

Cite this: *Digital Discovery*, 2026, 5, 910

# OBELiX: a curated dataset of crystal structures and experimentally measured ionic conductivities for lithium solid-state electrolytes

Félix Therrien,<sup>†\*a</sup> Jamal Abou Haibeh,<sup>†ab</sup> Divya Sharma,<sup>†ac</sup> Rhiannon Hendley,<sup>de</sup> Leah Wairimu Mungai,<sup>af</sup> Sun Sun,<sup>e</sup> Alain Tchagang,<sup>e</sup> Jiang Su,<sup>e</sup> Samuel Huberman,<sup>†b</sup> Yoshua Bengio,<sup>†ac</sup> Hongyu Guo,<sup>†\*e</sup> Alex Hernández-García<sup>†\*ac</sup> and Homin Shin<sup>†\*e</sup>

Solid-state electrolyte batteries are expected to replace liquid electrolyte lithium-ion batteries in the near future thanks to their higher theoretical energy density and improved safety. However, their adoption is currently hindered by imperfect electrode–electrolyte interfaces and a lower effective ionic conductivity, a quantity that governs charge and discharge rates. Identifying highly ion-conductive materials using conventional theoretical calculations and experimental validation is both time-consuming and resource-intensive. While machine learning holds the promise to expedite this process, relevant ionic conductivity and structural data is scarce. Here, we present OBELiX, a database of ~600 synthesized solid electrolyte materials and their experimentally measured room temperature ionic conductivities gathered from literature and curated by domain experts. Each material is described by their measured composition, space group and lattice parameters. A full-crystal description in the form of a crystallographic information file (CIF) is provided for ~320 structures for which atomic positions were available. We discuss various statistics and features of the dataset and provide training and testing splits carefully designed to avoid data leakage. Finally, we benchmark seven existing ML models on the task of predicting ionic conductivity and discuss their performance. The goal of this work is to facilitate the use of machine learning for solid-state electrolyte materials discovery.

Received 1st October 2025  
Accepted 23rd December 2025

DOI: 10.1039/d5dd00441a

rsc.li/digitaldiscovery

## 1 Introduction

Lithium-ion batteries (LIBs) used in most consumer electronics and electric vehicles have seen immense progress in terms of energy density, power density, safety and durability. However, their performance is reaching a plateau. Solid-state batteries are regarded as the next generation of batteries that may allow significant improvement over these characteristics.<sup>1,2</sup> The key difference between these two technologies is their electrolyte, the medium which allows the transport of ions during charge and discharge. A solid-state electrolyte (SSE)—as opposed to a liquid electrolyte in LIBs—permits new design choices that

ultimately lead to better battery properties,<sup>3</sup> let alone the fact that they are not flammable unlike their liquid counterparts.

Ionic conductivity ( $\sigma$ ), expressed in siemens per centimeter ( $S\text{ cm}^{-1}$ ), measures how easily ions can move through a medium or material. Ideal SSEs, also called “superionic” or “fast-ionic” conductors, are electrolytes that exhibit ionic conductivity comparable to those observed in liquid electrolytes and molten solids ( $>1\text{ mS cm}^{-1}$ ). Only a handful of *room temperature* ideal SSEs are known thus far within a small number of classes of materials: LISICON (*e.g.*,  $\text{Li}_{14}\text{ZnGe}_4\text{O}_{16}$ ), NASICON (*e.g.*,  $\text{Li}_{1.3}\text{Al}_{0.3}\text{Ti}_{1.7}(\text{PO}_4)_3$ ), garnet (*e.g.*,  $\text{Li}_7\text{Li}_3\text{Zr}_2\text{O}_{12}$ ), perovskite (*e.g.*,  $\text{Li}_{0.5}\text{La}_{0.5}\text{TiO}_3$ ), and argyrodite (*e.g.*,  $\text{Li}_6\text{PS}_5\text{Cl}$ ).<sup>2</sup>

Until now, the discovery of novel SSEs has largely relied on an incremental, experimental approach which consists, for example, of substituting atoms and elements in known compounds. This has allowed the discovery of some highly ion-conductive materials, but greatly limits the search space given that the experimental synthesis and characterization of a new, stable, inorganic solid-state electrolyte is a difficult and costly process that can take months to years.<sup>4</sup>

Computational discovery, on the other hand, requires time-consuming atomistic simulations, such as *ab initio* molecular dynamics (AIMD) which is based on density functional theory

<sup>a</sup>Mila, Montréal, Canada. E-mail: felix.therrien@mila.quebec; alex.hernandez-garcia@mila.quebec

<sup>b</sup>Department of Chemical Engineering, McGill University, Montréal, Canada

<sup>c</sup>Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada

<sup>d</sup>Department of Chemistry and Biomolecular Science, University of Ottawa, Ottawa, Canada

<sup>e</sup>National Research Council Canada, Ottawa, Canada. E-mail: hongyu.guo@nrc-cnrc.gc.ca; homin.shin@nrc-cnrc.gc.ca

<sup>f</sup>Technical University of Kenya, Nairobi, Kenya

<sup>†</sup> These authors contributed equally to this work.



(DFT), to accurately capture the complex relationship between ionic conductivity and the material's structure and composition.<sup>13–15</sup> These calculations can take from several hours to a few days for a single ionic conductivity and their parameters are often materials specific. Therefore, they are not well suited for large-scale explorations of hypothetical materials.

