Peertechz

ENGINEERING GROUP

Trends in
Computer Science and Information Technology
OPEN ACCESS JOURNAL

ISSN: 2641-3086 Journal DOI: https://dx.doi.org/10.17352/tcsit

Review Article

Regularization Strategy for Battery Materials Using Lattice Scaling

Eduardo Abenza[1#], César Alonso[1#], Isabel Sobrados[2,3], José M Amarilla[2,3], Javier L Rodríguez[1], José A Alonso[2,3], Roberto GE Martín[1*] and María C Asensio[2,3*]

[1]Department of Artificial Intelligence, HI-Iberia, 28036, Madrid, Spain

[2]Materials Science Institute of Madrid (ICMM/CSIC), Cantoblanco, E-28049, Madrid, Spain

[3]MATINÉE, the CSIC Associated Unit between the Materials Science Institute (ICMUV) and the ICMM, Cantoblanco, E-28049, Madrid, Spain

[#]These authors contributed equally

Submitted : 04 May, 2026
Accepted : 11 May, 2026
Published : 12 May, 2026

***Corresponding author**: María C Asensio, Materials Science Institute of Madrid (ICMM/CSIC), Cantoblanco, E-28049, Madrid, Spain,
E-mail: mc.asensio@csic.es

Keywords: Lattice scaling; Data augmentation; Regularization strategy; Battery materials; Machine learning; Crystal graph convolutional neural networks; Materials informatics; Lithium-Ion batteries

https://www.engineegroup.us

Check for updates

Abstract

Artificial iIntelligence (AI) has emerged as a powerful tool for accelerating materials discovery; however, its effectiveness in battery research remains constrained by the limited size and heterogeneity of available materials datasets. In this work, we introduce a physically informed regularization strategy based on lattice scaling of crystalline structures to reduce overfitting and enhance machine-learning performance in lithium-ion battery materials prediction. The proposed framework generates structurally perturbed variants through controlled isotropic and anisotropic modifications of unit-cell volumes within experimentally realistic limits (±5-30%), mimicking chemo-mechanical variations occurring during electrochemical cycling. The method is evaluated using Crystal Graph Convolutional Neural Networks (CGCNN) trained on electrode materials from the Materials Project Battery Explorer database to predict seven electrochemical and four intrinsic material properties. We demonstrate that lattice scaling - particularly anisotropic scaling within the physically meaningful range of ±5% volume variation - systematically improves predictive accuracy compared with both baseline oversampling and established crystal regularization techniques, achieving improvements of up to 10.5% in the MAD/MAE ratio for key electrochemical descriptors. The results highlight the importance of physically meaningful transformations in materials model regularization and reveal a property-dependent response to regularization strategies. The proposed approach provides a simple, computationally efficient, and architecture-independent pathway to improve data efficiency and generalization in AI-driven battery materials discovery.

Abbreviations

AI: Artificial Intelligence; ML: Machine Learning; CGCNN: Crystal Graph Convolutional Neural Network; GNN: Graph Neural Network; DA: Data Augmentation; LIBs: Lithium-Ion Batteries; CIF: Crystallographic Information File; MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; MAD: Mean Absolute Deviation; DFT: Density Functional Theory

Introduction

The development of advanced functional materials is central to addressing global challenges associated with sustainable energy generation, storage, and conversion. In particular, rechargeable batteries, photovoltaic technologies, and low-power electronic devices critically depend on continuous improvements in materials performance, stability, and efficiency [1,2]. Despite substantial advances in computational materials science and high-throughput simulations, the discovery of optimized materials remains a slow and resource-intensive process due to the combined cost of experimental validation and first-principles calculations.

AI and Machine Learning (ML) methodologies have emerged as transformative tools for accelerating materials discovery, enabling rapid exploration of complex chemical and structural spaces [3-8]. Modern ML frameworks [9] are increasingly capable of predicting materials properties directly from atomic structure and closed-loop active learning strategies [10] and generative deep learning models [11] are being developed

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

to identify stable materials with targeted functionalities. However, the predictive performance and reliability of these models remain fundamentally limited by the availability and quality of training data. Although materials repositories have expanded significantly in recent years [12], materials datasets remain orders of magnitude smaller and substantially more heterogeneous than those available in conventional AI domains such as computer vision or natural language processing [13]. This limitation is particularly critical in battery materials research, where datasets often combine theoretical calculations and experimental measurements obtained under diverse conditions. The resulting data scarcity and variability hinder model generalization and frequently lead to overfitting or reduced transferability.

Addressing limited data availability without compromising physical realism therefore represents a major challenge in materials informatics. Regularization strategies, including those inspired by Data Augmentation (DA), offer an attractive approach to enhance model robustness and reduce overfitting. While DA has proven highly effective in image and signal processing applications, its extension to crystalline materials remains nontrivial. Unlike images, crystal structures obey strict physical and chemical constraints, and naïve geometric transformations - such as arbitrary rotations, translations, or symmetry-breaking distortions - may generate configurations that are structurally unrealistic or inconsistent with underlying structure-property relationships [14-18]. Developing regularization strategies that preserve physical meaning while enriching structural variability thus constitutes an open problem in AI-driven materials science.

Recent efforts have explored regularization approaches based on random perturbations or unit-cell distortions to increase dataset size and mitigate overfitting [19,20]. Although these methods can improve learning efficiency, they often lack a direct connection to experimentally accessible structural variations. From a physical chemistry perspective, effective regularization should ideally reproduce structural fluctuations naturally experienced by materials under realistic operating conditions. Insertion-type battery electrodes provide a compelling framework for such physically grounded regularization. During electrochemical cycling, these materials undergo reversible chemo-mechanical expansion and contraction arising from ion insertion and extraction processes [21,22]. These volume variations modify interatomic distances while largely preserving crystallographic topology, suggesting a natural pathway for generating physically meaningful structural diversity (Figure 1).

