Materials Horizons

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/materials-horizons

View Article Online
DOI: 10.1039/D5MH01525A

Wider Impact Statement

This review addresses the critical intersection of machine learning and solid-state electrolyte development, a field experiencing unprecedented growth with hundreds of publications emerging in recent years. Key developments discussed include the evolution from classical ML screening approaches to sophisticated deep learning architectures like graph neural networks, the emergence of ML interatomic potentials enabling large-scale dynamics simulations, and the transition toward generative models for de novo materials design. The field's significance extends beyond academic interest: solid-state electrolytes are essential for next-generation batteries that promise enhanced safety, energy density, and sustainability for electric vehicles and grid storage applications. The rapid pace of innovation has created both opportunities and challenges: while ML has accelerated SSE discovery timelines from decades to years, the proliferation of disparate approaches, limited data availability for non-Lithium systems, and lack of standardized evaluation metrics have hindered systematic progress. This review's forward-looking perspective on autonomous discovery platforms, physics-informed generative models, and integrated experimental-computational workflows will shape the field's trajectory toward predictive materials design. By providing strategic directions for addressing current limitations, from developing universal descriptors to establishing closed-loop discovery systems, this work positions the materials science community to realize the transformative potential of AI-driven SSE innovation, ultimately accelerating sustainable energy storage technology development.

View Article Online
DOI: 10.1039/D5MH01525A

Data Availability

No primary research results, software or code have been included, and no new data were generated or analysed as part of this review.

View Article Online
DOI: 10.1039/D5MH01525A

ARTICLE

Machine Learning Pipelines for the Design of Solid-State Electrolytes

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Vinamr Jain,[a] Zhilong Wang,[a,b] Fengqi You*[a,b,c,d]

The development of solid-state electrolytes (SSEs) is critical for enabling safer, high-energy-density batteries. However, the discovery of new inorganic SSEs is hindered by vast chemical search spaces, complex multi-property requirements, and limited experimental data, especially for multivalent systems. This review presents the first systematic framework mapping five interconnected challenges in SSE discovery to emerging AI solutions, providing a strategic roadmap for practitioners. We comprehensively survey machine learning pipelines from data resources and feature engineering to classical models, deep learning architectures, and cutting-edge generative approaches. Key breakthroughs include: (1) machine learning interatomic potentials enabling microsecond-scale molecular dynamics simulations at near-DFT accuracy, revealing non-Arrhenius transport behavior and overturning established transport mechanisms; (2) advanced neural network architectures achieving unprecedented accuracy in ionic conductivity prediction across diverse chemical spaces, including transformer-based and graph neural network approaches; (3) generative models successfully proposing and experimentally validating novel SSE compositions through diffusion-based design frameworks; and (4) autonomous closed-loop discovery platforms integrating ML predictions with experimental synthesis, achieving order-of-magnitude efficiency gains over traditional approaches. Unlike previous reviews focused on Li-ion systems, we explicitly address the critical data gap for multivalent conductors (Mg^{2+}, Ca^{2+}, Zn^{2+}, Al^{3+}) and provide concrete strategies through transfer learning and active learning frameworks. We bridge conventional computational methods (DFT, molecular dynamics) with modern ML techniques, demonstrating hybrid workflows that overcome individual limitations. The review concludes with actionable recommendations for multi-objective optimization, explainable AI implementation, and physics-informed model development, establishing a comprehensive roadmap for the next generation of AI-accelerated solid-state battery materials discovery.

1. Introduction

Renewable energy growth and electrified transportation are creating an urgent demand for efficient, safe energy storage.[1,2] Rechargeable lithium-ion batteries (LIBs) have dominated portable electronics and electric vehicles due to their high energy density and long cycle life[3]. However, conventional LIBs rely on liquid electrolytes that are flammable and volatile, raising serious safety concerns (fires and leakage) especially in large-scale applications.[4,5] These liquid electrolytes also have limited electrochemical stability windows, effectively capping the LIB energy density by constraining high-voltage cathodes and prohibiting the use of lithium metal anodes.[6–8] Dendritic lithium growth and side reactions in liquid electrolytes pose risks of short-circuit and cell failure, highlighting the need for alternative electrolyte technologies to enable safer, higher-energy batteries.[9]

All-solid-state electrolytes are being intensively explored as a next-generation solution to overcome the limitations of liquid electrolytes.[10–12] By replacing the flammable liquid with a non-combustible solid, SSE-based batteries promise vastly improved safety and thermal stability.[13] Moreover, the mechanical rigidity of inorganic SSEs can suppress dendrite propagation, potentially allowing the pairing of high-capacity lithium metal anodes with high-voltage cathodes for higher energy density cells.[14] SSE materials fall into two broad classes: inorganic crystalline or glassy ceramics (oxide or sulfide based) and solid polymers (or polymer–ceramic hybrids).[15,16] Inorganic SSEs such as oxide "garnet" $Li_7La_3Zr_2O_{12}$ and sulfide $Li_{10}GeP_2S_{12}$ have achieved room-temperature Li^+ conductivities on the order of 10^{-3}–10^{-2} S cm^{-1},[17–19] approaching those of liquid electrolytes. Polymer SSEs (e.g., PEO-based systems) offer flexibility and facile processing, but typically display lower ionic conductivities (~10^{-8}–10^{-6} S cm^{-1} at ambient temperature) and often require heating to 60–80°C to reach optimal conduction.[20–22] Each SSE family has its own challenges: ceramic electrolytes can suffer from grain-boundary resistance and brittle interfaces, whereas polymer electrolytes tend to have narrower electrochemical stability windows and lower transference numbers.[23,24] Ongoing research is addressing these issues (e.g., novel glassy

[a] College of Engineering, Cornell University, Ithaca, New York 14853, USA
[b] Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, USA
[c] Cornell University AI for Science Institute, Cornell University, Ithaca New York 14853, USA
[d] Cornell AI for Sustainability Initiative (CAISI), Cornell University, Ithaca New York 14853, USA.
* Corresponding author: Fengqi You (fengqi.you@cornell.edu)

View Article Online DOI: 10.1039/D5MH01525A

sulfide compositions and composite electrolytes) to realize the full safety and performance advantages of SSEs.[25,26]

Prior to the rise of ML, researchers relied on first-principles computations and atomistic methods have been widely used to predict phase stability and Li^+ chemical potentials, and to calculate migration barriers via nudged elastic band (NEB) pathways for candidate electrolytes.[27,28] These calculations yield valuable atomistic insights – for example, clarifying ion conduction mechanisms in fast-ion conductors and screening thermodynamically stable electrolyte/electrode simulations to guide SSE discovery and optimization.[29–31] DFT calculations combined with other computational approaches have proven valuable for materials discovery.[32,33] Molecular dynamics (MD) simulations (both classical and *ab initio* (AIMD)) are another important tool, enabling the computation of ionic diffusivities and conductivities in SSE frameworks.[34] Indeed, AIMD simulations on prototypical superionic solids like $Li_{10}GeP_2S_{12}$ and cubic $Li_7La_3Zr_2O_{12}$ have reproduced experimental ionic conductivities, confirming the capability of simulations to evaluate candidate SSE performance.[35] However, DFT and MD are computationally intensive and scale poorly to the enormous compositional space of solid materials.[36] High-throughput DFT screening is typically limited to evaluating hundreds of candidates at best, after preliminary filtering by simpler models.[37] This bottleneck has motivated the emergence of ML approaches in electrolyte research, which can learn complex composition–structure–property relationships from data and make rapid property predictions.[38] For instance, ML interatomic potentials trained on DFT data can act as surrogates to rapidly estimate ion migration barriers or perform MD simulations at a fraction of the cost.[39,40] More broadly, regression and classification models have been trained to predict SSE ionic conductivity or stability from compositions and structures, enabling fast screening of thousands of unexplored chemistries.[41,42] Early studies using data-driven models have already identified new Li-ion conductors that were missed by intuition or limited DFT searches,[43,44] underscoring the promise of ML in accelerating materials discovery.

Despite this progress, several key research gaps and challenges remain, which form the motivation for this review. A fundamental hurdle is the limited availability of comprehensive datasets, particularly for solid conductors beyond well-studied Li^+ systems, such as those for multivalent ions (Mg^{2+}, Ca^{2+}, Zn^{2+}, Al^{3+}).[45,46] This scarcity impedes the ability of supervised ML models to generalize effectively.[47,48] Relatedly, a significant concern is the limited transferability of models, as those trained on known compounds, may perform poorly when extrapolated to novel crystal structures or to different ion chemistries.[47] Furthermore, designing practical materials requires a holistic, multi-objective approach. While most studies have focused on optimizing a single property like ionic conductivity,[49–51] practical SSEs must simultaneously satisfy multiple criteria, including a wide electrochemical stability window and sufficient mechanical strength to suppress dendrite formation.

Another challenge is the "black-box" nature of many advanced ML models, which limits their utility when they cannot provide insights into the underlying factors governing material properties.[52,53] Finally, there is a pressing need to move beyond the passive screening of predefined candidate materials toward proactive, generative design. This requires employing generative algorithms to propose novel electrolyte compositions and structures[54–56] and developing closed-loop "predictive synthesis" pipelines, which iteratively couple ML predictions with DFT validation and experimental feedback to accelerate the discovery of new materials.[57,58] Addressing these five interconnected challenges – data limitations, multi-criteria optimization, interpretability, model generalization, and generative design – is crucial for unlocking the next wave of breakthroughs in solid-state electrolyte development.