Machine learning (ML) has the potential to greatly accelerate the discovery of novel SSEs. Naturally, it can be used to predict ionic conductivity directly using, for example, graph neural networks (GNNs), which have been used extensively and successfully in materials science.<sup>16,17</sup> Machine-learned force fields or interatomic potentials (MLFF or MLIP) can also be used to obtain ionic conductivity through molecular dynamics in the “classical” way while using significantly less resources.<sup>18</sup> Finally, generative frameworks can accelerate dynamics simulations<sup>19</sup> and, provided that good ionic conductivity models are developed, there exists a wide range of frameworks that could generate new materials conditioned on that property.<sup>20–23</sup> However, the main obstacle to the development and validation of these models—and to some extent theoretical models—is the scarcity of relevant experimental ionic conductivity and structural datasets. Indeed, as detailed in the next section, the few datasets that exist contain partial material descriptions and ionic conductivity measurements at various or unspecified temperatures. To the best of our knowledge there does not exist another open access dataset of experimental room temperature ionic conductivities with corresponding full crystal descriptions.

In this work, we assembled OBELiX (Open solid Battery Electrolytes with Li: an eXperimental dataset), a curated database of 599 synthesized solid electrolyte materials and their experimentally measured room temperature ionic conductivity along with descriptors of their space group, lattice parameters, and chemical composition.‡ The database is analyzed in terms of the distribution of ionic conductivity, space groups, elements, and repeated compositions. We also propose a training and testing split that avoids data leakage between similar entries while balancing distributions of properties across splits. We use this split to benchmark the performance of 7 machine learning models at directly predicting room temperature ionic conductivity ( $\sigma_{RT}$ ).

We believe that this dataset and benchmark can significantly spur the use of ML for the discovery of novel solid-state battery materials. It may be small but it is important to realize that the database represents a large fraction of all materials whose ionic conductivity has been characterized experimentally. Importantly, this database has been carefully curated by domain experts and formatted by machine learning scientists to facilitate its use by this community. Finally, we believe that this benchmark can encourage novel machine learning research tailored to low-data regimes.

## 2 Related work

Crystal structure databases such as the Materials Project<sup>24</sup> or the Inorganic Crystal Structure Database (ICSD)<sup>25,26</sup> contain large amounts of potential candidates for solid-state electrolytes. For example, Sendek *et al.*<sup>5</sup> screened more than 12 000 Li-containing crystals for Li-ion SSEs using multiple criteria, thereby identifying 317 candidates, among which 21 crystals that showed promise as SSEs were selected from an ML-guided model. The ionic conductivity of these 21 structures was estimated theoretically. Jalem *et al.*<sup>6</sup> annotated 318 compounds by calculating ion migration energy barriers ( $E_b$ ), a less accurate but computationally lighter property that relates to ionic conductivity. Bayesian optimization was employed to screen candidate compounds with low  $E_b$ . He *et al.*<sup>7</sup> compiled a database of over 90 000 crystal structures, including more than 7000 structures with preliminary ion-transport data obtained through geometric analysis, and 12 000 activation energy values ( $E_b$ ) calculated using the bond valence site energy method. Additionally, they manually extracted 75 CIF files from literature data. They employed empirical and geometrical methods to estimate the minimum energy paths of these structures and obtain  $E_b$ , but they did not predict  $\sigma$ .

On the experimental side, the Liverpool Ionics (LiIon) Dataset<sup>8</sup> reports 820 entries containing chemical composition, structural family, and ionic conductivity at different temperatures (from 5 to 873 °C) measured by alternating current impedance spectroscopy, among which 465 entries were at room temperature. Laskowski *et al.*<sup>9</sup> gathered a dataset of 1346 entries with compositions, space group, and corresponding  $\sigma_{RT}$ , with a subset of 344 compounds whose structures are manually matched with an ICSD ID. The full dataset, including references, is only available as a pdf file. While we were preparing OBELiX the same group of authors published a new dataset partially based on Laskowski *et al.*<sup>9</sup> that contains a total of 571 compounds with experimentally measured ionic conductivities at room temperature.<sup>10</sup> Since OBELiX is also based on Laskowski *et al.*,<sup>9</sup> it has a significant overlap with McHaffie *et al.*<sup>10</sup> that will be discussed in Section 3.2.

Shon and Min<sup>11</sup> used text mining to extract more than 4000 ionic conductivity measurements from 1457 papers. Each ionic conductivity measurement is associated with a composition and about 350 are also associated with a “structure type”. Measurement temperature is not specified and compositions are not always fully described. A recent study by Yang *et al.*<sup>12</sup> introduced the Dynamic Database of Solid-State Electrolyte (DDSE) to facilitate the exploration of structure–performance relationships and accelerate the discovery of high-performance solid-state electrolytes (SSEs). The database contains performance data for 2448 materials (at time of writing), including ionic conductivity obtained from experimental reports, across a broad temperature range (132.40–1261.60 K). Ionic conductivity data is only available upon request to the authors.

These recent reports greatly increased the amount of readily available experimental ionic conductivity data. However, they contain limited structural information: the databases by Shon

‡ OBELiX is available here: <https://github.com/NRC-Mila/OBELiX>.



and Min<sup>11</sup> and Yang *et al.*<sup>12</sup> contain only a qualitative structure description for some materials, the LiIon dataset only includes the structural family and the dataset by Laskowski *et al.*<sup>9</sup> is limited to space group information. Although the full crystallographic information of the 344 compounds of the Laskowski dataset for which the ICSD ID is provided could be retrieved, the proprietary ICSD is not available to most researchers in the ML community. Table 1 summarizes the differences in terms of available features across the databases discussed above.

