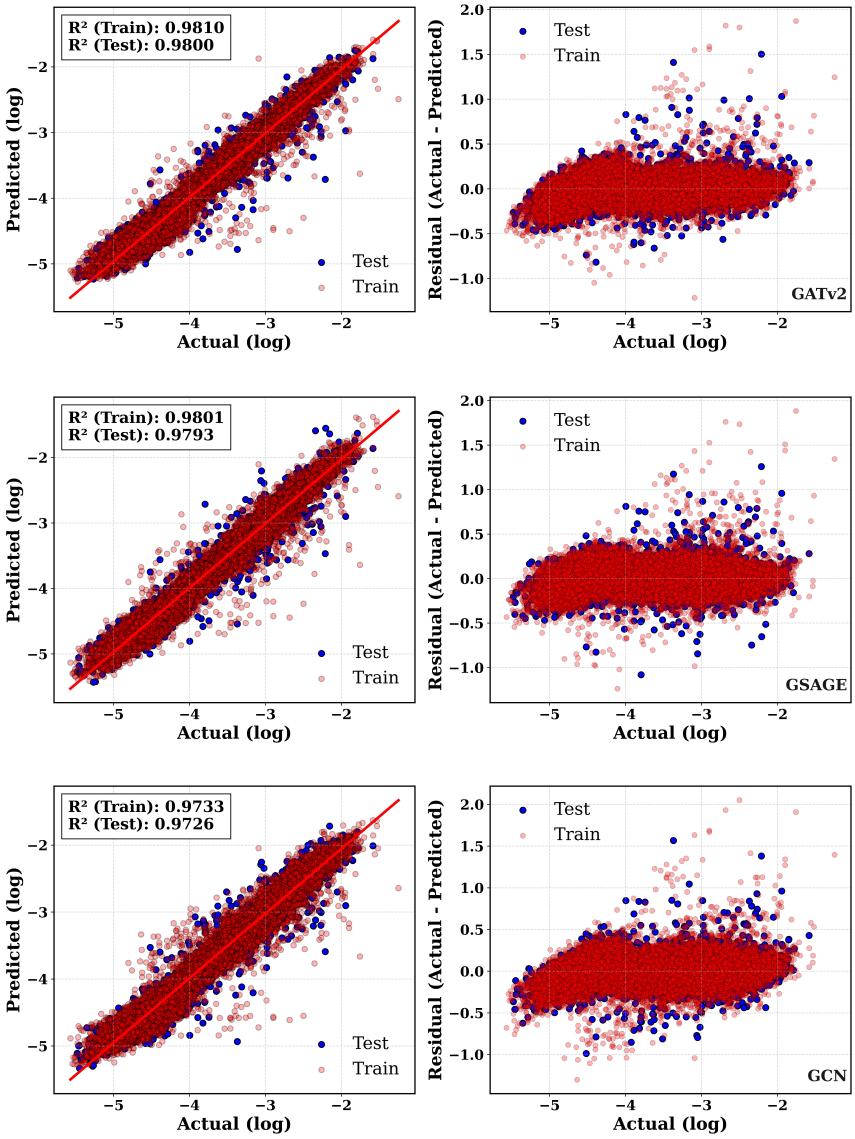
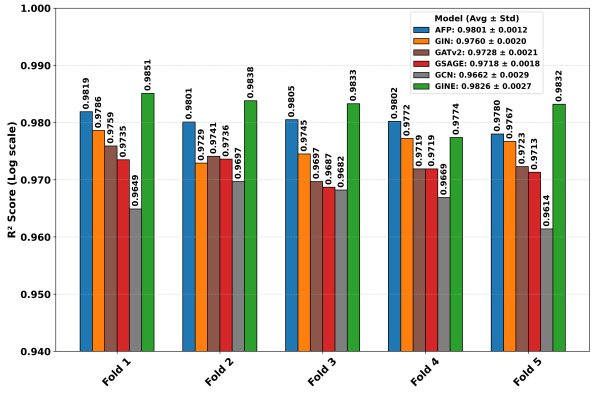
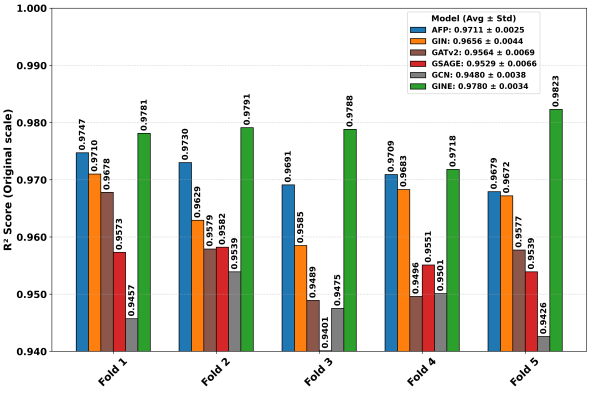