Here, we introduce a lattice scaling regularization framework in which crystal structures are systematically perturbed through isotropic and anisotropic modifications of their unit cell volume. A key strength of this approach lies in the use of ±5% volume variations, which directly mimic the chemo-mechanical expansion experimentally observed in insertion-based electrode materials during electrochemical cycling. This physically grounded perturbation range provides realistic structural variability while maintaining

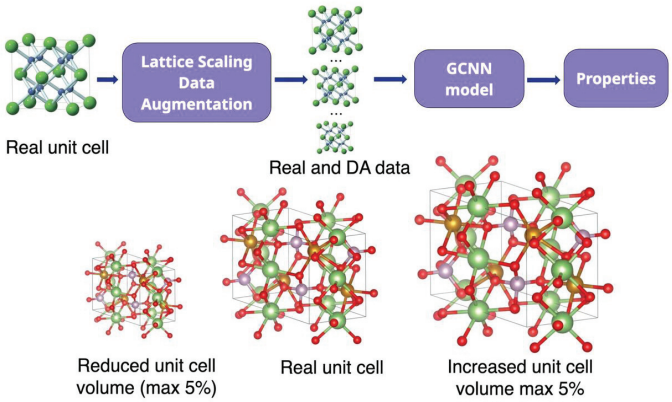

Figure 1: General framework describing the lattice scaling data augmentation method and the pipeline using this data augmentation for the predictive CGCNN model.

physicochemical relevance and most reliably preserving property labels, making it an especially effective and safe regularization choice for battery materials prediction. The proposed strategy is model-agnostic and can be incorporated as a plug-in into both conventional ML algorithms and modern deep-learning architectures, including Graph Neural Networks (GNNs). In this work, we evaluate its effectiveness using the Crystal Graph Convolutional Neural Network (CGCNN) framework [26], applied to cathode materials relevant to metal-ion batteries. Within CGCNN, crystalline structures are represented as atomistic graphs derived directly from Crystallographic Information Files (CIFs), enabling property prediction from structural information alone. By comparing models trained on original and lattice-perturbed datasets, we quantify the impact of physically informed regularization on the prediction of electrochemical and intrinsic material properties. By embedding experimentally grounded structural variability into machine-learning workflows, this work establishes lattice scaling as a physically motivated route toward improving data efficiency, model generalization, and predictive reliability in AI-assisted battery materials discovery. More broadly, the proposed framework highlights the importance of integrating physical insight into data-centric methodologies for next-generation materials design.

Figure 1 illustrates the workflow of the proposed pipeline. The chemical structures obtained from individual CIF files are used as input data, followed by an automated lattice-scaling procedure to generate the regularization dataset. As schematically shown, the CGCNN model is trained using both the original crystal structures and the lattice-perturbed data produced through lattice scaling, representing datasets relevant to battery materials and their associated chemical and electrochemical properties.

The resulting prediction error metrics demonstrate that the lattice-scaling regularization strategy significantly enhances the effectiveness and predictive performance of the CGCNN model. Furthermore, the proposed approach is systematically compared with previously reported crystal-structure regularization methods, including random perturbation, random rotation, random translation, and axis-swapping

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

transformations, using identical datasets and computational protocols. To facilitate adoption and promote dissemination of the lattice-scaling regularization methodology, we provide an open-source Python-based library that can be readily implemented with only a few lines of code. The package includes tutorials and example scripts for testing and application, and it is available upon request from the corresponding authors.

Modeling design, motivation, and methods

The proposed regularization framework applies systematic modifications of crystal unit cells to generate structurally realistic training variants. Rather than primarily aiming to expand dataset size, the central objective is to reduce overfitting by exposing the model to physically plausible structural perturbations during training. We propose a regularization transformation that can be automatically applied to a materials dataset by slightly modifying the unit cell of the materials which constitute the original dataset. Transformations based on volume variations produced by small bond distance changes, distortion of angles, or anisotropic bond distance modifications are widespread in crystal structures and can occur spontaneously. Therefore, regularization based on slightly modifying the size and shape of the unit cells provides a physically meaningful mechanism to improve model generalization in property prediction [27].

This work uses the properties reported at the Battery Explorer - Materials Project platform [28], which is a customized search tool for Li-ion electrode materials. This platform offers information about existing materials and predicts the properties of new ones related to cathodes and anodes of LIBs. In it, relevant properties like the voltage, capacity, energy density, specific density, voltage profile, oxygen evolution, and lithium positive ion diffusivity, among others, describe the materials. This Database mainly integrates results from high-throughput computing research based on quantum mechanics, solid state physics and statistical mechanics, and selected experiments. This approach, led by the Ceder group, is widely described in the following scientific reports, [29-33]. This work uses this database with and without the regularization process to train the graph-based predictive model CGCNN. The selected input structures are materials with application in insertion-based batteries, such as metal-ion batteries, in the discharged state (i.e., where the cathode is "full" of the mobile ion). For the lattice scaling regularization technique, the volume of the materials has been modified in different ranges of percentages. We consider that the most physically meaningful range of percentages has been from -5% to +5%. Performing the regularization transformation with such a small volume variation is aimed at increasing the 'safety' of the transformation. The label-preservation assumption underlying this approach deserves careful justification. A ±5% change in unit cell volume induces small perturbations of interatomic distances that do not alter the chemical identity, oxidation states, or coordination topology of the material. For electrochemical properties such as voltage and capacity, which are primarily governed by the thermodynamic energy difference between charged and discharged states, minor isotropic or anisotropic volume changes are unlikely to significantly modify these values. For intrinsic properties such as Fermi energy and band gap energy, which are more sensitive to structural details, a ±5% volume variation may induce small shifts in the electronic band structure; however, these perturbations remain within the range of thermal and quantum fluctuations that materials experience under experimental conditions. Consequently, while label preservation is not strictly guaranteed for all properties - particularly for structure-sensitive quantities - it constitutes a reasonable approximation within the ±5% volume variation regime employed here. In data augmentation contexts, the safety of a transformation has been defined as the likelihood that the transformed data retains the same label as the original data [21]. While label preservation is not guaranteed, a change of ± 5% in volume will be more likely to maintain the labels than a greater modification (e.g., ± 30%). Furthermore, volume changes up to 5% have been experimentally observed in insertion-based LIBs during the charge/discharge steps of their life cycle, supporting the notion that such small volume changes can preserve the labels of the data [23-25].