The lack of precise structural information labeled with ionic conductivity makes it difficult (1) to compare experimental values with theoretical predictions which require full crystal descriptions and (2) to train machine learning models to accurately predict ionic conductivity.

## 3 Data

### 3.1 Background

While the combination of the composition, lattice parameters and space group is often sufficient to qualify materials, they do not fully describe the crystal structure because in general they do not specify the positions of each atom. Some, but not all, experimental papers perform an additional analysis (Rietveld refinement) of the X-ray powder diffraction pattern to estimate atomic positions. Only in these cases is it possible to obtain a full description of the crystal including atomic positions which is necessary to build a crystal information file (CIF). This is why it is not possible to obtain CIF files for all entries in our dataset. The full crystal description including atomic positions is the information required to perform, for example, molecular dynamics simulations or density functional theory calculations.

In contrast to theory-based data found in the Materials Project, for example, experimental compositions often feature fractional numbers (real numbers rather than integers) resulting from partially vacant sites or disorder associated with partial cation substitution. Consider, for example, composition  $K_{0.1}Li_{0.9}SbO_3$ . At a specific location in the crystal (a site) there is a 90% probability of finding a lithium (Li) atom and a 10% probability of finding a potassium (K) atom. Site occupancy

does not need to add up to one since sites are often partially empty.

Such partial occupancy is ubiquitously observed in Li-ion SSEs<sup>27</sup> and it plays a crucial role in creating diffusion pathways. For example, the  $\sigma_{RT}$  of tetragonal  $Li_7La_3Zr_2O_{12}$  with a space group of  $I4_1/acd$  (no. 142) is two orders of magnitude smaller than that of the same garnet framework of cubic  $Li_7La_3Zr_2O_{12}$  with  $Ia\bar{3}d$  (no. 230) (see Fig. 1a). In this case, the disordering and partial occupation of Li (at the 96h site) promotes the Li-ion conduction. In the halide structure of  $Li_3InCl_6$  (Fig. 1b), the substitution of one  $Li^+$  with the  $In^{3+}$  cation introduces two intrinsic vacancies, to which is attributed the high  $\sigma_{RT}$  of that material. In sum, in order to screen SSEs with high  $\sigma_{RT}$ , it is highly desirable to include partial occupancy as a key feature of the materials.

### 3.2 Data collection

We built our dataset starting from the Liverpool Ionics Dataset and the Laskowski dataset by selecting materials for which the experimental room temperature ionic conductivity, space group and lattice parameters could be obtained. We manually retrieved missing information (*e.g.* lattice parameters or  $\sigma_{RT}$ )

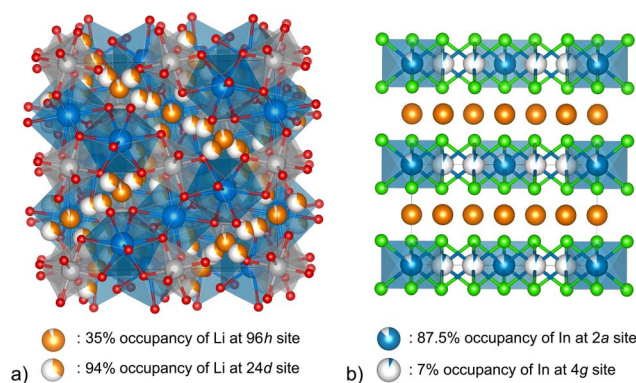


Fig. 1 Examples of solid state electrolyte materials with partial occupancies. (a)  $Li_7La_3Zr_2O_{12}$  with space group  $Ia\bar{3}d$  (b)  $Li_3InCl_6$  with space group  $C2/m$ .

**Table 1** Comparison of our dataset (OBELiX) with existing ones based on key features and labels. For features, the numbers represent the number of entries with that feature that are labeled with at least one experimental or computational ion transport property (not necessarily ionic conductivity). Numbers in parentheses represent proprietary or private data

Dataset	Labels		Features			
	$\sigma_{RT}^{exp} < \sigma^{exp}$		Comp.	Spg	Lattice	CIFs
Sendek <i>et al.</i> <sup>5</sup>	0	0	317	317	317	317
Jalem <i>et al.</i> <sup>6</sup>	0	0	318	318	318	318
He <i>et al.</i> <sup>7</sup> (SPSE)	0	0	75 (12k)	75 (12k)	75 (12k)	75 (12k)
Hargreaves <i>et al.</i> <sup>8</sup> (LiIon)	465	820	820	0	0	0
Laskowski <i>et al.</i> <sup>9</sup>	1346	1346	1346	0 (344)	0 (344)	0 (344)
McHaffie <i>et al.</i> <sup>10</sup>	571	571	571	0 (571)	0 (571)	0 (571)
Shon and Min <sup>11</sup>	n.a.	4032	4032	0	0	0
Yang <i>et al.</i> <sup>12</sup> (DDSE)	(1939)	(2448)	2448	0	0	0
OBELiX	599	599	599	599	599	321





### 3.3 Data splits

Experimental papers in this field often measure ionic conductivity for several variations of the same materials while changing the composition slightly. This can lead to multiple entries that are very similar and often have similar ionic conductivities. There are also several entries in our dataset that have the same composition, which may also lead to similar ionic conductivities. To avoid data leakage when testing machine learning models on OBELiX and to fairly compare new models in the future, we provide a split of the data where entries from the same paper or that have the same composition must be in the same set (training or testing).