This review addresses several critical gaps that distinguish it from existing literature on ML-driven SSE discovery. While previous reviews have largely focused on cataloguing ML techniques applied to battery materials broadly or examining specific electrolyte systems[59] within traditional experimental and computational frameworks, we provide the first systematic framework that maps specific challenges in SSE discovery to emerging AI solutions, offering a strategic roadmap for practitioners. Most existing reviews emphasize Li-ion systems exclusively, whereas we explicitly address the critical data scarcity for multivalent ion conductors and provide concrete strategies for extending ML approaches to these underexplored but technologically important systems. Importantly, we bridge the gap between traditional computational methods (DFT, MD, KMC) and modern ML techniques, demonstrating how hybrid workflows can overcome individual limitations while leveraging complementary strengths. Rather than merely surveying available techniques, we provide actionable guidance for data collection priorities, validation strategies, and implementation of explainable AI methods specifically tailored to solid-state electrolyte discovery. Finally, we emphasize emerging paradigms like autonomous discovery platforms and physics-informed machine learning that represent the next frontier in AI-accelerated materials discovery, going beyond conventional property prediction to enable true generative design of novel SSE materials.

We begin by examining the traditional computational methods that have historically guided SSE discovery, including NEB, molecular dynamics, and kinetic Monte Carlo simulations. We then detail the data resources and feature engineering strategies critical to enabling ML in this domain, followed by a survey of classical and deep learning models, including graph neural networks and ML-based interatomic potentials. We explore how these models have been applied to predict key properties (such as ionic conductivity, phase stability, and electrochemical compatibility), perform high-throughput screening to discover promising SSE candidates, and model ion diffusion mechanisms. Next, we address key challenges in ML-driven SSE discovery, including data scarcity, limited model transferability, and multi-objective optimization. We then discuss emerging solutions such as active and transfer learning, explainable AI, and physics-informed models. Finally, we highlight opportunities for autonomous discovery through generative design, ML interatomic potentials, and closed-loop pipelines integrating computation and experiment. Through

View Article Online
DOI: 10.1039/D5MH01525A

this synthesis, we aim to clarify the evolving role of machine learning in SSE development and highlight strategic directions for the field's continued advancement.

2. Conventional Computational Methods

Before the rise of ML, computational approaches including nudged elastic band (NEB) calculations, kinetic Monte Carlo (KMC) simulations, and molecular dynamics (MD) have been instrumental in SSE discovery. These methods provide the foundational data and physical insights that now enable ML-driven discovery. Understanding their capabilities and limitations is essential for designing effective hybrid computational workflows that combine traditional physics-based methods with modern ML techniques. A comparative summary of all computational methods discussed in this section is provided in Table S1 (Supplementary Information), highlighting their primary applications, advantages, limitations, and typical system sizes for SSE design.

2.1. Nudged Elastic Band (NEB) Method

NEB is an algorithm designed to find the Minimum Energy Path (MEP) and the associated saddle point (transition state) between known initial and final states on a potential energy surface. Its primary benefit is the direct calculation of the activation energy barrier (E_a) for specific atomic or ionic hops, providing crucial atomistic details of migration mechanisms. A crucial refinement, climbing image NEB (CI-NEB), addresses the challenge of accurately locating the true saddle point by driving one image uphill to converge precisely onto the saddle point.[60] This is vital for screening materials and dopants based on ion mobility. The E_a values derived from NEB calculations are also essential inputs for higher-scale simulations like Kinetic Monte Carlo.

The method has evolved from characterizing single materials to enabling high-throughput discovery. Early work mapped anisotropic Li-ion diffusion pathways in β-Li_3PS_4,[61] while automated path search methods have efficiently evaluated activation energies.[62] Automated high-throughput DFT workflows integrated with materials databases like the Materials Project, AFLOW, OQMD, and NIST-JARVIS have transformed materials discovery, allowing systematic exploration of thousands of potential SSE compositions with standardized protocols for convergence and property extraction.[63] Recent integration of NEB into high-throughput workflows enables screening of entire material classes like antiperovskites.[64] Modern implementations incorporate ML-guided path initialization using graph neural networks to generate superior initial guesses, dramatically improving convergence rates and reducing spurious local minima,[65] alongside adaptive sampling techniques with Gaussian process regression for efficient high-dimensional configuration space exploration. NEB can be combined with different levels of theory. DFT-NEB provides high accuracy but is computationally expensive, while classical NEB using empirical potentials offers computational efficiency at the cost of accuracy dependent on force field quality. Critical implementation challenges are discussed in detail in SI Section S1.1.

2.2. Kinetic Monte Carlo (KMC) simulations

KMC is a stochastic simulation technique modeling system evolution through discrete events with known rate constants. KMC excels at accessing experimentally relevant timescales (microseconds to seconds or longer), far exceeding typical MD simulations. This enables study of slow diffusion phenomena, SEI layer growth, or defect kinetics while efficiently bridging atomistic event rates to macroscopic properties like diffusion coefficients and ionic conductivity.

Recent methodological advances have significantly enhanced KMC capabilities for materials simulations. Adaptive kinetic Monte Carlo (aKMC) methods such as the kinetic Activation-Relaxation Technique (k-ART)[66] and Self-Evolving Atomistic Kinetic Monte Carlo (SEAKMC)[67] eliminate the need for pre-defined event catalogs by identifying transitions on-the-fly, enabling simulations of complex disordered systems. Accelerated techniques including the Mean Rate Method and First Passage Time Analysis have been developed to overcome kinetic trapping in superbasins,[68] extending the accessible timescales for materials with complex energy landscapes. Applications include active learning integration with KMC to explore SEI formation reaction barriers[69] and ab initio-based KMC investigating polyanion mixing effects on Na-ion transport in NASICON electrolytes.[70] Implementation considerations are discussed in SI Section S1.2.

2.3. Molecular dynamics (MD) simulations

2.3.1. Classical MD simulations

Classical MD simulates the atomic-scale motion of particles by numerically integrating Newton's equations of motion. MD allows for the simulation of significantly larger systems (10^3 — 10^6+ atoms) and longer timescales (nanoseconds to microseconds) compared to ab initio methods. It directly simulates ion dynamics at finite temperatures, enabling the calculation of transport properties (diffusion coefficients D, ionic conductivity σ, activation energies E_a), structural analysis via RDFs and coordination numbers, and prediction of mechanical properties.

The primary limitation is that accuracy hinges entirely on force field quality and transferability—the "force field bottleneck." Classical force fields do not explicitly treat electrons, precluding description of electronic phenomena like charge transfer or bond breaking/formation unless specialized reactive force fields are used. Applications include studying ion transport in polymer-argyrodite interfaces using newly developed OPLS-AA based force fields,[71] analyzing how Li vacancies or interstitials in β-Li_3PS_4 enhance conductivity by facilitating three-dimensional diffusion pathways,[72] and examining Li^+ transport in dilithium ethylene dicarbonate (Li_2EDC), a primary SEI component.[73] Software packages and implementation considerations are provide in SI sections S1.3-S1.5.

View Article Online
DOI: 10.1039/D5MH01525A

2.3.2. AIMD for ionic conductivity validation

AIMD combines molecular dynamics with quantum mechanical calculations (typically DFT) to determine interatomic forces on-the-fly at each simulation time step. This avoids empirical force field requirements, making AIMD particularly useful for novel or complex materials. It can implicitly account for electronic effects like dynamic polarization and charge distribution during ion motion, potentially offering higher accuracy than classical MD where these are prominent. AIMD serves as a crucial tool for benchmarking and parameterizing classical force fields or machine learning potentials.

However, AIMD is extremely computationally expensive. This restricts simulations to small system sizes (typically a few hundred atoms) and very short physical timescales (picoseconds to a few nanoseconds). Consequently, to observe sufficient diffusion events for calculating transport properties, AIMD simulations of SSEs are often run at very high temperatures, with room-temperature properties extrapolated via the Arrhenius relation, which can be unreliable if diffusion mechanisms change, or phase transitions occur. The accuracy of AIMD also remains dependent on the approximations within the underlying DFT calculation (e.g., the exchange-correlation functional). Applications include investigating lithium-ion diffusion in garnet-type materials[74] and studying chemical processes at the Li/Li_6PS_5Cl interface at different temperatures.[75] Sampling considerations are discussed in SI section S1.6.

3. Machine Learning Algorithms and Model Architectures for SSEs

In recent years, ML has emerged as a powerful paradigm to accelerate the design and discovery of novel SSEs. By learning complex relationships between material features and target properties, ML techniques can efficiently screen vast numbers of candidate materials, predict key performance metrics, and guide experimental synthesis efforts. This section reviews the key ML algorithms, model architectures, and the essential data resources that underpin the application of ML in the search for high-performance inorganic SSEs.

3.1. Data Resources for SSE Machine Learning

The efficacy and reliability of any ML model are inextricably linked to the quality, quantity, and relevance of the underlying data used for training and validation. In the context of SSE discovery, acquiring sufficient high-quality data presents a significant challenge, particularly for experimentally measured properties like ionic conductivity. This data scarcity can limit the predictive power and generalizability of ML models. SSE research leverages data from diverse sources, broadly categorized into large-scale computational databases and smaller, curated experimental datasets.