In symmetric lattice scaling, the volume of the unit cell has been increased or decreased randomly (within limits), but the relationships between the three lengths of the unit cell (a,b,c) are kept constant. In asymmetric lattice scaling, unit cell lengths (a,b,c) have also been increased and decreased randomly (and therefore also the volume) but changing the values of a, b, and c independently. Results are shown for two cases where the maximum limits of volume variation were chosen to be 5% and 30%. To feed CGCNN, the crystalline systems are represented as graphs, where atoms constitute the graph's nodes. The bonds in these systems are represented by the edges between the nodes, where all atoms within an appropriate radius are considered bonded. Nodes and edges are associated with vector representations (attributes) that enhance the model training. In fact, node attributes encode chemical information about the chemical element in each node. In contrast, edge attributes encode the bond distance between the corresponding pair of atoms. The lattice scaling regularization method slightly changes the interatomic distances. Hence the CGCNN graphs are only modified at the edge level, representing the bonds between atoms. Because the atomic bonds are based on the interatomic distance, some atoms might be considered bonded when they previously were not, or vice versa. Furthermore, the edge attributes will also be modified as the distance varies. These changes at the edge level are the reason why the lattice scaling method increases the amount of data present in the dataset. For each initial material, there will be multiple different crystal graphs that will be processed by the CGCNN model.

We have focussed on the results of the recent high-throughput search on cathode materials. The CGCNN model has been applied to predict the seven electrochemical properties and four material properties described in Table 1. The model has learned from a,b initio computational results available in Battery Explorer. Briefly, the properties of potential cathodes for LIBs are obtained through first principles computations and high-throughput computational screening approaches described in previous reports [34,35].

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

Table 1: List of the electrochemical and material properties predicted using the CGCNN model.

Property category	Property	Description
Electrochemical	Average voltage	Amount of electrical potential a battery holds. Units: V.
Electrochemical	Gravimetric capacity	Amount of electric charge an electrode material can store normalized by weight. Units: mAh/g.
Electrochemical	Gravimetric energy	The specific energy of a battery is a measure of how much energy a battery can store normalized by weight. Units: Wh/kg.
Electrochemical	Maximum voltage	The maximum voltage of a battery is the highest voltage among the different steps between total charge and total discharge. Units: V.
Electrochemical	Minimum voltage	The minimum voltage of a battery is the lowest voltage among the different steps between total charge and total discharge. Units: V.
Electrochemical	Volumetric capacity	Amount of electric charge an electrode material can store normalized per volume. Units: Ah/l.
Electrochemical	Volumetric energy	Energy density of a battery is a measure of how much energy a battery can store normalized per volume. Units: Wh/l.
Material	Formation energy	Energy of the material with respect to standard states (elements), normalized per atom. Units: eV/atom.
Material	Energy	Potential energy of the material, normalized per atom. Units: eV/atom.
Material	Fermi energy	Energy required to add an electron to the material. Units: eV.
Material	Band gap energy	Energy difference between the top of the valence band and the bottom of the conduction band in the electronic structure of the material. Units: eV.

Datasets

The initial dataset was formed by 4401 electrode active materials for metal-ion batteries, obtained from the Battery Explorer of the Materials Project database on 2021-12-21. To train a CGCNN model, this initial dataset is split into a training set, validation set and test set. The training set, composed by 80% of the materials of the initial set, is used to directly train the model. The validation set, composed by 10% of the materials, guides the training of the model. Finally, the remaining 10% of the materials form the evaluation set, which will be used for the final evaluation of the trained model's performance.

For each regularization technique, the initial training set was augmented, generating a new perturbed training set. In each dataset, the initial training set was repeated 10 times, and the materials were subjected to the pertinent regularization transformation. This results in a dataset in which, apart from the original materials present in the initial training set, there are ten new materials for each original material. The validation and evaluation sets remained non-augmented, encouraging a fairer evaluation of the technique.

Four datasets have been generated for the lattice scaling regularization technique.

- First, two datasets were generated with randomly sampled volumes between 95 and 105% of the volume of the original material. This ±5% volume range is highlighted as the most physically meaningful regularization choice, directly reflecting the volume changes experimentally observed in insertion-based LIBs during charge/discharge cycling, and most reliably preserving material property labels. Regarding these two datasets, transformations were applied in both an isotropic and an anisotropic fashion. In isotropic regularization, all lattice lengths change in the same proportion, while in anisotropic transformations the lattice lengths vary in different proportions.
- The remaining two datasets were generated with randomly sampled volumes between 70 and 130% of the volume of the original material. Transformations were also applied in an isotropic and anisotropic manner.

All lattice scaling transformations were applied using the Python library "pymatgen" [36].

In summary, the size of the training, validation, and evaluation sets before and after the lattice scaling regularization is indicated in Table 2.