To obtain this split, we used a Monte Carlo method that moved groups of entries from one set to the other to minimize (1) the difference between the distribution of log ionic conductivity between the two sets and (2) the difference between their respective subsets containing CIF files. The algorithm also ensured that the final test set represented between 20% and 30% of the data. This algorithm is available on our public repository.

The obtained distribution of log ionic conductivity in each set and subset is presented in Fig. 2a along with the proportion of each crystal family and space group. The test set represents 20.2% of the full dataset and 20.9% of the subset that has CIF files.

The distributions in log space of ionic conductivity for the two sets are very similar. Note that the entries plotted at  $10^{-15}$  were reported as having a conductivity of “less than  $10^{-10}$ ,” without a quantitative value. The proportion of crystal families and space groups is also fairly similar between the two sets, except for space group 167, which is much more prevalent in the training set. This is due to the fact that a large group of entries (106) with space group 167 were either from the same paper or had the same composition. This meant that the entire group could not be split between the two sets without leaking either a paper or a composition.

The dataset contains 55 space groups, 4 of which are only in the test set. Fig. 2c shows the prevalence of the 55 different elements that are present in the dataset. All entries contain lithium (by design) and most of them contain oxygen. Phosphorus, lanthanum, sulfur and titanium follow as the most prevalent elements. Silver is the only element that is not found in the training set (it is only in the test set).

About 75% (245/321) of the entries with atomic information have some level of partial occupancy (disorder). The proportion of partially occupied structures in each split was not controlled for explicitly, but it is similar in the test (53/67) and train (192/254) splits. For the rest of the entries, when atomic positions and occupations are unknown, it is not always possible to tell if a structure is disordered.

## 4 Benchmarks

In this section, we benchmarked how well existing models perform on the new dataset. This evaluation is essential for determining whether these models can be effectively applied or

if there is a need to develop new models better suited for the task.

We note that experimental data intrinsically embeds errors and uncertainty associated not only with various sources of measurement techniques but also with data extraction from figures and inconsistent labeling (*e.g.*, bulk, grain boundary, or total ionic conductivity are often indistinguishably reported). Before assessing the performance of predictive models it makes sense to quantify the uncertainty (“performance”) of experimental data acquisition. Thankfully, our dataset contains 48 sets of compositions and space groups that have multiple entries, spanning a total of 122 entries. These entries and their corresponding ionic conductivities are plotted in Fig. 3. The color represents the maximum difference in lattice parameters between any two entries of a same set. The maximum difference is of only 1.2% for all sets, which gives us confidence that grouped materials are in fact the same. This means that these materials were synthesized and their ionic conductivity measured two or more times, most likely by different researchers. This represents a unique opportunity to quantify experimental uncertainty and reproducibility. The inset of Fig. 3 shows the distribution of log ionic conductivities with respect to the mean of each set of repeated materials. The root mean squared deviation from the set averages of the  $\log(\sigma_{RT})$  is of 0.63 and the mean absolute deviation from the set medians is of 0.41. The latter can be compared to the model’s mean absolute error when predicting log ionic conductivity and represents its lower bound. Therefore, any model that would be reported as having lower MAE than that value would most likely be over-trained.

### 4.1 Baselines

To evaluate the performance of ML models on OBELiX, we tested five widely adopted graph neural networks developed specifically for materials science applications, PaiNN,<sup>29</sup> SchNet,<sup>30</sup> M3GNet,<sup>31</sup> SO3Net,<sup>32</sup> and CGCNN<sup>33</sup> on the subset of the dataset that contains CIF files. These graph-based models, where each node represents an atom, effectively capture atomic interactions while preserving molecular invariance, enabling accurate material property predictions when trained on large datasets.<sup>34</sup> On the full dataset, where atomic positions are not always available, we also tested two standard machine learning models, a random forest (RF) and a multilayer perceptron (MLP).

The RF and the MLP use the composition, space group and lattice parameters as inputs where the composition is a vector containing the occurrence ( $\in \mathbb{R}$ ) of each element of the periodic table. The 3D geometric models use the crystal structure as their input and build different representations from that structure. The crystal structure contains the composition and space group information implicitly, but the models are not given that information explicitly. The way atomic and structural information is processed and aggregated into a single learned representation for each material is a defining aspect of each of the models. Therefore, for most experiments, we did not alter the models’ representations beyond what could be modified with