3.1.1. Computational Databases

These repositories primarily contain material properties derived from computational methods, most notably DFT and MD simulations. They serve as invaluable resources for high-throughput computational screening (HTS), allowing researchers to filter vast numbers of candidate materials based on predicted fundamental properties such as thermodynamic stability, electronic structure (e.g., band gap), crystal structure, and mechanical properties. While these databases contain diverse materials beyond SSEs, they serve as critical sources for identifying promising SSE candidates and training predictive models.

Materials Project (MP): The most prominent open-source database with DFT-calculated properties for hundreds of thousands of inorganic compounds.[63] MP provides formation energies, band gaps, elastic tensors, and crystal structures—all accessible via web interface and API. Its integration with pymatgen[76] and matminer[77] facilitates automated data retrieval and feature generation for ML workflows. MP is frequently used to identify Li-containing structures as initial SSE candidates.

Inorganic Crystal Structure Database (ICSD): Contains over 300,000 experimentally determined crystal structures,[78] providing reliable crystallographic information that serves as starting points for DFT calculations or structural descriptor generation.

AFLOW, OQMD, and NIST-JARVIS: These repositories offer additional DFT-calculated properties across millions of materials. AFLOW provides extensive electronic, thermodynamic, and mechanical properties via its REST API (AFLOWLIB).[79] OQMD focuses on thermodynamic stability through formation energies relative to the convex hull.[80] JARVIS offers comprehensive properties including elastic tensors, dielectric constants, and phonon properties for tens of thousands of materials.[81]

Other Computational Repositories: Additional databases contribute to the materials data ecosystem. The Computational Materials Repository (CMR) aggregates electronic structure data from various projects, including C2DB and QPOD.[82] Materials Cloud supports reproducible computational workflows and integrates with AiiDA for provenance tracking.[83] The Crystallography Open Database (COD) aggregates over 520,000 crystal structures of organic, inorganic, and metal-organic compounds.[84] GNoME, developed by DeepMind, has used deep learning to predict the stability of over 2 million inorganic crystals.[85] The Alexandria database provides DFT-calculated properties for millions of materials and is used to train large-scale ML models.[86]

3.1.2. Experimental and Curated Datasets

While computational databases offer breadth, datasets containing experimentally measured properties, particularly ionic conductivity, are essential for training models to predict real-world performance. These datasets are often smaller, compiled through painstaking literature surveys or expert curation.

- **LiIon Dataset:** An expert-curated collection focusing on lithium-ion conductors, containing 820 entries from 214 literature sources.[87] Each entry includes chemical composition, an assigned structural label (e.g., garnet,

LISICON), and AC impedance-measured ionic conductivity at specific temperatures. With 403 unique compositions having near-room-temperature conductivity data, it has been instrumental in training ML classifiers (like CrabNet) to distinguish between high and low conductivity compositions.[87]

- **OBELiX Dataset:** A more recent effort specifically designed for benchmarking ML models for SSE conductivity prediction. It comprises approximately 600 synthesized solid electrolyte materials with experimentally measured room-temperature ionic conductivity, along with composition, space group, lattice parameters, and, for about half the entries, full crystallographic information files (CIFs).[88]
- **Literature-Mined Datasets:** Several studies have employed natural language processing (NLP) and text mining techniques to automatically extract relevant data (e.g., ionic conductivity values, synthesis parameters, structural types) directly from the vast body of scientific literature. While powerful for data aggregation, these approaches face challenges related to the heterogeneity of reported data, inconsistencies in experimental conditions, and the accuracy of automated extraction.[89] An example includes the work by Shon & Min (2023), which extracted over 4000 conductivity measurements from nearly 1500 papers.[90]

3.1.3. Data Challenges

The effective application of ML in SSE discovery is often hampered by several data-related challenges. As mentioned, experimental data, especially reliable room-temperature ionic conductivity measurements, remains relatively scarce compared to the vastness of the chemical space being explored. Data heterogeneity is another issue, arising from differences between computational predictions and experimental realities, variations in experimental protocols and measurement conditions across different studies, and the diverse formats used for data reporting. Furthermore, both computational and experimental data contain inherent uncertainties and potential errors as DFT calculations rely on approximations, while experimental measurements are subject to synthesis variations and characterization limitations.[91] These issues often result in datasets with missing values and significant class imbalance, where high-performing electrolytes are severely underrepresented. To mitigate these challenges, researchers employ various strategies, including data imputation to estimate missing entries and resampling techniques such as the Synthetic Minority Over-sampling TEchnique (SMOTE) to create more balanced training sets.[92] Finally, data accessibility varies, with some key databases requiring subscriptions while others are open access.

The landscape of data resources reveals a complementary relationship between large-scale computational databases and smaller, targeted experimental datasets. Computational databases like MP, AFLOW, OQMD, and JARVIS provide the necessary breadth for initial high-throughput screening, enabling the filtering of millions of hypothetical compounds based on fundamental properties like thermodynamic stability (formation energy, energy above hull), electronic insulation (band gap), and potentially relevant structural or mechanical characteristics. However, accurately predicting ionic conductivity, the key performance metric for an SSE, directly from first principles is computationally demanding, often requiring expensive MD simulations. This is where curated experimental datasets like LiIon and OBELiX become critical. Although smaller in size, they contain the direct experimental measurements needed to train and validate ML models specifically designed to predict ionic conductivity. This often leads to a multi-stage ML workflow: initial screening using models trained on large computational datasets to identify stable and electronically suitable candidates, followed by conductivity prediction for the down-selected candidates using models trained on experimental data. Table S2 provides a summary of prominent datasets commonly used in machine learning studies for solid-state electrolyte research, including their primary data sources, key material properties covered, accessibility, and relevant references. The development of accurate and efficient machine learning interatomic potentials (MLIPs, discussed in Section 3.4) represents a significant effort to bridge this gap, aiming to enable faster calculation of dynamic properties like ionic conductivity for the vast number of candidates identified through computational screening.

3.2. Classical Machine Learning Algorithms and Descriptors

Before the widespread adoption of deep learning, classical machine learning algorithms formed the backbone of data-driven materials discovery efforts, including the search for novel SSEs. These methods remain valuable tools for establishing baseline models, interpreting feature importance, and tackling problems with limited data. They typically operate on a set of pre-defined features, known as descriptors, which numerically encode relevant material characteristics.

3.2.1. Descriptors (Features): The Language of Materials for ML

Descriptors translate the chemical and physical nature of a material into a numerical format that ML algorithms can process. The selection, generation, and quality of these descriptors are paramount, directly influencing model accuracy, interpretability, and generalizability. A significant challenge in the field is the development of descriptors that are both universally applicable across different material classes and accurately capture the underlying physics governing the target property. Descriptors used in SSE research can be grouped into several categories:

- **Compositional Descriptors:** These features are derived solely from the material's chemical formula (stoichiometry) and the intrinsic properties of its constituent elements. Examples include average atomic mass, mean electronegativity, variance of atomic radii, elemental fractions, and specific stoichiometric ratios. They are computationally inexpensive to generate but ignore the crucial

View Article Online
DOI: 10.1039/D5MH01525A

influence of atomic arrangement and bonding. For instance, one study utilized a set of 145 "Chemical Descriptor" features based on stoichiometry and elemental properties[93]. While simple, compositional descriptors alone can sometimes yield reasonable predictive performance, particularly for classification tasks or when combined with more sophisticated algorithms.

- **Structural Descriptors:** These capture information about the geometric arrangement of atoms in the crystal lattice. They can range from simple parameters like lattice constants, cell volume, space group number, and packing fraction to more complex representations like radial distribution functions (RDFs), coordination numbers, bond angles, polyhedral volumes, local atomic environment motifs (e.g., using Voronoi analysis), and topological indices. Structural descriptors are vital as many key SSE properties, including ionic conductivity pathways and mechanical stability, are intimately linked to the crystal structure. Generating these features typically requires crystallographic information (e.g., from CIF files obtained via ICSD or MP) and specialized analysis tools. Examples include employing Voronoi tessellation features to improve graph neural networks,[94] or using Smooth Overlap of Atomic Positions (SOAP) descriptors to represent local atomic environments.[95]
- **Electronic Descriptors:** These features quantify aspects of the material's electronic structure, which governs electrical conductivity, electrochemical stability, and chemical bonding. Common examples include the electronic band gap (Eg), position of valence and conduction band edges, density of states near the Fermi level, work function, electron affinity, ionization potential, and measures of bond ionicity or covalency. Electronic descriptors are crucial for screening potential SSEs, as ideal candidates must be good ionic conductors but poor electronic conductors (i.e., possess a wide band gap) and exhibit stability within the battery's operating voltage window. These descriptors are often derived from computationally intensive DFT calculations.
- **Physicochemical/Thermodynamic Descriptors:** This broad category includes various calculated or tabulated physical and chemical properties. Examples relevant to SSEs include formation energy, energy above the convex hull (Ehull) for thermodynamic stability assessment, density, ionic radii, melting point, and mechanical properties like bulk modulus (K) and shear modulus (G). These descriptors relate to a material's stability, processability, and mechanical robustness against issues like dendrite penetration. Formation energy and Ehull are standard outputs from DFT databases (MP, OQMD) used for initial stability screening, while mechanical moduli, predicted using ML or DFT, are critical for assessing dendrite suppression capabilities.
- **Kinetic/Dynamic Descriptors:** These features aim to capture aspects related to ion transport dynamics. Examples include activation energy barriers for ion migration (Eb or Ea), diffusion coefficients (D), attempt frequencies, and properties derived from phonon calculations (e.g., vibrational density of states, phonon band structure features). These descriptors are most directly related to ionic conductivity (σ), often following an Arrhenius-type relationship $\sigma \propto exp\left(\frac{E_a}{k_b T}\right)$. However, they are typically challenging and computationally expensive to obtain, requiring methods like NEB calculations for migration barriers or extensive MD simulations for diffusion coefficients. Recent work has shown phonon-related features derived from DFT phonon calculations can be important predictors for ionic conductivity in ML models.[96]

The different categories of descriptors, along with their generation methods and significance, are summarized in Table 1.