To compare the efficiency of the lattice scaling regularization method presented in this work, we have carried out similar experiments from the same initial dataset described in Table 2 but employing four additional state-of-the-art regularization transformations for crystalline chemical structures. These four regularization transformations were (1) Random Perturbation, (2) Random Rotation, (3) Random Translate, and (4) Swap Axes. The same splits described in Table 2 have been used for all tested regularization transformations.

These transformations, performed using the Python library "AugLiChem", are briefly described as follows:

- Random Perturbation: This regularization transformation perturbs all the sites of the crystalline compound unit cell by a short distance between 0 and 0.5 Å. This approach is very effective because, with such a small perturbation, the main symmetry elements of the system are usually radically changed.
- Random Rotation: In the Random Rotation regularization transformation, all the sites in the crystal are randomly rotated between 0 and 360 degrees.
- Random Translate: In this regularization transformation, only a few crystal sites are displaced by a distance ranging from 0 to 0.5 Å.
- Swap Axes: In this regularization transformation, the coordinates of the sites in the crystal are swapped

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

Table 2: Overview of the datasets used as input for property prediction using the CGCNN model, with and without Regularization. Regularized materials are used only for training.

Datasets	Training dataset N° of materials	Validation dataset N° of materials	Evaluation dataset N° of materials
Initial Dataset	3515	439	440
After Regularization	38665	439	440

between two axes, for example, between the x- and y-axes. With this transformation, the locations of all the crystal sites are considerably displaced.

Additionally, we have also prepared a baseline dataset using oversampling, in which each instance of the initial training dataset has been repeated ten times, but without any regularization transformation. The motivation of this baseline dataset is to have a dataset that is not transformed, but that contains the same amount of training samples as the regularized datasets. Therefore, the regularization techniques will be compared with this baseline dataset, so that the difference in performance cannot be attributed to differences in the amount of training data employed by the CGCNN model.

To obtain a more robust performance evaluation, the entire process has been replicated three times for each method (e.g., for anisotropic lattice scaling 5%, three different datasets have been created). In each replicate, we have followed the procedure described above, only changing the random seed. This affects which materials belong to the training, validation, and evaluation sets; the regularization transformations; and the training of the CGCNN model.

The complete dataset generation is described in Figure 2.

Results and discussions

The enhancement in property prediction performance of the CGCNN model due to the lattice scaling regularization has been evaluated by carrying out experiments to predict seven electrode materials' electrochemical properties and four material properties based on the chemical structure datasets listed in Table 2. These eleven properties have been documented in Table 1. The same procedure has been performed using the baseline dataset, and the regularized datasets. The first metric considered to evaluate the models is the Mean Absolute Error (MAE), defined as the mean of the errors (in absolute value) that the model commits when predicting the corresponding property. The lower the MAE value, the better the model behaves. In addition, the Root Mean Squared Error (RMSE) has been calculated as another performance metric. It is the square root of the mean of the errors (in quadratic value) that the model commits when predicting the respective property. A lower RMSE value indicates a better performance of the model.

Furthermore, since each property has associated different units and variance, the Mean Absolute Deviation (MAD) of each predicted property has been calculated in the evaluation set to facilitate an unbiased comparison of the behaviour of the model applied to properties with different natures. MAD measures how spread out the data is in the evaluation set, like the standard deviation; it is the mean of the absolute difference between all the data and the mean value of the corresponding property [37]. The value of MAD represents the MAE that would be obtained by a random model that predicts the mean value of the dataset for any instance. Therefore, the MAD/MAE ratio has been evaluated, as it allows a better comparison between different properties. A model with a high MAD/MAE ratio (5 or more) is considered an excellent predictive model. Finally, all metrics and predictions have been obtained using the same CGCNN model architecture but applying the different regularization methods to the dataset.

Regarding the prediction of the electrochemical properties, Table 3 compares the lattice scaling regularization methods and the previous state-of-the-art regularization techniques with the baseline dataset. For the lattice scaling regularization, the anisotropic lattice scaling (5%) improves the prediction of every property except for gravimetric capacity, for which this technique is neutral. The anisotropic lattice scaling (30%) improves every property prediction, except for minimum voltage, where it is neutral, and gravimetric capacity, where it worsens the performance. On the other hand, isotropic lattice scaling (5%) improves the prediction of every electrochemical property except for minimum voltage, where it is neutral, or gravimetric energy, where it performs worse than the baseline dataset. Lastly, isotropic lattice scaling (30%) only improves the predictions of gravimetric capacity, volumetric capacity, and maximum voltage, worsening the predictions of the rest of the properties.

For every electrochemical property except gravimetric energy, one or multiple lattice scaling regularization methods give the best MAD/MAE improvement out of all the regularization techniques, performing better than the state-of-the-art approaches. It is particularly noteworthy that the ±5% anisotropic lattice scaling - grounded in experimentally observed chemo-mechanical volume changes in electrode materials - consistently delivers strong improvements across electrochemical properties, underscoring the value of this physically motivated regularization choice. We see remarkable

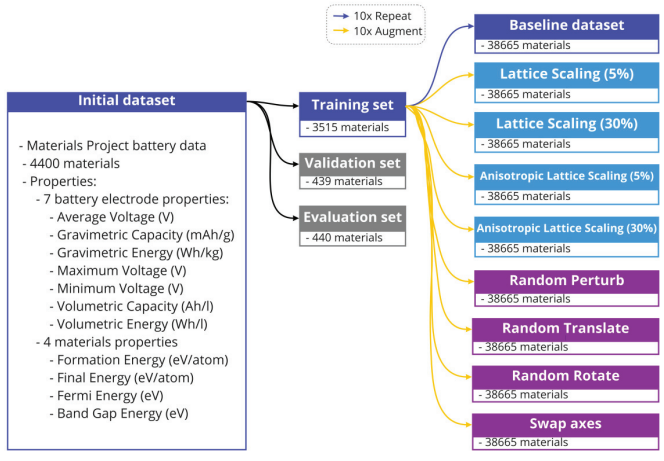

Figure 2: General framework describing the lattice scaling data augmentation method and the pipeline using this data augmentation for the predictive CGCNN model.