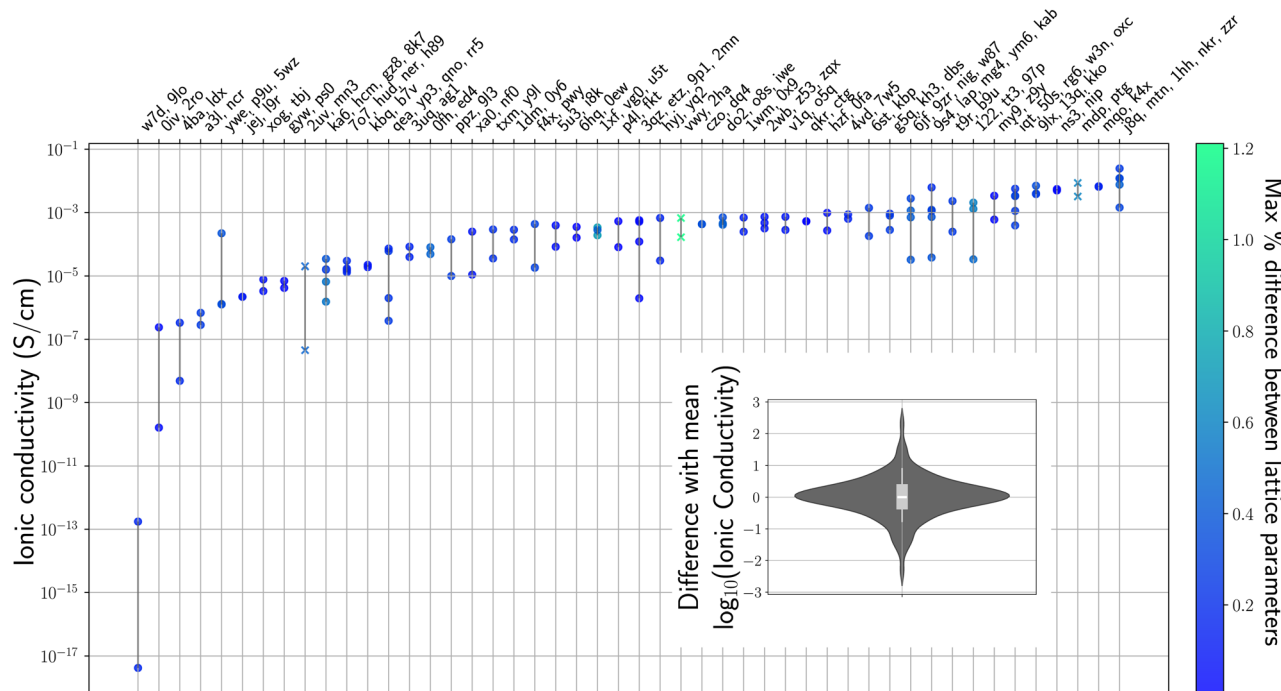


Fig. 3 Ionic conductivity of entries in the dataset that have the same composition and space group. The color shows the largest relative difference between lattice parameters within a set of entries with same space group and composition. The inset shows the distribution of differences with the mean ionic conductivity of the sets in log scale. It is scaled proportionally to the rest of the plot.

hyperparameters. However, since none of the models could take into account partial occupancy of the sites and a large portion of entries contain such disorder, we created disordered versions of CGCNN and SO3Net with a slightly modified atomic embedding. For all other experiments, occupations were rounded to the nearest integer before being fed to the models.

## 4.2 Setup

To optimize the training process and assess the stability of the models, we implemented a 5-fold cross-validation strategy. For hyperparameter optimization, we employed a grid search strategy across a predefined space of 100 randomly sampled hyperparameter sets for each model. This number was selected to strike a balance between comprehensive exploration of the hyperparameter space and computational feasibility. In the case of RF and MLP where training is extremely fast; all hyperparameter sets were tested. The hyperparameter space was carefully designed for each model based on its unique architecture and requirements (see Table S2 in the SI for a complete list). For example, PaiNN's search space included parameters such as the cutoff distance, number of interactions, and batch size.

We computed the mean absolute error (MAE) between the predicted and the measured ionic conductivities to evaluate the performance of each configuration. Specifically, the average validation MAE across all folds in the cross-validation process was used to assess each setup's effectiveness. The hyperparameter set that achieved the lowest average validation MAE was selected as the best-performing configuration. After

choosing the best hyperparameters, each model was retrained on the entire training set and evaluated on the test set. A detailed table of the selected hyperparameters for each model is included in the SI (Table S2).

Pretraining can enhance model performance by initializing weights with knowledge from larger datasets and related tasks, which is then fine-tuned on a smaller, task-specific dataset. We pretrained PaiNN and SchNet on the Materials Project with a band gap prediction task. In this case we fixed the trained representation (PaiNN or SchNet) and trained the output model (an MLP followed by a pooling layer) on OBELiX. For M3GNet and CGCNN we use pretrained models that were available on their public repositories. The M3GNet model was trained on formation energy per atom whereas CGCNN was trained on Fermi energy both from the Materials Project. As recommended in their respective documentation, we fine-tuned the models by training all model parameters starting from the trained models.

## 4.3 Discussion

Fig. 4 and Table S1 present our benchmarking results and Fig. S1 presents the corresponding parity plots. The MLP and the RF were trained on the full training set, but tested on both the full test set (in orange) and the subset of the test set that has CIF files (in red). The goal is to be able to compare their performance directly with geometric models, given that the variance of the CIF subset is larger.

The two simple models, RF and MLP, outperform all 3D geometric models both in the cross-validation and the test performance even when comparing with the subset of the test