Libraries and Tools for Featurization

The automated generation of descriptors, or "featurization," is facilitated by an ecosystem of open-source Python libraries. Pymatgen[76] provides the core data structures and tools for materials analysis. Built upon this, Matminer[77] offers a high-level interface for computing a comprehensive suite of compositional, structural, and electronic descriptors from standard material representations. For more advanced models, libraries such as DeepChem[97] are valuable for generating the graph-based representations required by architectures like Graph Neural Networks. These toolkits are instrumental for automating the creation of robust and reproducible feature sets for machine learning.

Table 1: Common Descriptors Used in Machine Learning for Solid-State Electrolytes

Descriptor Category	Specific Descriptor Example	Information Encoded	Generation Method	Pros/Cons
Compositional	Average Electronegativity	Elemental chemical bonding tendency	Formula-based	Simple; Ignores structure
	Elemental Fractions	Stoichiometry	Formula-based	Simple; Basic composition info
Structural	Volume per Atom	Packing density, free volume	Structure Analysis (CIF)	Relates to ion mobility/stiffness; Requires structure

	Space Group Number	Crystal symmetry	Structure Analysis (CIF)	Captures overall symmetry; Coarse descriptor
	Radial Distribution Function (RDF)	Average local atomic density around a central atom	Structure Analysis (CIF)	Detailed local structure; Computationally more intensive
	Coordination Number	Number of nearest neighbours	Structure Analysis (CIF)	Local bonding environment: Definition can vary
Electronic structure	Band Gap (Eg)	Energy required to excite an electron	DFT	Key for electronic conductivity; Computationally expensive
	Formation Energy	Thermodynamic stability relative to elemental phases	DFT	Fundamental stability metric; Requires calculation
	Energy Above Hull (Ehull)	Thermodynamic stability relative to competing phases	DFT	Better stability indicator than formation energy; Requires phase diagram data
Physicochemical	Ionic Radii	Effective size of ions	Tabulated/Formula	Relates to packing and channel size; Simple approximation
	Shear/Bulk Modulus (G,K)	Resistance to shear/volume deformation	DFT / ML Prediction	Key for mechanical stability (dendrites); Requires calculation/prediction
Kinetic/Dynamic	Migration Barrier (Ea, Eb)	Energy barrier for ion hopping	DFT (NEB) / MD	Directly relates to conductivity; Computationally very expensive
	Phonon Properties	Lattice vibrational characteristics	DFT (Phonon Calc.)	Relates to ion dynamics/stability; Computationally expensive

Note: CIF = Crystallographic Information File; DFT = Density Functional Theory; MD = Molecular Dynamics; NEB = Nudged Elastic Band; ML = Machine Learning.

3.2.2. Classical ML Algorithms in SSE Research

Various classical ML algorithms have been applied to SSE research for tasks including property prediction, classification, and unsupervised exploration of materials space.

- **Regression:** Used to predict continuous target variables.
  - *Algorithms:* Simple Linear Regression, Polynomial Regression, Kernel Ridge Regression (KRR), Support Vector Regression (SVR), Gaussian Process Regression (GPR).
  - *Applications:* Predicting ionic conductivity (logσ), activation energies, elastic moduli (K,G) for mechanical stability assessment, and formation energies. For example, Ahmad et al. used Gradient Boosting Regressor (GBR) and KRR, trained on structural features, to predict shear and bulk moduli for over 12,000 inorganic solids in a screening study for dendrite suppression[98]. Zhao et al. used GPR-based Bayesian optimization to guide the experimental synthesis of LATP electrolytes towards optimal ionic conductivity[99].
- **Classification:** Used to assign materials to discrete categories.
  - *Algorithms:* Logistic Regression (LR), Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees (DT).
  - *Applications:* Xu et al. (2020) used Logistic Regression to classify SICON compounds as poor or good superionic conductors based on elemental descriptors.[47] Chen et al. (2021) employed Support Vector Machines to analyze relationships between manufacturing conditions and solid-state electrolyte film performance for evaluation and optimization.[100] Adhyatma et al. (2022) applied a tree-based LightGBM model to classify doped LLZO compounds by their ionic conductivity levels (high or low).[101]
- **Ensemble Methods:** These techniques combine predictions from multiple individual models (base learners) to improve overall performance, robustness, and reduce overfitting. They often achieve state-of-the-art results on tabular data.
  - *Algorithms:* Random Forest (RF), Gradient Boosting Machines (GBM, including variants like XGBoost and LightGBM).
  - *Applications:* RF and GB variants are frequently employed for both regression (predicting conductivity, formation energy) and classification (high/low conductivity, stability) in SSE research. For instance, Pereznieto et al. (2023) utilized a Random Forest algorithm to analyze experimental data and discover new potential Na-ion solid electrolytes exhibiting high ionic conductivity.[102] Kim et al.

View Article Online DOI: 10.1039/D5MH01525A

(2023) implemented an ensemble model of gradient boosting algorithms to classify over 3,500 NASICON structures, successfully identifying promising Na superionic conductor candidates with high accuracy.[103] Tang et al. (2024) applied an XGBoost algorithm to predict key properties such as band structure and stability, which enabled the screening and identification of 194 ideal solid-state electrolyte candidates from over 6,000 structures.[104] Zhang et al. (2024) developed Random Forest models alongside neural networks to predict ionic conductivity in NASICON materials and to identify influential factors, highlighting the role of Na stoichiometric count.[105]

- **Clustering:** Unsupervised learning algorithms group similar data points together without relying on predefined labels.
  - *Algorithms:* k-Means, Agglomerative clustering, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)
  - *Applications:* Park et al. (2024) used HDBSCAN to cluster over 12,000 Na-containing materials based on structural properties, identifying 12 groups and revealing shared characteristics in high-conductivity clusters.[106] Laskowski et al. (2023) applied agglomerative clustering to ~26,000 Li-containing structures to identify promising superionic conductor candidates for further screening.[95] Gallo-Bueno et al. (2022) used unsupervised outlier detection models to automatically classify computed Li-argyrodite crystal structures based on their structural distortion.[107]

The successful application of classical ML algorithms is heavily dependent on the process of "feature engineering" – the careful selection, transformation, and combination of descriptors to best represent the underlying material physics relevant to the target property. The frequent high performance reported for ensemble methods like Random Forest and Gradient Boosting variants (XGBoost, LightGBM)[108–111] underscores the difficulty in capturing the complex, often non-linear, structure-property relationships in SSEs using single, simpler models acting on these hand-crafted features. Ensemble methods offer robustness by averaging out errors from individual base learners (like decision trees) and implicitly handling feature interactions, making them well-suited to the high-dimensional and potentially noisy descriptor spaces common in materials informatics. However, their complexity can sometimes make direct physical interpretation of the learned relationships challenging compared to simpler models like linear regression.

Despite these interpretability challenges, classical ensemble methods remain preferable in scenarios with limited training data where deep learning models would overfit, or when transparent decision-making is critical for materials design insights. For instance, Decision tree models can readily identify feature importance rankings,[106] while XGBoost provides built-in interpretability tools that can reveal which structural descriptors most strongly influence ionic conductivity predictions.[112–114] These advantages make classical approaches particularly valuable in early-stage SSE discovery when datasets are small or when researchers need to understand and communicate the physical basis underlying model predictions to experimental collaborators. Unsupervised clustering techniques, such as HDBSCAN, provide a valuable alternative or complementary approach.[106] By grouping materials based on similarities in their descriptor vectors (often structural features derived from large computational databases), clustering can reveal inherent patterns and identify promising material families even when labeled target data (like experimental conductivity) is sparse. This capability allows researchers to leverage the vastness of computational datasets to guide exploration before focusing on more data-intensive supervised prediction tasks. This reliance on feature engineering and the success of complex ensembles sets the stage for deep learning approaches (Section 3.3), which aim to automate the feature learning process itself.

3.3. Neural Network Architectures and Deep Learning Models

While classical ML methods have proven valuable, their reliance on hand-crafted descriptors limits their ability to capture complex, non-linear interactions and spatial correlations within crystal structures that govern SSE properties. Deep learning (DL), characterized by artificial neural networks with multiple layers, enables hierarchical feature learning directly from raw data, reducing the need for manual feature engineering.