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

Table 3: Comparison of electrochemical property prediction with different techniques of Regularization strategy, expressed as the mean MAD/MAE improvement with respect to the baseline dataset. For MAD/MAE ratio, a higher metric is indicative of a better performance. The higher, the better. For a recapitulative of the comparative metrics on the prediction of the CGCNN model for all methods (MAE, RMSE, and MAD/MAE ratio), see Tables S1-S7.

	Average Voltage (V)	Gravimetric Capacity (mAh/g)	Gravimetric Energy (Wh/kg)	Maximum voltage (V)	Minimum voltage (V)	Volumetric capacity (Ah/l)	Volumetric Energy (Wh/l)
Anisotropic Lattice Scaling (5%)	4.6%	0.5%	4.2%	3.1%	3.3%	7.3%	9.2%
Anisotropic Lattice Scaling (30%)	6.0%	-1.1%	6.9%	10.5%	0.7%	2.4%	7.8%
Isotropic Lattice Scaling (5%)	2.5%	1.4%	-2.2%	4.4%	-0.3%	3.0%	7.8%
Isotropic Lattice Scaling (30%)	-3.0%	1.8%	0.6%	2.4%	-5.0%	4.8%	-1.5%
Random Translate	-1.8%	-3.8%	-6.1%	-2.4%	-1.9%	1.6%	2.7%
Random Perturb	-3.4%	-7.8%	10.7%	9.3%	-0.2%	-0.6%	4.3%
Random Rotate	-8.1%	-2.5%	-3.1%	-0.8%	-8.9%	0.7%	1.8%
Swap Axes	-7.5%	-1.0%	-7.2%	-7.1%	-12.8%	-1.6%	-2.2%

improvements, up to 10.5% in the case of anisotropic lattice scaling (30%) applied to maximum voltage prediction. Regarding gravimetric energy, although anisotropic lattice scaling performs better than the baseline, the best results are obtained by the random perturbation regularization method.

An additional model was also trained with the initial dataset, but it consistently showed a worse performance than the other methods, since it contained 10 times less data than the baseline or the regularized datasets.

It should be noted that, except for a few cases, the state-of-the-art regularization techniques for crystalline materials perform worse than the baseline dataset, where we are simply oversampling 10 times the initial dataset without any additional transformation. This topic will be discussed later (Table 3).

The same strategy has been followed to evaluate the effectiveness of the lattice scaling regularization transformation on the enhancement of the materials' properties prediction using the CGCNN model. Apart from the electrochemical properties of the electrode materials, the CGCNN models were also trained to predict four materials properties: formation energy, energy, Fermi energy and band gap energy. The applicability of the proposed regularization strategy differs importantly across these properties, and it is necessary to clearly delineate its limitations. For formation energy, every regularization technique performed worse than the baseline. This result has a clear physical interpretation: formation energy depends strongly on interatomic distances and the exact geometry of the coordination environment. Since the lattice scaling transformation directly modifies these structural parameters, the assigned labels no longer accurately reflect the true formation energy of the scaled structures. In other words, the label-preservation assumption breaks down for formation energy, and any structural regularization transformation necessarily introduces label noise that degrades predictive accuracy. The method is therefore not recommended for formation energy prediction, and caution should be exercised for any property that is tightly coupled to specific bond lengths or angles. For energy per atom, the impact of every regularization technique was either neutral or positive, with isotropic lattice scaling (5%) providing the greatest improvement, followed by isotropic lattice scaling (30%). In the case of Fermi energy and band gap energy - both intrinsic electronic properties known to be sensitive to crystal structure - the label-preservation assumption requires careful consideration. The Fermi energy depends on the electronic density of states, which shifts as the band structure changes with lattice parameter variations. Similarly, band gap energy is affected by crystal field splitting and orbital hybridization, both of which are modulated by interatomic distances. Despite these sensitivities, both types of lattice scaling at ±5% had a positive impact on Fermi energy prediction, and anisotropic lattice scaling (5%) provided the greatest improvement for band gap energy. This indicates that within the physically meaningful ±5% range, the regularization benefit - reduced overfitting - outweighs the label noise introduced by the perturbation. In contrast, most state-of-the-art regularization methods impaired performance for these properties, suggesting that physically grounded, small-magnitude perturbations are essential when regularizing predictions of structure-sensitive electronic properties. For larger volume variations (30%), results become mixed, consistent with the increased label inconsistency expected at greater structural distortions beyond the physically realistic regime. In the case of Fermi energy, both types of lattice scaling (5%) had a positive impact, while most of the state-of-the-art regularization methods impaired the performance of the model. Finally, band gap energy improved the most with anisotropic lattice scaling (5%), having mixed results with the other techniques (Table 4).

In summary, lattice scaling regularization, and especially anisotropic lattice scaling with a maximum volume change of 5%, generally improve the performance of the CGCNN predictive model in comparison to the baseline dataset and to the other state-of-the-art regularization techniques (with some exceptions). The ±5% volume variation range is particularly highlighted as a key strength of this framework: it directly mimics the chemo-mechanical expansion observed in insertion-based battery electrodes during cycling, constitutes a physically meaningful perturbation, and most reliably preserves material property labels - making it the recommended regularization setting for battery materials prediction.