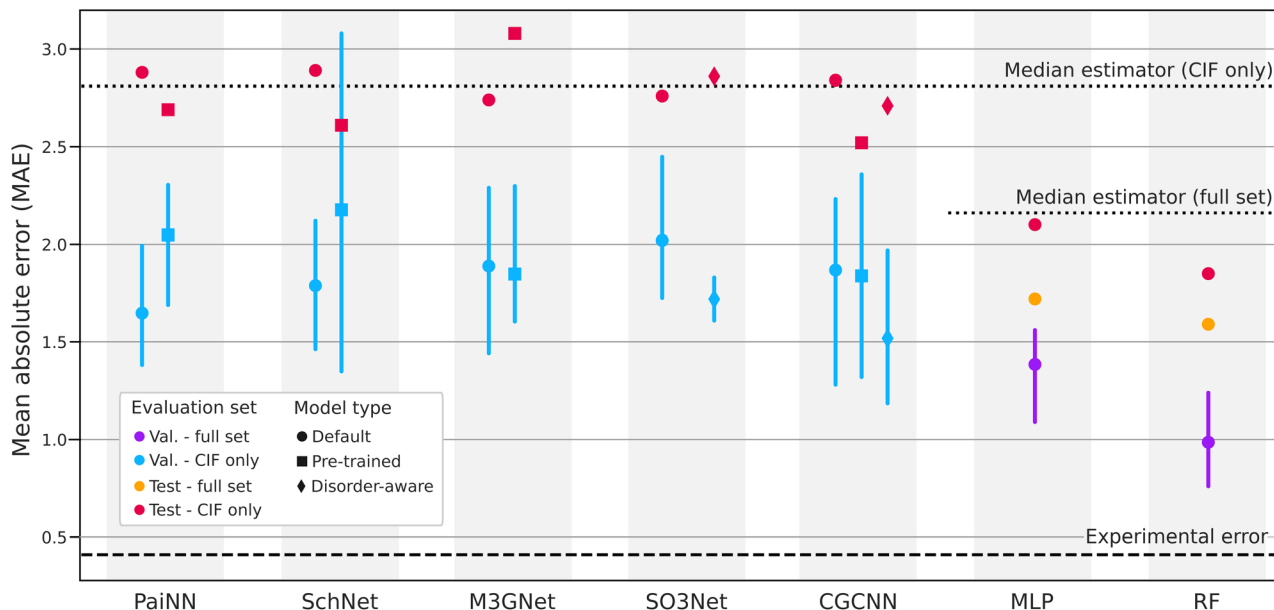


Fig. 4 Benchmarking of various ML models. The same data is tabulated in Table S1. Simpler models outperform geometric GNNs.

set that has CIF files. There are two factors that could explain this result. First, the RF and the MLP used the full training set of 478 structures while the other models were limited to the subset of 254 entries that have CIF files. Second, the geometric models use crystal information to infer properties of the crystal, but they do not properly handle partial occupancies which, as discussed before, are very common in SSE materials and are present in about 3/4 of our CIF files. In order to use these models without modification on our dataset we rounded occupancies to the nearest integers which can lead to important changes in the composition.

To partly verify the above claim that dataset size and the presence of partial occupancy can explain the increased performance of the simple models, we retrained them on the subset of entries that have CIF files only. Doing so, the MAE of the MLP increased to 3.15 while that of the RF was maintained at 1.87. Therefore, dataset size does seem to have a significant impact on the MLP and may explain the difference in performance between that model and the larger models. Random forest still performs well even given less data. Rounding compositions to the nearest integer on the other hand, had little effect on both the RF and the MLP. Rounding compositions is *similar* to rounding site occupancy, but it does not have exactly the same effect. Nevertheless, it indicates that the absence of partial occupancy likely does not explain the difference in performance between the simple models and the more complex ones.

To further explore the effects of partial occupancy, which, as explained in Section 3.1, is an important concept in this field, we introduce new implementations of both CGCNN and SO3Net (dis-CGCNN and dis-SO3Net) that take into account partial occupation (disorder). In both cases, the atomic embedding is replaced with a *site* embedding that is an average over the element embeddings weighted by occupancy. We trained these

models using the same optimal hyperparameters as their original version. The results presented in Fig. 4 and at the bottom of Table S1 show a small improvement in cross-validation performance but it does not translate into significantly better test performance.

The 3D geometric models not only performed poorly compared to simple ML models using less structural information, but their performance on the test set was barely better or sometimes worse than predicting the median of the training set (dotted line in Fig. 4). This shows that these large models can easily overfit small experimental datasets which was also observed in other studies.<sup>35</sup> Moreover, given that the cross-validation splits were chosen randomly within the training set and that the test set was built using the method described in Section 3.3, the relatively large difference in performance between the validation and testing sets illustrate the importance of carefully building leakage-free test sets and that choosing the test set randomly would have most likely led to a false impression of performance.

It is important to note that there may exist more recent GNN architectures that perform better on this task, however given the fact that some of the more recent models tested here still perform equally or close to state-of-the-art models on scalar predictive tasks<sup>34</sup> we do not believe these newer models would perform significantly better on OBELiX and would likely suffer from the same limitations.

Pretraining of 3D geometric models offers some marginal improvements for PaiNN, SchNet and CGCNN. As mentioned in Section 4.2, the pretraining of PaiNN and SchNet restricts the trainable model size which may reduce accuracy while increasing generalizability. This would explain their slightly higher validation MAE and lower training MAE. To measure the effect of the fine-tuning strategy alone we also fine-tuned PaiNN and SchNet by allowing all parameters to change. Under this



strategy, PaiNN and SchNet had validations MAEs of  $1.66 \pm 0.21$  eV and  $1.81 \pm 0.31$  eV while their test MAEs were of 2.60 eV and 2.81 eV respectively. From this limited study, the fine-tuning strategy seems to explain the increased validation MAEs of PaiNN and SchNet in Fig. 4. Therefore, in concert with a more restricted fine-tuning strategy, a better pretrained representation might compensate for the reduced expressivity and increase both accuracy and generalizability, but a much more in-depth analysis of the possible pretraining labels and datasets would be required. In the case of CGCNN and M3GNet which were fine-tuned by allowing all model parameters to change, it is possible that the pretraining property used for CGCNN was “closer” (or more relevant) to ionic conductivity which allowed it to stay in the same weight “basin” and take advantage of the pretrained model's generalizability.