The simplest deep learning architecture, Feedforward Neural Networks (FNNs) or Multi-Layer Perceptrons (MLPs), consists of an input layer, one or more hidden layers, and an output layer, processing information in one direction. They operate on pre-defined descriptors similar to classical algorithms (Figure 2a) and have been used as components within ensemble models, baseline comparisons, or for property prediction based on manually selected features in SSE research.[88,105,115]

Graph Neural Networks (GNNs) represent a more sophisticated approach, naturally operating on graph representations of materials where atoms are nodes and bonds or interatomic proximity define edges. This allows GNNs to learn representations that explicitly incorporate atomic connectivity and local chemical environments, automatically identifying features relevant to predicting material properties. Capturing crystal structure nuances, such as periodicity and 3D geometry (SE(3) invariance/equivariance), is crucial for effective GNN design. Crystal Graph Convolutional Neural Network (CGCNN) represents crystals as graphs and uses convolutional layers to aggregate information from neighboring atoms and bonds to learn atom-level features, which are then pooled to predict material properties (Figure 2b). It has been applied to predict thermodynamic stability and mechanical properties of SSEs.[116,117] Improved versions like iCGCNN incorporate Voronoi tessellation information and explicit many-body interactions to enhance performance.[118] Materials Graph Network (MEGNet) extends the graph network concept by including global state variables (like temperature or pressure) alongside atomic

(node), bond (edge), and global features, allowing for more versatile property predictions (Figure 2b). MEGNet and related architectures like M3GNet[119] have been trained on large datasets (e.g., Materials Project) for broad applicability in materials property prediction and can be applied to predict SSE stability or mechanical properties.[120] SchNet employs continuous-filter convolutional layers to model quantum interactions in atomistic systems without using explicit graph representations, and has been used to predict formation energies of bulk crystals and potential energy surfaces.[121] The field continues evolving rapidly, with newer architectures like ALIGNN (Atomistic Line Graph Neural Network),[122] k-NAGCN (k-Nearest Atom Graph Neural Network),[123] and transformer-based models like CrystalFramer (which introduces dynamic, attention-based coordinate frames)[124] continuously advancing accuracy and representational power for crystal structures.

Distinct from structure-based approaches, some deep learning models prioritize elemental composition, offering advantages when structural information is unavailable, computationally expensive to obtain, or for rapid initial screening across vast chemical spaces. ElemNet learns material properties directly from elemental compositions represented as fractional counts, bypassing structural information for rapid composition-based screening.[125] CrabNet, a transformer-based model using attention mechanisms, operates primarily on compositional data but implicitly learns interactions between elements[126] (Figure 2c). It demonstrated success when trained on the LiIon dataset for classifying compositions by their likelihood of exhibiting high lithium-ion conductivity.[87] More broadly, transformer architectures—inspired by their success in natural language processing and relying heavily on self-attention mechanisms—can capture long-range interactions within crystal graphs or learn complex relationships between constituent elements, as seen in CrabNet[126] and CrystalFramer.[124] Transformer architectures are also being used to develop powerful interatomic potentials like GPTFF.[127]

While most ML models predict properties of given materials (forward problem), generative models solve the inverse problem: generating novel material structures likely to possess desired properties. Techniques like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models are being explored for materials discovery.[55,128] These models learn the underlying distribution of known stable materials and can sample this distribution or be conditioned to generate new candidates meeting specific criteria (e.g., high stability, target band gap, specific crystal structure). MatterGen, a diffusion model operating on 3D crystal geometry, has demonstrated the ability to generate novel, stable materials with target properties by learning from large databases like MP and Alexandria.[56] Such approaches hold significant promise for generating entirely new SSE candidates beyond modifications of known structures. Other generative approaches like SHAFT utilize hierarchical generation based on symmetry constraints.[129]

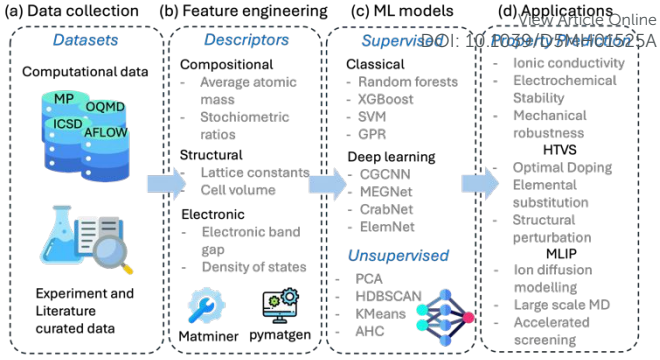

Figure 1: Overview of a machine learning pipeline for the design and discovery of SSEs. (a) The pipeline begins with data resources such as the Materials Project, ICSD, and JARVIS, which provide structural and property data for a wide range of inorganic materials. (b) These data are transformed into meaningful descriptors: composition-based, structural, and electronic, using tools such as Matminer and pymatgen. (c) Machine learning models, organized by learning paradigm (supervised, unsupervised, deep learning), are then trained on these descriptors. Classical models (e.g., random forests, SVMs) and deep learning architectures (e.g., CGCNN, MEGNet, CrabNet) are (d) employed to predict key properties such as ionic conductivity, electrochemical stability, and mechanical robustness. These models also enable applications including ML-based interatomic potentials and high-throughput virtual screening for novel multivalent SSEs.

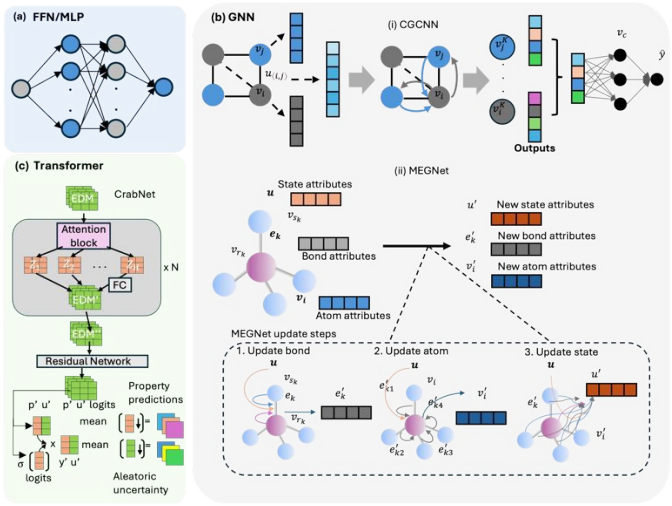

Figure 2: Schematic overview of representative deep learning architectures for SSE property prediction. (a) FFN or MLP, which maps a fixed-length vector of engineered features to a target property. (b) GNN architectures that operate on graph representations of crystal structures. (i) The CGCNN updates atom features (v_i) by aggregating information from its local atomic neighborhood. (ii) The MEGNet framework, which iteratively updates atom (v_i), bond (e_k), and global state (u) attributes to learn a comprehensive representation of the material. (c) The CrabNet architecture, a transformer-based model that uses a self-attention mechanism on elemental composition to predict properties and quantify aleatoric uncertainty.

3.4. Machine Learning Interatomic Potentials (MLIPs) for Dynamics (MLMD)

A major breakthrough enabled by deep learning is the development of highly accurate Machine Learning Interatomic Potentials (MLIPs), also known as ML force fields. These models learn the complex relationship between atomic configurations and the potential energy surface (PES) – including energies, forces on atoms, and stresses on the simulation cell – directly from large datasets generated by high-fidelity quantum mechanical calculations (typically DFT). Once trained, MLIPs can perform MD simulations, termed MLMD, with an accuracy

View Article Online DOI: 10.1039/D5MH01525A

approaching that of DFT but at a computational cost orders of magnitude lower (closer to classical force fields).

This capability is particularly transformative for SSE research. Simulating ion transport dynamics – the diffusion pathways, diffusion coefficients (D), activation energies (Ea), and ultimately ionic conductivity (σ) – requires tracking atomic motion over long timescales (nanoseconds or more) and large system sizes (thousands of atoms) to capture statistically relevant events and collective motion. Ion transport in SSEs involves rare events such as defect formation, migration, and collective rearrangements that occur over vastly different timescales: while individual atomic hops happen on picosecond timescales, macroscopic diffusion processes and phase transformations relevant to battery operation occur over seconds to minutes. Such simulations are often computationally prohibitive using traditional AIMD. MLIPs overcome this limitation, enabling routine MLMD simulations that provide direct insights into the mechanisms governing ionic conductivity in complex SSE materials.