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

In most of the cases, the state-of-the-art regularization techniques tend to perform worse than the baseline dataset. In this study, we have used the default configuration for the different regularization methods, which could be one of the reasons for this substandard performance. Using hyperparameter tuning to select the best configuration for each technique could be essential in improving their performance in material and electrochemical property prediction. Likewise, this tuning could be applied to our lattice scaling method to further increase the predictive performance of the CGCNN model (Figure 3).

It should be noted that the performance of each regularization technique seems to be related to the property that is being predicted. The prediction of some material properties, such as formation energy, is always impaired when using structural regularization, independently of the transformation. As discussed above, this is not a coincidental observation but reflects a fundamental limitation: lattice scaling regularization modifies the very structural parameters that govern formation energy, invalidating the label-preservation assumption for this property. Practitioners should therefore be aware that the applicability of lattice scaling regularization is property-dependent: it is most beneficial for properties governed by long-range structural features or overall material composition, and least suitable for properties tightly coupled to specific short-range bonding geometries. On the contrary, properties such as volumetric energy improve with almost every regularization transformation when comparing it to the baseline. By applying a regularization transformation while maintaining the same labels, we are biasing the model into considering that this transformation does not affect the predicted property. For instance, in the case of formation energy, we hypothesize that the volume of the unit cell and, in general, atomic displacements, affect the final value of the property. Therefore, when applying any regularization technique, the prediction would not improve on this property, because it depends on the parameters that we are modifying. The same reasoning could be applied to the other materials and electrochemical properties.

Furthermore, the initial dataset from Materials Project contains electrode active materials of different metal-ion batteries. A recent work on image regularization via data augmentation shows that certain regularization techniques can enhance the prediction on certain classes, while impairing the prediction on other classes [38]. In this work, we evaluate the regularization techniques without distinguishing between the different inserted metal ions. However, it is known that insertion-based batteries with different metal ions experiment different volume changes during their charge and discharge steps, as seen in the Materials Project database [39-42]. Given the difference in volume changes and the fact that the lattice scaling works by modifying the volume of the materials, it would be interesting to analyse the performance of this technique on the batteries with different mobile ions to evaluate whether it depends on the type of ion. This knowledge could be useful for practical applications, to choose the method that better improves the prediction on LIBs instead of metal-ion batteries in general.

Table 4: Comparison of material property prediction with different techniques of Regularization strategy, expressed as the mean MAD/MAE improvement with respect to the baseline dataset. For MAD/MAE ratio, a higher metric is indicative of a better performance. For a recapitulative of the comparative metrics on the prediction of the CGCNN model for all methods (MAE, RMSE, and MAD/MAE ratio), see Tables S8-S11.

	Formation energy (eV/atom)	Energy (eV/atom)	Fermi energy (eV)	Band gap energy (eV)
Anisotropic Lattice Scaling (5%)	-0.6%	0.2%	2.6%	3.7%
Anisotropic Lattice Scaling (30%)	-15.6%	-1.9%	-0.1%	1.6%
Isotropic Lattice Scaling (5%)	-8.3%	3.8%	2.2%	0.1%
Isotropic Lattice Scaling (30%)	-1.1%	1.8%	-3.6%	-4.0%
Random Translate	-9.2%	0.6%	1.6%	1.4%
Random Perturb	-16.9%	-4.5%	-6.2%	0.2%
Random Rotate	-8.0%	-0.8%	-3.2%	-4.5%
Swap Axes	-4.1%	-2.9%	-4.8%	-5.4%

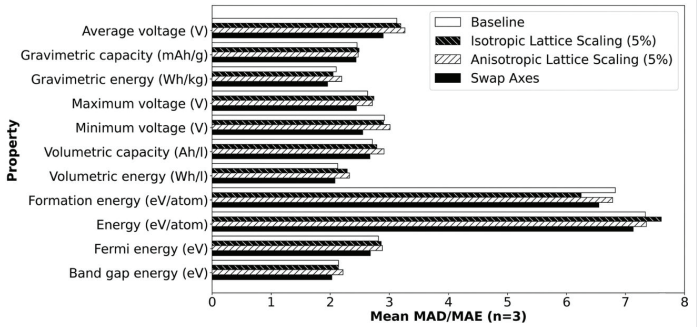

Figure 3: Comparison of the mean MAD/MAE ratio between isotropic and anisotropic lattice scaling (5%), the Swap Axes DA transformation, and the baseline dataset.

Limitations of the study

Although the proposed lattice-scaling regularization strategy demonstrates significant improvements for several electrochemical and intrinsic material properties, certain limitations should be acknowledged. First, the effectiveness of the method is strongly dependent on the structural sensitivity of the target property. Properties such as formation energy, which are highly dependent on precise interatomic distances and local bonding environments, may experience degraded predictive performance due to violations of the label-preservation assumption.

Second, the present study focuses primarily on insertion-based battery materials obtained from the Materials Project Battery Explorer database and does not explicitly distinguish between different mobile-ion chemistries. Since different battery systems may exhibit distinct chemo-mechanical behaviors and volume expansion characteristics, the optimal regularization regime may vary across material classes.

Additionally, the regularization parameters employed in this work were selected based on physically motivated volume ranges rather than systematic hyper-parameter optimization. Further tuning of augmentation magnitude and transformation

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

strategies may yield additional improvements in predictive performance.

Finally, while the proposed approach improves model generalization, the augmented structures should not be interpreted as fully independent ground-truth configurations. Instead, they represent controlled perturbative approximations intended to improve learning robustness within physically realistic structural regimes.