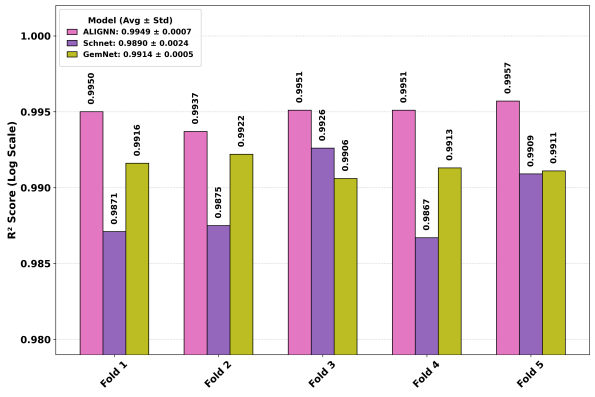
FIG. S4: 5-fold cross-validation R^2 scores for 2D and 3D molecular graph models predicting ω. (a) 2D models on log-transformed ω. (b) 2D models on the original ω scale. (c) 3D models on log-transformed ω. (d) 3D models on the original ω scale.



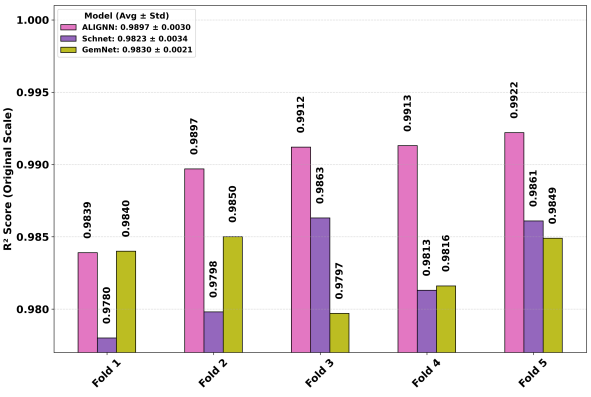
(a)



(b)

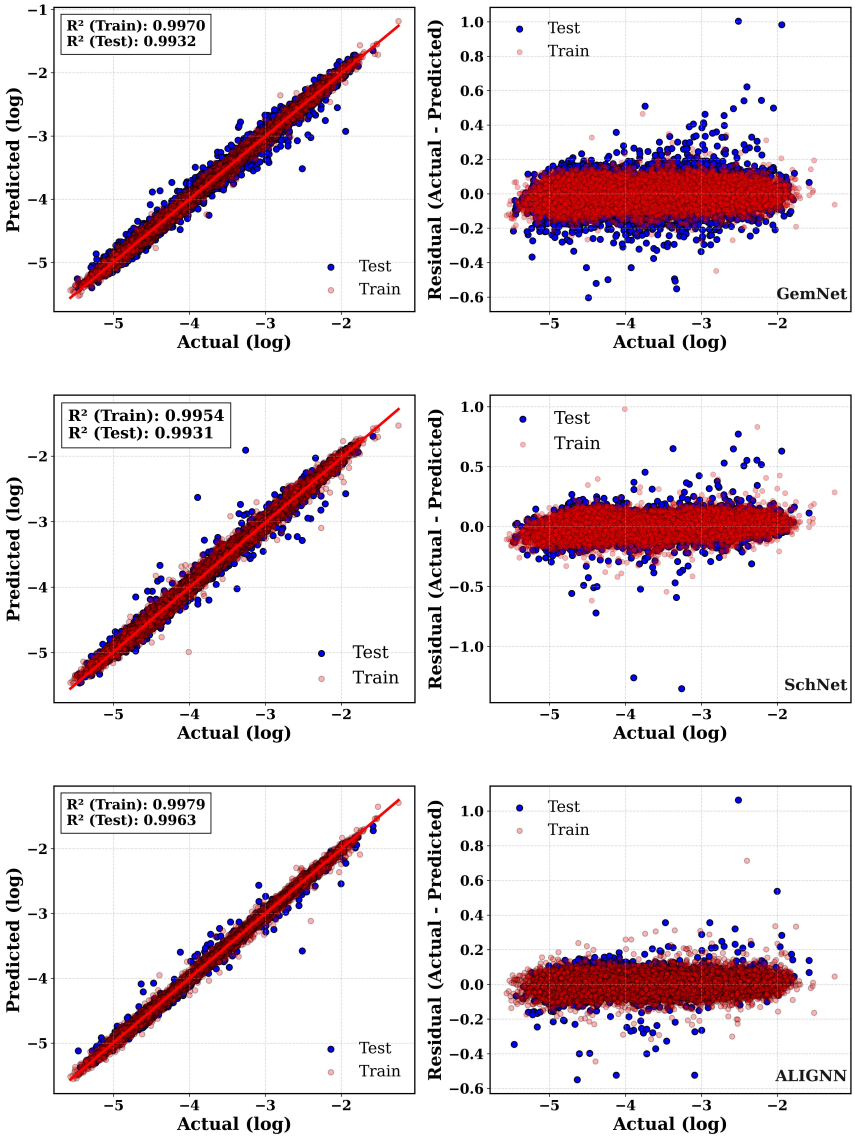


(c)