Several MLIP frameworks have been applied to study SSEs:

- **Gaussian Process Regression and Sparse GPR (SGPR) approaches:** Traditional GPR methods provide a Bayesian framework for learning interatomic potentials with built-in uncertainty quantification, but their $O(n^3)$ computational scaling with dataset size becomes prohibitive for large training sets. SGPR addresses this limitation through low-rank approximations using reduced "inducing sets" of representative local environments, achieving computational scaling comparable to linear methods while retaining the probabilistic advantages of GPR.[130] SGPR has been successfully applied to survey Li diffusivity across hundreds of ternary crystals and create transferable universal potentials for complex electrolytes like $Li_{10}GeP_2S_{12}$.[131,132]
- **Gaussian Approximation Potential (GAP):** Based on Gaussian process regression. A near-universal GAP was developed for the Li-P-S (LPS) material class, enabling studies of conductivity in both crystalline (e.g., Li_3PS_4, $Li_7P_3S_{11}$) and glassy phases and revealing the importance of anion dynamics.[133]
- **Deep Potential Molecular Dynamics (DeePMD / DeePMD-kit):** A deep neural network-based potential that has seen wide application.[134] It has been used to model Li diffusion in amorphous Li_3PO_4,[135] superionic conductors like $Li_{10}GeP_2S_{12}$ (LGPS) and Nb-doped garnets, and importantly, to perform microsecond-long simulations revealing the lack of a significant "paddle-wheel" effect from polyanion rotations on Li diffusion in crystalline $Li_7P_3S_{11}$ and $Li_2B_{12}H_{12}$ at room temperature.[136]
- **Crystal Hamiltonian Graph Network (CHGNet):** A GNN-based universal MLIP pre-trained on the extensive Materials Project trajectory dataset, uniquely incorporating electronic charge and magnetic moment information.[106] It has been demonstrated for charge-informed MD simulations of Li intercalation (Li_xMnO_2) and Li diffusion in garnet SSEs.[137]
- **M3GNet (Materials 3-Body Graph Network):** Another GNN-based universal potential trained on the Materials Project database, designed for broad applicability in structural relaxation and dynamics simulations.[119]
- **GPTFF (Graph-based Pre-trained Transformer Force Field):** A recent transformer-based force field trained on a massive dataset (billions of force components), aiming for high accuracy and generalizability across diverse inorganic systems.[127]

MLMD simulations driven by these potentials have provided crucial insights, such as identifying non-Arrhenius diffusion behavior in LGPS,[135] elucidating specific diffusion pathways,[137] and quantifying the impact of structural features like defects or anion dynamics on conductivity.[133] The significant speed-up factors highlight the potential of MLIPs to dramatically accelerate the computational assessment of ionic transport.[138]

The progression from classical ML to deep learning marks a significant evolution in the computational toolkit for SSE discovery. GNNs, in particular, represent a paradigm shift away from manual feature engineering towards automated learning of structure-property relationships directly from the atomic graph representation. This allows models to potentially uncover more complex and subtle correlations than might be captured by human-designed descriptors. However, these advances come with important practical considerations. GNN architectures like CGCNN and MEGNet require high-quality crystal structure files (CIFs) with precise atomic positions as inputs, as they construct graph representations directly from atomic arrangements and bonding information[117,120]. The incorporation of both atomic and bond-level descriptors introduces numerous hyperparameters, necessitating larger training datasets (typically $>10^3$ samples) and substantial computational resources compared to classical ML approaches that rely on pre-computed scalar descriptors[139]. In contrast, SGPR-based approaches can achieve comparable accuracy with smaller training datasets due to their efficient use of training data and adaptive sampling strategies, making them particularly suitable for data-scarce regimes where generating extensive DFT training sets is computationally expensive.[130,140]

Perhaps even more impactful is the development and application of MLIPs. While classical ML and standard GNNs often focus on predicting static properties (stability, band gap, moduli) or rely on computationally expensive methods (AIMD, NEB) to infer dynamics, MLIPs provide a computationally tractable route to directly simulate the crucial dynamic processes governing ionic conductivity. This enables the field to move beyond predicting prerequisites for good conductivity towards simulating and understanding the transport phenomenon itself over timescales reaching microseconds—a significant computational achievement.[141] However, MLIPs require careful validation to ensure transferability across different thermodynamic conditions and structural motifs, as their accuracy is fundamentally limited by the quality and coverage of the underlying DFT training set. Additionally, the

View Article Online
DOI: 10.1039/D5MH01525A

computational overhead of generating sufficient training data for MLIPs can be substantial, particularly for complex multi-component systems. Despite these advances, current MLMD simulations still remain far from capturing the experimentally relevant timescales (seconds to minutes) over which macroscopic ionic transport and device-relevant processes occur, and bridging to true experimental scales may require hybrid approaches combining MLMD with adaptive KMC methods.

Models trained predominantly on computational data face inherent challenges when predicting experimentally observed ionic conductivities due to systematic discrepancies between DFT calculations and experimental measurements. Effective validation strategies require testing against independent experimental datasets rather than computational holdouts, implementing cross-validation with available experimental data, and developing calibration methods that account for temperature-dependent Arrhenius behavior and experimental measurement uncertainties.[142] For SGPR-based approaches, the inherent uncertainty quantification provides additional validation capabilities by identifying regions where model predictions may be unreliable, enabling more robust assessment of model confidence and guiding iterative improvement through active learning protocols.[140] Furthermore, hybrid training approaches that incorporate both computational and experimental data during model development can significantly improve predictive accuracy for experimental properties. As computational materials discovery matures, adopting rigorous experimental validation protocols will be critical for establishing ML models as reliable tools for guiding experimental synthesis efforts. Generative models represent a further step, shifting the focus from predicting properties of existing or hypothetical materials to designing entirely new structures optimized for target performance.

Furthermore, the emergence of large-scale, pre-trained models signifies a trend towards developing more universal and transferable tools in materials informatics. Models like MEGNet, M3GNet, CHGNet, and GPTFF, trained on vast and diverse datasets such as the Materials Project calculation database, encapsulate a broad understanding of chemical bonding and structural stability across the periodic table. This pre-training allows these foundational models to be potentially fine-tuned for specific downstream tasks, such as predicting properties within a particular class of SSEs, using smaller, task-specific datasets. This strategy leverages the massive amounts of existing computational data to build general knowledge, which can then accelerate research on specific material systems by reducing the burden of generating extensive training data for every new problem. Nevertheless, practitioners should be aware that even pre-trained universal models may require domain-specific fine-tuning and validation, particularly when applied to novel chemistries or extreme conditions not well-represented in the original training data. The success of these approaches ultimately depends on careful consideration of data quality, model selection criteria, and rigorous benchmarking against experimental observations. This approach promises to significantly enhance the efficiency of ML-driven materials discovery pipelines.

4. ML-Guided Applications in SSE Discovery and Design

4.1. Prediction of Key Material Properties

A primary application of ML in SSE research is the rapid and accurate prediction of crucial material properties. By learning from existing data, ML models can establish correlations between easily obtainable features (e.g., composition, crystal structure) and target properties that are typically expensive or slow to determine. Ideal SSEs should possess a suite of desirable characteristics, including high ionic conductivity (often targeting >1 mS cm^{-1} at room temperature), a wide electrochemical window to ensure stability against high-voltage cathodes and low-voltage anodes (like Li metal), and sufficient mechanical strength to suppress lithium dendrite penetration.

4.1.1. Ionic Conductivity

Ionic conductivity is arguably the most critical performance metric for an SSE. ML models have been developed to predict this property, often by correlating structural and chemical descriptors with experimentally measured or computationally derived conductivity values. These models can significantly expedite the identification of promising high-conductivity candidates from large databases.

The foundational work by Sendek et al. (2017) established the viability of ML-driven conductivity screening through a logistic regression classifier trained on 40 lithium-containing compounds.[143] Despite the limited training set, their model effectively distinguished fast from slow Li-ion conductors using atomistic descriptors including Li-Li coordination numbers, sublattice bond ionicity, and anion coordination environments. The practical validation of this approach emerged when high-throughput screening of 12,000 Materials Project compounds identified 21 fast-conductor candidates, with subsequent DFT-MD simulations confirming superionic behavior in several materials, notably Li_3InCl_6, which achieved experimental verification.[143,144] This early success demonstrated that even simple ML models, when coupled with physically meaningful features, could effectively navigate vast chemical spaces.

Building on these classification successes, recent efforts have focused on regression-based conductivity prediction with enhanced accuracy. The comparative analysis by Mishra et al. (2023) systematically evaluated eight predictor models including Random Forest Regressor, Support Vector Machine, and shallow neural networks using activation energy, operating temperature, and lattice parameters as features.[110] Their findings highlighted the superior robustness of ensemble methods like Random Forest, while demonstrating that model stacking prevents overfitting, a critical insight for conductivity prediction where data scarcity remains a persistent challenge.

The transition toward more sophisticated approaches is exemplified by studies targeting specific electrolyte chemistries

View Article Online DOI: 10.1039/D5MH01535A

with optimized algorithms and novel descriptors. Jaafreh et al. (2024) developed a targeted framework for Mg-ion electrolytes by leveraging phonon density of states (PhDOS) data to calculate "total phonon band center" as a conductivity proxy.[145] Their systematic comparison of Extra Random Trees, Gradient Boosting, and Extreme Gradient Boosting algorithms revealed that Extra Random Trees achieved superior performance (R^2 = 0.964), enabling predictions across ~9,000 Mg compounds. The chemical insights derived from this model, particularly the identification of Mg-Se systems as exhibiting the lowest median band centers (27.5 meV) compared to Mg-S (40.5 meV) and Mg-O (55.5 meV), demonstrate how ML can simultaneously accelerate screening and provide mechanistic understanding.[145]

Addressing the critical data gap for multivalent systems, Dong et al. developed a generalizable ML framework specifically designed for screening Na, Mg, and Al garnet electrolytes.[146] Utilizing carefully designed chemical descriptors, their XGBoost models achieved 94% accuracy for thermal stability and 89% for band gap prediction across 43,732 compounds. The framework identified 1,764 compounds meeting both thermal stability and electronic criteria, which were further filtered to yield 44 economically viable candidates with high performance potential. Interpretability analysis revealed that mean electronegativity is the most critical factor for thermal stability, while atomic radius range governs band gap properties, providing actionable design principles for multivalent conductor development.