Conclusion

In conclusion, lattice scaling is a powerful regularization technique to improve the prediction of some electrochemical and material properties. Its effectiveness is most pronounced within the physically meaningful ±5% volume variation range, which stands out as a key strength of the proposed framework by directly reflecting real chemo-mechanical processes in insertion-based battery electrodes. It could have many possible near-future applications in predicting electrochemical properties, as it can ensure substantial gains over other standard regularization approaches, given that, with some exceptions, the standard state-of-the-art regularization transformations produce a worse performance than simply oversampling the initial dataset.

Acknowledgements

The present research has been undertaken in the context of the Associated Research Unit MATINÉE of the CSIC (Spanish Scientific Research Council), created between the Institute of Materials Science (ICMUV) of Valencia University and the Materials Science Institute of Madrid (ICMM). The authors acknowledge financial support from the project INGENIOUS (TED2021-132656B-C21 & TED2021-132656B-C22) entitled "Lithium-ion batteries: an effective methodology for the challenge of the optimization and regeneration of their main components," granted by the Call 2021 - "Ecological Transition and Digital Transition Projects." Promoted by the Ministry of Science and Innovation, Funded by the European Union within the "NextGeneration" EU program, the Recovery, Transformation, and Resilience Plan, and the State Investigation Agency. HI-IBERIA has fully supported the computing work.

Data availability

The baseline dataset used in this study is publicly available from the Materials Project repository, https://materialsproject.org/, Battery Explorer section. The raw data were accessed and transformed for the purposes of our analysis. The transformed data, the scripts used for data download, transformation, and processing, are available from the corresponding authors (RGEM or MCA) upon request.

Author contributions

EA, CA and MCA conceived the idea and designed the calculations. MCA wrote the first draft of the paper, which was enriched with further contributions from all the authors. RGEM and MCA organized the research project, and EA and CA participated in the computational design and discussion with IS, JMA, JLR, JAA, RGEM, and MCA. All the authors were deeply involved in the analyses of input data and computational results.

Additional information

Regularization strategy for battery materials using lattice scaling _SM.pdf

(Supplementary-Tables)

References

1. Goodenough JB, Kim Y. Challenges for Rechargeable Li Batteries. Chem Mater. 2010;22:587-603. Available from: https://doi.org/10.1021/cm901452z
2. Buriak JM, Toro C. Rising to the Challenge: John B. Goodenough and Youngsik Kim, and "Challenges for Rechargeable Li Batteries". Chem. Mater. 2015;27:5149-5150. Available from: https://pubs.acs.org/doi/10.1021/acs.chemmater.5b02863
3. Sendek AD, Yang Q, Cubuk ED, Duerloo KN, Cui Y, Reed EJ. Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. Energy Environ Sci. 2017;10:306-320. Available from: https://doi.org/10.1039/C6EE02697D
4. Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. Nat Rev Mater. 2020. Available from: https://doi.org/10.1038/s41578-020-00255-y
5. Lombardo T, Duquesnoy M, El-Bouysidy H, Årén F, Gallo-Bueno A, Jørgensen PB, et al. Artificial Intelligence Applied to Battery Research: Hype or Reality? Chem Rev. 2022;122)(12):10899-10969. Available from: https://doi.org/10.1021/acs.chemrev.1c00108
6. Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning, J Mat. 2017;3:159-177. Available from: https://doi.org/10.1016/j.jmat.2017.08.002
7. Barrett DH, Haruna A. Artificial intelligence and machine learning for targeted energy storage solutions. Curr Opin Electrochem. 2020;21:160-166. Available from: https://doi.org/10.1016/j.coelec.2020.02.002
8. Materials Genome Initiative for Global Competitiveness. 2011. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.mgi.gov/sites/mgi/files/materials_genome_initiative-final.pdf
9. Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. NPJ Comput Mater. 2019;5:83. Available from: https://doi.org/10.1038/s41524-019-0221-0
10. Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11:5966. Available from: https://doi.org/10.1038/s41467-020-19597-w
11. Woo S, Shenvi RA. Natural Product Synthesis through the Lens of Informatics. Acc Chem Res. 2021;54(5):1157-1167. Available from: https://doi.org/10.1021/acs.accounts.0c00791
12. Morgan D, Jacobs R. Opportunities and Challenges for Machine Learning in Materials Science. Annu Rev Mater Res. 2020;50(1):71-103. Available from: https://doi.org/10.1146/annurev-matsci-070218-010015
13. Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. NPJ Comput Mater. 2018;4:25. Available from: https://doi.org/10.1038/s41524-018-0081-z
14. Puchala B, Tarcea G, Marquis EA, Hedstrom M, Jagadish HV, Allison JE. The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community. JOM. 2016;68:2035-2044. Available from: https://doi.org/10.1007/s11837-016-1998-7

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110

15. Bruno I, Gražulis S, Helliwell JR, Kabekkodu SN, McMahon B, Westbrook J. Crystallography and Databases. Data Sci J. 2017;16:38. Available from: https://doi.org/10.5334/dsj-2017-038

16. Allen FH, Shields GP. Crystallographic Databases and Knowledge Bases in Materials Design. Nato Science Series. 1999. Available from: https://doi.org/10.1007/978-94-011-4653-1_21

17. Groom CR, Bruno IJ, Lightfoot MP, Ward SC. The Cambridge Structural Database. Acta Crystallogr B Struct Sci Cryst Eng Mater. 2016;72:171-179. Available from: https://doi.org/10.1107/S2052520616003954?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle

18. Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sánchez-Carrera RS, Gold-Parker A, et al. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. J Phys Chem Lett. 2011;2:2241-2251. Available from: https://doi.org/10.1021/jz200866s

19. Magar R, Wang Y, Lorsung C, Liang C, Ramasubramanian H, Li P, et al. AugLiChem: Data Augmentation Library of Chemical Structures for Machine Learning. 2021). Available from: https://doi.org/10.48550/arXiv.2111.15112