It is important to bear in mind that the variability of the prediction accuracy is high in this small data regime as illustrated by the validation MAEs' standard deviations and that much of the difference between models falls within that variability. Performance is dependent on the (random) choice of cross-validation splits which ultimately dictate the choice of hyperparameters. Complex GNNs with more hyperparameters are more prone to overfitting hyperparameters to a specific set of splits which makes them particularly difficult to tune and compare. Indeed, the variability across folds is smaller for the RF and MLP than for the GNNs.

## 5 Limitations

We have built OBELiX as carefully as possible making sure that all features match the measured ionic conductivity correctly. However, since data is reported and measured in very different ways across journals and decades, there most probably remains inconsistencies between some of the entries especially in terms of atomic positions which are particularly difficult to measure and report. We will continue to improve the dataset as these issues come to light.

Varying factors outside the composition and crystal structure including the measurement conditions (frequency, pressure, measuring device, metalization, *etc.*) and the microstructure (grain size, porosity, phase purity, *etc.*) that depend on the fabrication process (sintering, cold/hot press, pulverization, heat treatment, *etc.*) may have important effects on the measured ionic conductivity. The absence of these factors in OBELiX sets a bound to the performance of the models presented here that is partially, but not fully captured by the experimental uncertainty discussed in Section 4. The repeated materials presented in Fig. 3 could serve as a useful starting point to identify which of these numerous factors have the most impact on the measured conductivity and dictate what additional features could be added to the dataset.

OBELiX is small for ML standards. The difficulty of building an experimental dataset is that there is only a limited number of experiments that were actually performed. Section 4 shows how challenging it is to train existing models on such a small data regime. Ultimately, it highlights the need for models, training architectures and benchmarks tailored for small data regimes,

that could benefit numerous applied fields with similarly limited experimental data (*e.g.* ref. 36). Moreover, OBELiX can be used as a tool to validate and improve molecular dynamics (MD) based methods which are widely applicable across materials science and could later serve as a way to generate a significantly larger computational database of ionic conductivity. For example, in subsequent work, we are currently using a subset of OBELiX to compare the performance of MLFFs and ab-initio methods when predicting ionic conductivity with various MD simulation conditions. Our dataset provides an opportunity to quantitatively test the performance of MLFFs on long timescale MD simulations or ML methods such as LiFlow<sup>19</sup> aimed at accelerating them.

We benchmarked ionic conductivity prediction on our dataset with popular existing models *as is* and using standard training and hyperparameter tuning. We are aware that performance could be improved by modifying the model architectures, training procedure or with data augmentation, but we consider that these methods would not be “baselines” and are outside the scope of this paper.

## 6 Conclusion and outlook

In this paper, we presented OBELiX, a dataset of 599 materials with experimental room temperature ionic conductivities curated by domain experts, including 321 structures with full crystallographic information. We gathered these materials from existing databases and manually extracted data from the literature to build a consistent, easy-to-access database of solid-state electrolyte materials. We benchmarked several ML models and found that the simple random forest model had the best predictive performance. Modern geometric GNNs on the other hand, likely over-fit and were unable to perform well on our carefully designed test set. These findings highlight the immense opportunity for improvement in ML methods specific to this task and tailored for low data regimes.

We hope that OBELiX will serve as a reference point to train and test ionic conductivity models for the ML and computational materials science community in general, ultimately advancing solid-state battery technology.

## Author contributions

Conceptualization: H. S., A. H.-G., H. G., F. T.; data curation: R. H., J. A. H., F. T., H. S.; formal analysis: F. T., J. A. H., D. S.; funding acquisition: Y. B., H. G., H. S.; investigation: all authors; methodology: F. T., J. A. H., D. S.; project administration: F. T., A. H.-G., H. G., H. S.; resources: Y. B., H. G., H. S.; software: F. T., J. A. H., D. S., L. W. M., A. H.-G.; supervision: F. T., A. H.-G., H. G., H. S., S. H.; validation: F. T., J. A. H., D. S., A. H.-G., H. S.; visualization: F. T., J. A. H., H. S.; writing – original draft: F. T., J. A. H., H. S.; writing – review & editing: all authors.

## Conflicts of interest

There are no conflicts of interest to declare.



## Data availability

All data is freely available on our public repository (<https://github.com/NRC-Mila/OBELiX>) as a single csv or xlsx file accompanied by a set of 321 CIF files, including 291 with added random noise. The same data is also available on Kaggle (<https://www.kaggle.com/datasets/flixtherrien/obelix>). Code for benchmarking, configuration files for each experiment as well as data analysis and processing scripts are available on our public repository. All experiments were performed with OBEliX version 1.0.0 (<https://doi.org/10.34740/kaggle/dsv/11789455>). We will continue to update OBEliX with new data. Contributions to the dataset are encouraged through a form on our repository.

Supplementary information (SI): the benchmark models, parity plots for each experiment, hyperparameters and details about resource usage. See DOI: <https://doi.org/10.1039/d5dd00441a>.

## Acknowledgements

The authors acknowledge support from the National Research Council Canada (NRC) through a collaborative R&D grant (AI4D-core-132), Calcul Québec and the Digital Research Alliance of Canada. This project was undertaken thanks to funding from IVADO and the Canada First Research Excellence Fund.