Kharbouch et al. (2024) achieved exceptional accuracy for ionic conductivity prediction (R^2 = 0.85) for LLZO-type garnets through meticulous data curation and hyperparameter optimization using CatBoostRegressor with Optuna framework tuning.[147] Their emphasis on rigorous preprocessing, including stoichiometric verification and KNN imputation, underscores the critical importance of data quality in achieving reliable conductivity predictions.

Recent developments have integrated pre-trained graph neural network potentials to generate physics-informed descriptors. Maevskiy et al. (2025) employed M3GNet to analyze potential energy surfaces under frozen framework approximation, deriving heuristic descriptors correlated with lithium mobility.[148] This approach achieved efficiency gains of approximately 50× faster than MLIP-driven MD and >3,000× faster than AIMD, with eight out of ten highest-ranked materials confirmed as superionic conductors through first-principles calculations.[148] The significance of this work lies in its demonstration of how powerful, pre-trained "foundation" models can be adapted to generate specialized, physically meaningful features for predicting properties like ionic conductivity, enabling rapid and reliable large-scale screening.

Models trained predominantly on computational data face inherent challenges when predicting experimentally observed ionic conductivities due to systematic discrepancies between DFT calculations and experimental measurements. Effective validation strategies require testing against independent experimental datasets rather than computational holdouts, implementing cross-validation with available experimental data, and developing calibration methods that account for temperature-dependent Arrhenius behavior and experimental measurement uncertainties.[142] Furthermore, hybrid training approaches that incorporate both computational and experimental data during model development can significantly improve predictive accuracy for experimental properties.[149] As computational materials discovery matures, adopting rigorous experimental validation protocols will be critical for establishing ML models as reliable tools for guiding experimental synthesis efforts.

4.1.2. Electrochemical Stability

Electrochemical stability is vital for the practical application of SSEs, ensuring they do not decompose when in contact with highly reactive electrodes (e.g., Li metal anode) or at the operating voltages of the battery. ML models contribute by predicting properties indicative of stability, such as formation energy (a proxy for thermodynamic stability against decomposition into competing phases) and band gap (often correlated with the electrochemical window).

The critical importance of accurate structural sampling for stability predictions is demonstrated by Ataya et al., who revealed that conventional Coulomb methods fail to identify the most stable, low-energy LLTO configurations after DFT geometry relaxation.[150] This structural misrepresentation led to overestimated electrochemical stability windows (3.1 V versus the correct 2.5 V), with prediction errors reaching 0.67 eV. To address this sampling challenge, the authors developed a SOAP-KRR machine learning model trained on only 40 DFT-relaxed structures that accurately predicts energy rankings, providing a computationally efficient alternative for sampling disordered materials.[150]

Complementing these structural considerations, comprehensive screening approaches have emerged that integrate stability assessments within broader materials discovery pipelines. Chen et al. (2025) developed a hierarchical screening strategy starting with 20,717 Li-containing compounds from the Materials Project database.[51] Their multi-stage process applied thermodynamic stability and electronic band gap pre-screening, followed by ML classification and regression models trained on 468 samples to identify high-conductivity candidates. After electrochemical stability window assessment and AIMD validation, this approach identified three promising candidates (Li_3BiS_3, Li_5BiS_4, and $Li_{10}ZnP_4S_{16}$) with high room-temperature ionic conductivities, low activation energies, and favorable interfacial compatibility with common cathodes.[51]

The relationship between composition, structure, and electrochemical performance has been further elucidated through targeted studies of specific electrolyte families. Kireeva et al. investigated garnet-structured solid electrolytes by combining experimental data analysis with machine learning, identifying an optimal lattice constant range of 12.950-12.965 Å for maximum ionic conductivity in LLZO-type garnets.[151] Their quantitative regression models using SVM, LSTM, GP, and XGBoost algorithms revealed that Li and La content, atomic scattering factors at the C site, and Shannon ionic radii of dopants were the most influential parameters affecting ionic

View Article Online
DOI: 10.1039/D5MH01525A

conductivity, providing quantitative guidance for compositional optimization.[151]

4.1.3. Mechanical Stability

The mechanical properties of SSEs are critical, particularly for their ability to suppress the growth of lithium dendrites, which can cause short circuits and battery failure, especially when using Li metal anodes. ML models have been developed to predict mechanical properties such as bulk modulus (K) and shear modulus (G), which are key inputs for theories of dendrite suppression.

Early applications of graph neural networks for mechanical property prediction established the feasibility of high-throughput screening approaches. Ahmad et al. employed a CGCNN trained on 2041 crystal structures with DFT-calculated elastic moduli to predict mechanical properties for over 12,000 inorganic solids.[98] These ML-predicted moduli were then integrated with the Monroe-Newman stability parameter (χ) framework to assess dendrite initiation propensity at Li metal/SSE interfaces, identifying over 20 mechanically anisotropic interfaces involving six solid electrolytes predicted to suppress dendrite growth.[98]

The challenge of limited training data has been systematically addressed through active learning strategies that optimize data acquisition. Choi et al. trained a LightGBM model on 14,238 elasticity structures, initially achieving modest performance (R^2 = 0.633 for shear modulus prediction).[152] However, their active learning approach, which iteratively added materials with high prediction uncertainty to the training set, improved the R^2 score to 0.802 with only 1,600 strategic additions compared to 2,800 required for random selection.[152] This efficiency gain highlights the critical importance of intelligent data acquisition strategies, particularly given the computational expense of DFT elasticity calculations.

Building on these methodological advances, comprehensive screening workflows have emerged that integrate mechanical property prediction with other critical SSE characteristics. Sun et al. developed a two-stage ML workflow starting with LGBM-based mechanical property screening of 5,329 LLZO-derived candidates, followed by superionic conductor classification and AIMD validation.[50] This hierarchical approach successfully identified 10 new tetragonal-phase materials combining superior mechanical properties with high ionic conductivity[50].
The interpretability of mechanical property predictions has been enhanced through feature analysis techniques that provide physical insight into structure-property relationships. Wang et al. developed an optimized LGBM model achieving R^2 ≈ 0.86-0.87 for both shear and bulk modulus prediction using 8,920 Materials Project samples.[153] Their integration of SHAP analysis revealed that volume per atom and valence band maximum are critical predictors, while extrapolation experiments to datasets containing elements (Mg, Al, K, Ni) absent from training demonstrated that model transferability to new chemical spaces can be significantly improved with strategic addition of diverse samples.[153]

4.2. High-Throughput Virtual Screening (HTVS)

HTVS leverages computational power to rapidly evaluate vast numbers of candidate materials for desired properties, significantly accelerating the materials discovery cycle. ML plays a crucial role in making HTVS more efficient and intelligent by acting as fast and inexpensive filters, prioritizing the most promising materials for further, more accurate investigation rather than relying solely on brute-force first-principles calculations. The integration of ML transforms HTVS from a potentially exhaustive search into a more guided exploration, employing classifiers to identify materials belonging to desired classes (e.g., "superionic conductor"), regression models to predict continuous property values, and active learning approaches that iteratively suggest the most informative candidates to evaluate next. Figure 3 shows a schematic of a typical ML-driven HVTS workflow.

The scale and sophistication of modern HTVS campaigns are exemplified by ultra-large screening efforts that combine multiple ML models in hierarchical filtering approaches. Chen et al. (2024) demonstrated this approach by screening over 32 million candidates for solid-state electrolytes.[154] Structure candidates generated via iso-valent substitutions were reduced to ~589,000 stable materials using ML potentials (M3GNet) for thermodynamic phase stability assessment. Subsequent funnel-based screening applied ML models for band gap (> 3 eV) and electrochemical stability filters, followed by higher-accuracy DFT calculations, yielding 18 final candidates with new compositions. The top candidates, the $Na_xLi_{3-x}YCl_6$ series, were synthesized and experimentally validated, confirming both structure and conductivity predictions.[154]

Complementing these massive screening approaches, targeted studies of specific material families have employed sophisticated multi-property optimization strategies. Lee et al. (2025) computationally screened 4,375 hypothetical Na-based argyrodites using DFT calculations to evaluate energy above hull, formation energy, band gap, and electrochemical stability window[155]. Their 4-dimensional Pareto sorting technique narrowed the field to 15 top candidates, with AIMD simulations ultimately identifying five promising virtual compositions, including $Na_6SiS_4Cl_2$ and $Na_{7.75}SiS_{5.75}Cl_{0.25}$.[155] This approach demonstrates how multi-objective optimization can efficiently navigate complex property trade-offs in materials design. Similarly employing multi-dimensional optimization, Lee et al. (2024) combined genetic algorithms with Bayesian optimization using GPR surrogate models to screen 18,133 hypothetical antiperovskite electrolytes. Their active learning framework reduced the computational burden to just 144 strategically selected DFT calculations while constructing a 4-dimensional Pareto frontier for thermodynamic stability, band gap, electrochemical window, and ionic conductivity, ultimately identifying 22 promising candidates with seven exhibiting superior room-temperature conductivity (>4 mS cm^{-1}).[156]