(d)

FIG. S5: Predicted vs. actual and residuals vs. actual of ω (log-transformed ω) for GemNet, SchNet, and ALIGNN models (top to bottom).



SECTION S1: EVALUATION METRICS FOR ELECTROPHILICITY INDEX PREDICTION (ω)

To assess the performance of regression models in predicting the ω, we employ widely used statistical metrics: MAE, MSE, RMSE, and the R^2. These metrics quantify the discrepancy between predicted and reference values, providing complementary perspectives on model accuracy. The MAE measures the average magnitude of prediction errors, irrespective of their direction, and

is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\omega_i - \hat{\omega}_i\right|, \tag{1}$$

where ω_i and $\hat{\omega}_i$ denote the true and predicted ω, respectively, and n is the total number of molecules. The MSE emphasizes larger deviations by squaring the errors:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\omega_i - \hat{\omega}_i\right)^2, \tag{2}$$

while the RMSE provides a measure in the same units as ω:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\omega_i - \hat{\omega}_i\right)^2}. \tag{3}$$

The Coefficient of determination, R^2, evaluates the proportion of variance in the reference data explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\omega_i - \hat{\omega}_i)^2}{\sum_{i=1}^{n}(\omega_i - \bar{\omega})^2}, \tag{4}$$

where $\bar{\omega}$ is the mean of the true ω. Values of R^2 closer to 1 indicate better predictive performance. During model training, these metrics were computed on both the training and validation sets after each epoch, with summary statistics recorded at regular intervals. Final evaluation was performed on the independent test set to ensure robust and unbiased assessment of model predictive capability. Collectively, these metrics provide a comprehensive evaluation of regression accuracy, reliability, and generalization for ω prediction.

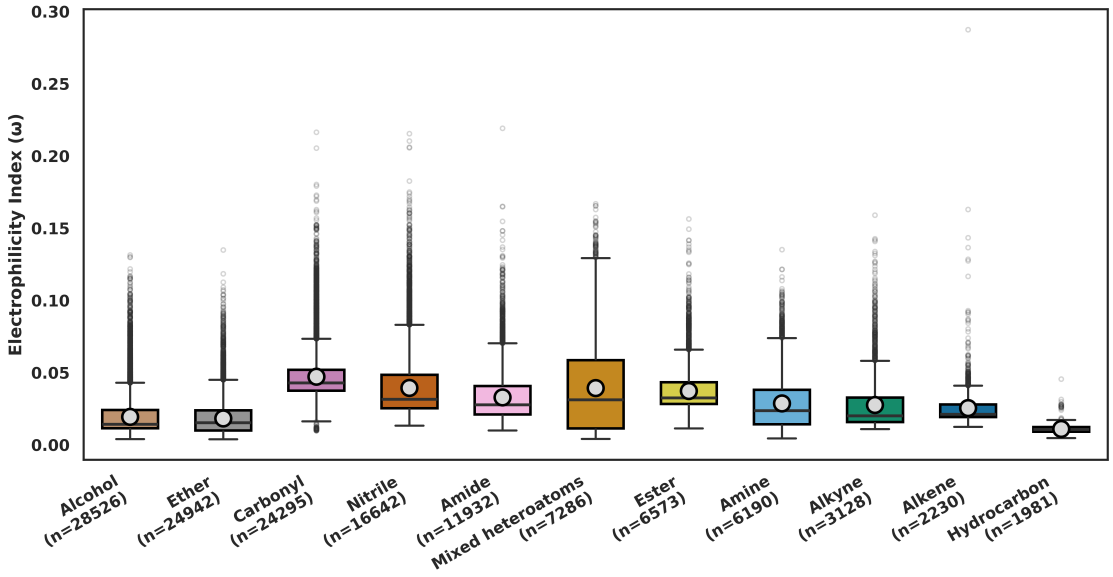


FIG. S6: Distribution of the electrophilicity index (ω) across representative functional groups. Boxes denote the interquartile range (IQR), the central line indicates the median, whiskers show the data range, and outliers are plotted as individual points. White markers represent the mean value for each group.