## Notes and references

- J. Janek and W. G. Zeier, *Nat. Energy*, 2016, **1**, 1–4.
- J. Janek and W. G. Zeier, *Nat. Energy*, 2023, **8**, 230–240.
- J. Betz, G. Bieker, P. Meister, T. Placke, M. Winter and R. Schmich, *Adv. Energy Mater.*, 2019, **9**, 1803170.
- S. Zhao, W. Jiang, X. Zhu, M. Ling and C. Liang, *Sustain. Mater. Technol.*, 2022, **33**, e00491.
- A. D. Sendek, Q. Yang, E. D. Cubuk, K.-A. N. Duerloo, Y. Cui and E. J. Reed, *Energy Environ. Sci.*, 2017, **10**, 306–320.
- R. Jalem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki and T. Saito, *Sci. Rep.*, 2018, **8**, 5845.
- B. He, S. Chi, A. Ye, P. Mi, L. Zhang, B. Pu, Z. Zou, Y. Ran, Q. Zhao, D. Wang, *et al.*, *Sci. Data*, 2020, **7**, 151.
- C. J. Hargreaves, M. W. Gaultois, L. M. Daniels, E. J. Watts, V. A. Kurlin, M. Moran, Y. Dang, R. Morris, A. Morscher, K. Thompson, *et al.*, *npj Comput. Mater.*, 2023, **9**, 9.
- F. A. Laskowski, D. B. McHaffie and K. A. See, *Energy Environ. Sci.*, 2023, **16**, 1264–1276.
- D. B. McHaffie, Z. W. Iton, J. M. Bienz, F. A. Laskowski and K. A. See, *Digital Discovery*, 2025, **4**, 1518–1533.
- Y.-J. Shon and K. Min, *ACS Omega*, 2023, **8**, 18122–18127.
- F. Yang, E. C. dos Santos, X. Jia, R. Sato, K. Kisu, Y. Hashimoto, S.-i. Orimo and H. Li, *Nano Mater. Sci.*, 2024, **6**, 256–262.
- G. Ceder, S. P. Ong and Y. Wang, *MRS Bull.*, 2018, **43**, 746–751.
- J. Qi, S. Banerjee, Y. Zuo, C. Chen, Z. Zhu, M. H. Chandrappa, X. Li and S. P. Ong, *Mater. Today Phys.*, 2021, **21**, 100463.
- A. Bielefeld, D. A. Weber and J. Janek, *ACS Appl. Mater. Interfaces*, 2020, **12**, 12821–12833.
- J. Schmidt, M. R. Marques, S. Botti and M. A. Marques, *npj Comput. Mater.*, 2019, **5**, 1–36.
- K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
- D. Wines and K. Choudhary, *arXiv*, 2024, preprint, arXiv:2412.10516, DOI: [10.48550/arXiv.2412.10516](https://doi.org/10.48550/arXiv.2412.10516).
- J. Nam, S. Liu, G. Winter, K. Jun, S. Yang and R. Gómez-Bombarelli, *arXiv*, 2024, preprint, arXiv:2410.01464, DOI: [10.48550/arXiv.2410.01464](https://doi.org/10.48550/arXiv.2410.01464).
- A. Hernandez-Garcia, A. Duval, A. Volokhova, Y. Bengio, D. Sharma, P. L. Carrier, Y. Benabed, M. Koziarski and V. Schmidt, *arXiv*, 2023, preprint, arXiv:2310.04925, DOI: [10.48550/arXiv.2310.04925](https://doi.org/10.48550/arXiv.2310.04925).
- R. Zhu, W. Nong, S. Yamazaki and K. Hippalgaonkar, *Matter*, 2024, **7**, 3469–3488.
- C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, S. Shysheya, J. Crabbé, L. Sun, J. Smith *et al.*, *arXiv*, 2023, preprint, arXiv:2312.03687, DOI: [10.48550/arXiv.2312.03687](https://doi.org/10.48550/arXiv.2312.03687).
- A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
- A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, *APL Mater.*, 2013, **1**, 011002.
- A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 364–369.
- M. Hellenbrandt, *Crystallogr. Rev.*, 2004, **10**, 17–22.
- J. C. M. Madrid and K. K. Ghuman, *Adv. Phys.:X*, 2021, **6**, 1848458.
- J. Riebesell, H. Yang, R. Goodall and S. G. Baird, Pymatviz: visualization toolkit for materials informatics, 2022, <https://github.com/janosh/pymatviz>, DOI: [10.5281/zenodo.7486816](https://doi.org/10.5281/zenodo.7486816).
- K. Schütt, O. T. Unke and M. Gastegger, *Proceedings of the 38th International Conference on Machine Learning*, 2021, pp. 9377–9388.
- K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko and K. Müller, *Neural Information Processing Systems*, 2017.
- C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- K. T. Schütt, S. S. Hessmann, N. W. Gebauer, J. Lederer and M. Gastegger, *J. Chem. Phys.*, 2023, **158**, 144801.
- T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
- S. Liu, W. Du, Y. Li, Z. Li, Z. Zheng, C. Duan, Z.-M. Ma, O. M. Yaghi, A. Anandkumar, C. Borgs, J. T. Chayes, H. Guo and J. Tang, *Adv. Neural Inf. Process. Syst.*, 2024, **36**, 66084–66101.
- V. Fung, J. Zhang, E. Juarez and B. G. Sumpter, *npj Comput. Mater.*, 2021, **7**, 84.
- J. Abed, J. Kim, M. Shuaibi, B. Wander, B. Duijf, S. Mahesh, H. Lee, V. Gharakhanyan, S. Hoogland, E. Irtem, *et al.*, *arXiv*, 2024, preprint, arXiv:2411.11783, DOI: [10.48550/arXiv.2411.11783](https://doi.org/10.48550/arXiv.2411.11783).

