Supplemental Information for OBELiX: A Curated Dataset of Crystal Structures and Experimentally Measured Ionic Conductivities for Lithium Solid-State Electrolytes

I. PARITY PLOTS AND DATA FROM FIGURE 4

Table S1 contains the information from Figure 4 in a tabulated form. Figure S1 presents parity plots for benchmarking experiments discussed in Section 4.

II. BASELINE MODELS

We trained and tested 7 ML models which are briefly described below:

1. RF, as an ensemble of decision trees, is robust to noisy data and provides feature importance insights, making it a strong baseline for structured datasets.
2. MLP, a neural network-based approach, captures complex nonlinear relationships, offering a comparison to deep learning-based methods.
3. PaiNN [1] enforces E(3)-equivariance, enabling accurate modeling of atomic interactions and force predictions.
4. SchNet [2] learns continuous filter representations, making it effective for capturing atomic environments.
5. M3GNet [3]integrates message passing with three-body interactions, improving property predictions for crystalline materials.
6. SO3Net [4] leverages spherical harmonics to enhance equivariant representations for molecular and solid-state systems.
7. CGCNN [5] models crystal structures directly as graphs, making it a strong baseline for learning structure-property relationships.

Table S2 shows the best hyperparameter sets for each model presented in Table S1

III. EFFECTS OF ADDED RANDOM NOISE

Table S3 presents the cross-validation and test MAEs for the models trained on randomized atomic positions. The models were trained on the randomized data and tested on the original CIFs. There is no significant difference in performance between models trained on the original data and models trained on data with added random noise on atomic positions.

IV. COMPUTATIONAL RESOURCES USED FOR BENCHMARKING

Table S4 presents the resources used to find optimal hyperparameters and train each model.

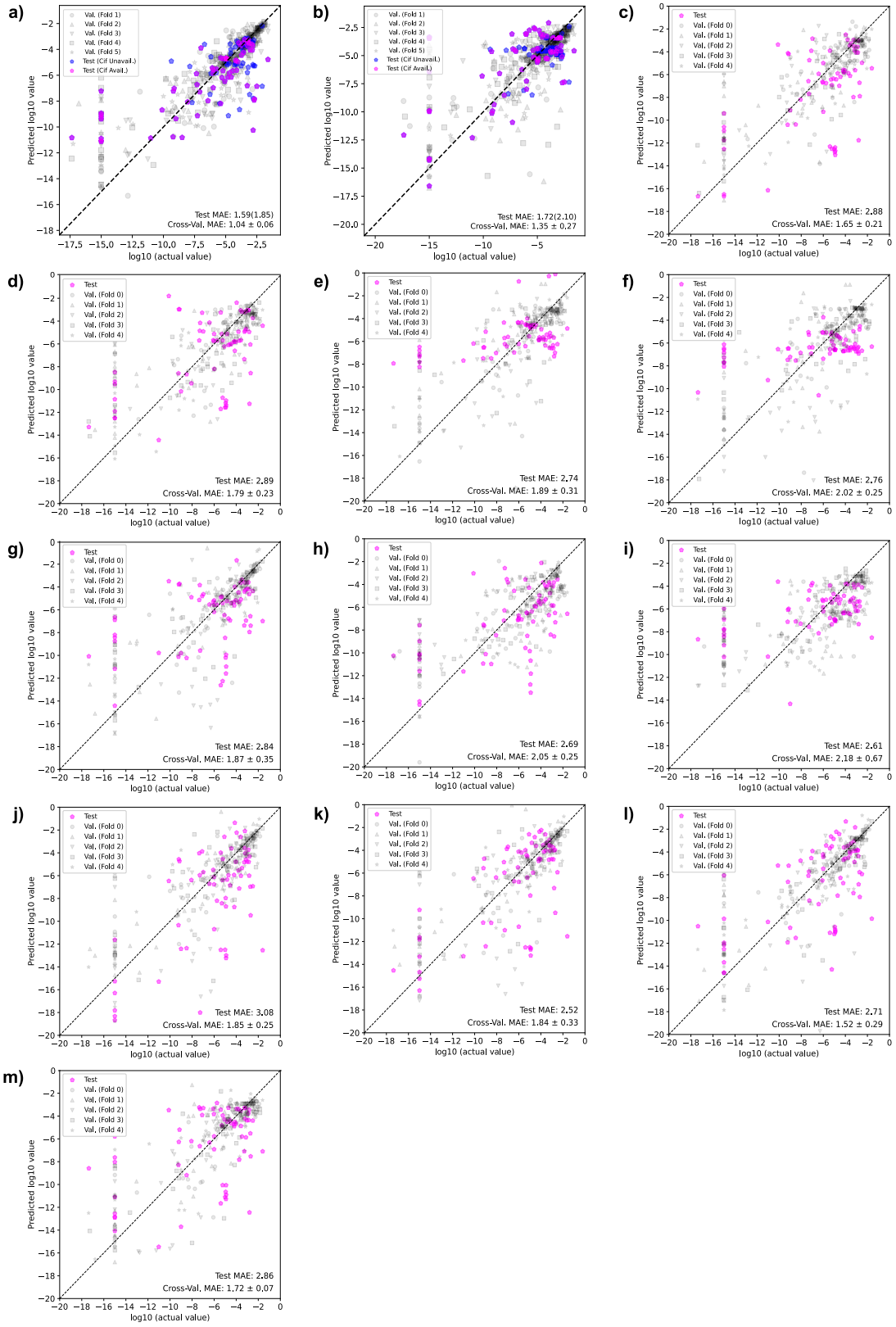

FIG. S1. Parity plots for benchmarked models. a) Random Forest b) Multilayer perceptron c) PaiNN d) SchNet e) M3GNet f) SO3Net g) CGCNN h) PaiNN with pretraining i) SchNet with pretraining j) M3GNet with pretraining k) CGCNN with pretraining l) CGCNN with disorder (partial occupancy) m) SO3Net with disorder (partial occupency)