20. Maharana K, Mondal S, Nemade B. A Review: Data Pre-Processing and Data Augmentation Techniques. Global Transitions Proceedings. 2022;3(1):91-99. Available from: https://doi.org/10.1016/j.gltp.2022.04.020

21. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019;6:60. Available from: https://doi.org/10.1186/s40537-019-0197-0

22. Magar R, Wang Y, Lorsung C, Liang C, Ramasubramanian H, Li P, et al. AugLiChem: Data Augmentation Library of Chemical Structures for Machine Learning ArXiv. 2021. Available from: https://ui.adsabs.harvard.edu/link_gateway/2022MLS&T...3d5015M/doi:10.1088/2632-2153/ac9c84

23. Ishidzu K, Oka Y, Nakamura T. Lattice volume change during charge/discharge reaction and cycle performance of Li[NixCoyMnz]O2. Solid State Ionics. 2016;288:176-179. Available from: https://doi.org/10.1016/j.ssi.2016.01.009

24. Kriston A, Ruiz V, Pfrang A, Kersys A, Marinaro M, Stegmaier P, et al. On the correlation between volume change of anode materials in Li-ion cells and their degradation and failure. Collection of open conferences in research transport. 2018;2018:128. Available from: https://www.scipedia.com/public/Kriston_et_al_2018a

25. Koerver R, Zhang W, de Biasi L, Schweidler S, Kondrakov AO, Kolling S, et al. Chemo-mechanical expansion of lithium electrode materials – on the route to mechanically optimized all-solid-state batteries. Energy Environ Sci. 2018;11:21422158. Available from: https://doi.org/10.1039/C8EE00907D

26. Xie T, Grossman JC. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. Phys Rev Lett. 2018;120:145301. Available from: https://doi.org/10.1103/PhysRevLett.120.145301

27. Magar R, Wang Y, Farimani AB. Crystal Twins: Self-supervised Learning for Crystalline Material Property Prediction. 2022NPJ Comput Mat. Available from: https://doi.org/10.1038/s41524-022-00921-5

28. Battery Explorer-Materials Project. Available from: https://materialsproject.org/batteries

29. Hautier G, Fischer CC, Jain A, Mueller T, Ceder G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. Chem Mater. 2010;22:3762-3767. Available from: https://doi.org/10.1021/cm100795d

30. Urban A, Seo DH, Ceder G. Computational understanding of Li-ion batteries. NPJ Comput Mater. 2016;2:16002. Available from: https://doi.org/10.1038/npjcompumats.2016.2https://doi.org/10.1038/npjcompumats.2016.2

31. Wang Y, Richards WD, Ong SP, Miara LJ, Kim JC, Mo Y, et al. Design principles for solid-state lithium superionic conductors. Nature Mater. 2015;14(10): 1026-1031. Available from: https://doi.org/10.1038/nmat4369

32. Urban A, Lee J, Ceder G. The Configurational Space of Rocksalt-Type Oxides for High-Capacity Lithium Battery Electrodes. Adv Energy Mater. 2014;4:1400478. Available from: https://doi.org/10.1002/aenm.201400478

33. Ceder G. Opportunities and challenges for first-principles materials design and applications to Li battery materials. MRS Bull. 2010;35:693-701. Available from: https://doi.org/10.1557/mrs2010.681

34. Hautier G. Prediction of new battery materials based on ab initio computations. 2016. Available from: https://doi.org/10.1063/1.4961901

35. Urban A, Seo DH, Ceder G. Computational understanding of Li-ion batteries. NPJ Comput Mater. 2016;2:16002. Available from: https://doi.org/10.1038/npjcompumats.2016.2

36. Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, et al. Python Materials Genomics (pymatgen): A robust, open-source Python library for materials analysis. Comput Mater Sci. 2013;68:314-319. Available from: https://doi.org/10.1016/j.commatsci.2012.10.028

37. Choudhary K, DeCost B. Atomistic Line Graph Neural Network for improved materials property predictions. NPJ Comput Mater. 2021;7:185. Available from: https://doi.org/10.1038/s41524-021-00650-1

38. Balestriero R, Bottou L, LeCun Y. The Effects of Regularization and Data Augmentation are Class Dependent. 2022. Available from: https://proceedings.neurips.cc/paper_files/paper/2022/hash/f73c04538a5e1cad40ba5586b4b517d3-Abstract-Conference.html

39. Ong SP, Jain A, Hautier G, Kang B, Ceder G. Thermal stabilities of delithiated olivine MPO4 (M=Fe, Mn) cathodes investigated using first principles calculations. Electrochemistry Communications. 2010;12:427-430. Available from: https://doi.org/10.1016/j.elecom.2010.01.010

40. Wang L, Maxisch T, Ceder G. A First-Principles Approach to Studying the Thermal Stability of Oxide Cathode Materials. Chem Mater. 2007;19: 543-552. Available from: https://pubs.acs.org/doi/10.1021/cm0620943

41. Zhou F, Cococcioni M, Marianetti CA, Morgan D, Ceder G. First-principles prediction of redox potentials in transition-metal compounds with LDA + U. Phys Rev B. 2004;7:235121. Available from: https://doi.org/10.1103/PhysRevB.70.235121

42. Adams S, Rao RP. High power lithium ion battery materials by computational design: High power Li ion battery materials by computational design. Phys Status Solidi A. 2011;208:1746-1753. Available from: https://doi.org/10.1002/pssa.201001116

Citation: Abenza E, Alonso C, Sobrados I, Amarilla JM, Rodríguez JL, Alonso JA, et al. Regularization Strategy for Battery Materials Using Lattice Scaling. Trends Comput Sci Inf Technol. 2026;11(2):49-57. Available from: https://dx.doi.org/10.17352/tcsit.000110