TABLE S1. Benchmarking of various ML models with and without pretraining. For the median prediction, the random forest (RF) and the multilayer perceptron (MLP), results are presented for the full dataset and numbers in parenthesis are results for the subset of the test set that has CIF files. All other results apply only to entries with CIF files. "p-" indicates a model that was pretrained and "dis-" indicates a model that was modified to take partial occupancy (disorder) into account.

Model	**Cross-val. MAE** Avg. ± SD	**Test MAE**
Experiment		0.41
Median pred.		2.16 (2.81)
RF	1.04 ± 0.06	1.59 (1.85)
MLP	1.35 ± 0.27	1.72 (2.10)
PaiNN	1.65 ± 0.21	2.88
SchNet	1.79 ± 0.23	2.89
M3GNet	1.89 ± 0.31	2.74
SO3Net	2.02 ± 0.25	2.76
CGCNN	1.87 ± 0.35	2.84
p-PaiNN	2.05 ± 0.25	2.69
p-SchNet	2.18 ± 0.67	2.61
p-M3GNet	1.85 ± 0.25	3.08
p-CGCNN	1.84 ± 0.33	2.52
dis-CGCNN	1.52 ± 0.29	2.71
dis-SO3Net	1.72 ± 0.07	2.86

TABLE S2. The selected hyperparameters for the baseline models.

Model	Hyperparameter	Value
RF	max_depth	36
	max_features	sqrt
	min_samples_leaf	1
	n_estimators	50
MLP	activation	relu
	batch_size	16
	early_stopping	True
	hidden_layer_sizes	[64, 64, 64, 64]
	learning_rate	adaptive
	learning_rate_init	0.01
	max_iter	1000
	n_iter_no_change	100
	solver	adam
PaiNN	cutoff	5
	n_interactions	2
	n_atom_basis	80
	batch_size	32
	max_epochs	100
	weight_decay	0.0001
Schnet	cutoff	5
	n_interactions	3
	n_atom_basis	80
	batch_size	32
	max_epochs	100
	weight_decay	0.01
M3GNet	cutoff	5.0
	threebody_cutoff	5.0
	is_intensive	True
	readout_type	"set2set"
	nblocks	3
	dim_node_embedding	128
	dim_edge_embedding	128
	units	64
	batch_size	35
	max_epochs	50
	lr	0.001
	weight_decay	0.01
SO3Net	cutoff	5.0
	is_intensive	True
	nmax	2
	lmax	1
	target_property	"graph"
	readout_type	"set2set"
	nblocks	3
	dim_node_embedding	64
	nlayers_readout	3
	units	32
	batch_size	35
	max_epochs	80
	lr	0.001
	weight_decay	0
CGCNN	n_conv	3
	n_h	1
	atom_fea_len	64
	h_fea_len	64
	batch_size	35
	epochs	50
	lr	0.001
	weight_decay	0

TABLE S3. Performance of the 5 geometric models on the public dataset with added random noise to the atomic positions. Test results are on the original test set.

Model	Cross-validation MAE	Test MAE
PaiNN	2.03 ± 0.27	2.95
SchNet	1.99 ± 0.22	2.78
M3GNet	1.83 ± 0.29	2.91
SO3Net	1.98 ± 0.23	2.79
CGCNN	1.94 ± 0.42	2.95

TABLE S4. Resource usage for benchmarking

Model	**Hardware**	**Hyperparameter tuning**	**Final training**
RF	AMD EPYC 7502 (1 core)	7min	1s
MLP	AMD EPYC 7502 (1 core)	50min	14s
PaiNN	NVidia A100 GPU	2h40min	2min
SchNet	NVidia A100 GPU	1h55min	2min
M3GNet	NVidia A100 GPU	3h35min	2min
SO3Net	NVidia A100 GPU	2h25min	2min
CGCNN	NVidia A100 GPU	1h40min	1min

[1] K. Schütt, O. T. Unke and M. Gastegger, Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, 2021, pp. 9377–9388.
[2] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko and K. Müller, *Neural Information Processing Systems*, 2017.
[3] C. Chen and S. P. Ong, *Nature Computational Science*, 2022, **2**, 718–728.
[4] K. T. Schütt, S. S. Hessmann, N. W. Gebauer, J. Lederer and M. Gastegger, *The Journal of Chemical Physics*, 2023, **158**, 144801.
[5] T. Xie and J. C. Grossman, *Physical review letters*, 2018, **120**, 145301.