The integration of experimental insights with computational screening has enabled more targeted materials design strategies. Sewak et al. trained a logistic regression model on 170 experimental NASICON materials, using PCA to identify 9 key features governing ionic conductivity.[157] The model revealed that low dopant electronegativity and increased Li

View Article Online
DOI: 10.1039/D5MH01525A

occupancy at M2 sites are critical for high conductivity, insights that guided dopant selection for the $LiGe_2(PO_4)_3$ system. Bond valence sum energy calculations further screened dopants by migration barrier estimation, leading to the design of $Li_2Mg_{0.5}Ge_{1.5}(PO_4)_3$ with a DFT-validated migration barrier of 0.261 eV.[157]

Advanced ML architectures have been developed specifically for ionic conductivity screening, leveraging physics-informed descriptors to enhance prediction accuracy. Xie et al. performed high-throughput screening of nearly 50,000 Li-containing compounds using bond-valence kinetic Monte Carlo simulations, identifying 329 materials meeting stability and conductivity thresholds.[158] Their graph convolutional network, trained to predict conductivity directly from bond valence energy landscapes, outperformed models learning from atomic structure alone and accelerated screening of 979 additional candidates generated via isovalent substitution, identifying 239 potential superionic conductors.[158]

Specialized neural network architectures have also emerged for targeted chemical space exploration. Wan et al. (2024) developed DopNetFC, which outperformed conventional ML approaches including Random Forest and GBDT for screening atom substitution schemes.[159] Applied to over 2,208 potential substitutions in $Li_{10}GeP_2S_{12}$, the most promising ML-identified candidates were validated through multi-step DFT calculations assessing thermodynamic, electronic, and mechanical stability.[159] This approach demonstrates the effectiveness of task-specific neural architectures for exploring well-defined chemical modification spaces.

Multivalent conductor screening has been advanced through comprehensive ML platforms addressing critical data gaps beyond Li-ion systems. Wang et al. developed AI-IMAE based on CGCNN, a platform providing real-time activation energy predictions across nine ionic species (Li^+, Na^+, Mg^{2+}, Zn^{2+}, Al^{3+}, Ag^+, Cu^{2+}, F^-, O^{2-}) with ~10^5x speedup over traditional methods.[160] Screening 144,595 compounds identified 316 SSE candidates and 129 cathode materials across the different ionic species. Similarly, Cai et al. used XGBoost algorithms to screen spinel structures for Mg/Zn cathodes, achieving 91.2% prediction accuracy and identifying six candidates ($MgNi_2O_4$, $MgMo_2S_4$, $MgCu_2S_4$, $ZnCa_2S_4$, $ZnCu_2O_4$, $ZnNi_2O_4$) with ionic diffusion coefficients >1×10^{-9} cm^2s^{-1} and volume expansions <22%.[161] These targeted approaches demonstrate ML's potential for accelerating discovery in underexplored multivalent systems.

4.3. Elucidating Ion Dynamics via ML Interatomic Potentials (MLIPs)

Understanding the atomistic mechanisms of ion diffusion is fundamental to designing SSEs with high ionic conductivity. Traditional methods like AIMD provide high accuracy but are computationally expensive, limiting simulations to small system sizes (hundreds of atoms) and short timescales (picoseconds to nanoseconds). Classical empirical potentials are much faster but often lack the accuracy and transferability needed for complex SSE chemistries or reactive environments. MLIPs have emerged as a transformative technology, bridging this accuracy-cost gap. Trained on extensive datasets of energies and forces generated by DFT calculations, MLIPs can reproduce the potential energy surface with near-DFT accuracy but at a computational cost order of magnitude lower, enabling large-scale (thousands to millions of atoms) and long-timescale (nanoseconds to microseconds) MD simulations.

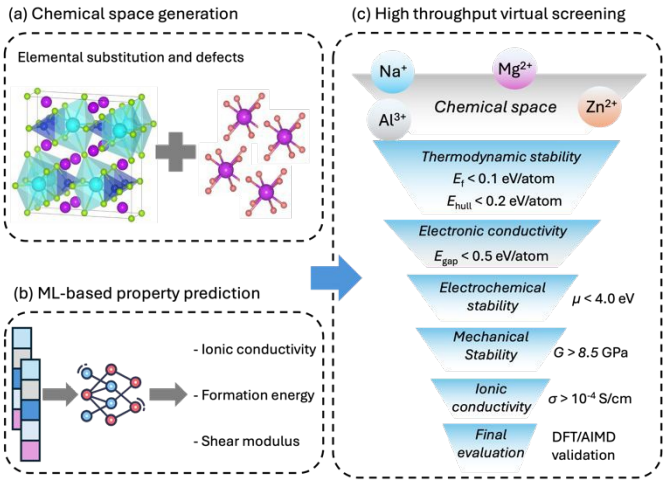

Figure 3: Schematic illustration of a HTVS workflow for the discovery and evaluation of SSEs. (a) The chemical space is generated via systematic elemental substitutions and defect engineering within known crystal structure prototypes. (b) ML models trained on precomputed datasets are employed to rapidly predict key properties such as ionic conductivity, formation energy, and shear modulus. (c) Candidate materials are filtered through a sequential funnel based on physical criteria including thermodynamic, electronic, electrochemical, and mechanical stability, followed by ionic conductivity thresholds. The most promising candidates undergo final validation using first-principles calculations (DFT and/or AIMD).

This capability has profound implications. MLIPs allow for the simulation of complex SSE systems, such as amorphous phases, grain boundaries, and interfaces, which are often intractable with AIMD due to their size and disorder. Furthermore, the extended simulation times accessible with MLIPs are crucial for capturing rare diffusion events, accurately calculating diffusion coefficients, and observing collective ionic motion, leading to unprecedented insights into ion transport pathways and the role of structural dynamics. Beyond these mechanistic studies, MLIPs also enable the high-throughput computational screening of vast design spaces to accelerate the discovery of entirely new SSE materials (Figure 4).

The theoretical foundation for this field was established by Behler and Parrinello (2007), who introduced high-dimensional neural network potentials using symmetry functions to describe local chemical environments in a rotationally and translationally invariant manner.[162] This pioneering approach laid the groundwork for modern MLIPs that enable DFT-accuracy simulations at significantly reduced computational cost.

Applications of MLIPs in SSE research have progressed from validating known properties to discovering new transport phenomena and challenging established mechanisms. Gigli et al. (2024) exemplified this evolution by investigating charge transport in all known phases (α, β, and γ) of Li_3PS_4 using three separate potentials trained on different DFT reference levels (PBEsol, r²SCAN, and PBE0).[163] Their large-scale (768-atom) and

View Article Online
DOI: 10.1039/D5MH01525A

long-timescale (up to 6 ns) simulations revealed that superionic behavior results from a structural transition from γ to mixed α-β phases, driven by thermal activation of correlated PS_4 flips that reduce Li-ion diffusion activation energy by up to 6-fold.[163] Crucially, they refuted the "paddle-wheel" mechanism by demonstrating that PS_4 flip timescales (nanoseconds) and Li-ion hopping (picoseconds) are separated by orders of magnitude, while also showing that the commonly used Nernst-Einstein approximation underestimates conductivity by more than a factor of two.[163]

The power of MLIPs in elucidating complex transport behaviors extends to understanding non-Arrhenius temperature dependence in garnet systems. Dai et al. (2022) studied $Li_xLa_3Zr_{x-5}Ta_{7-x}O_{12}$ garnets using MLIPs trained on DFT-MD trajectories, achieving superior accuracy compared to other computational models.[164] Their simulations revealed that ionic conductivity follows Vogel-Tammann-Fulcher rather than Arrhenius behavior, with maximum conductivity occurring at Li content between 6.6 and 6.8.[164] This work demonstrates how MLIPs can capture subtle temperature-dependent transport phenomena that require extensive sampling.

MLIPs have proven particularly valuable for studying amorphous systems and interfaces, where structural disorder demands large simulation cells and long equilibration times. Seth et al. (2025) investigated Li^+ transport in amorphous LiPON and at Li||LiPON interfaces using a neural equivariant interatomic potential (NequIP) trained on over 13,000 DFT structures.[165] Their simulations accurately reproduced experimental room-temperature conductivity in bulk LiPON while revealing that interfacial transport is one order of magnitude slower than bulk transport.[165] Similarly, Yang et al. (2025) combined AIMD with DeePMD MLIPs to study amorphous $Li_xAlO_yCl_{3+x-2y}$ electrolytes, revealing that Li^+ transport is facilitated by Cl atom rotation within tetrahedral frameworks and that oxygen doping enhances glass-forming ability while reducing mobile Cl atoms, requiring optimization of the O/Cl ratio for maximum conductivity.[166]

The integration of MLIPs with materials discovery workflows has enabled the exploration of composition-structure-property relationships across extended chemical spaces. Guo et al. (2022) demonstrated this approach by mapping the phase diagram of glass-ceramic lithium thiophosphate electrolytes using neural network potentials coupled with genetic algorithms to explore amorphous structures along the $(Li_2S)_x(P_2S_5)_{1-x}$ composition line.[167] Through unsupervised structure-similarity analysis, they identified that local Li environments resembling superionic β-Li_3PS_4 are energetically favorable around $x \approx 0.725$, leading to the design of a new candidate composition with predicted ionic conductivity exceeding 10^{-2} S cm^{-1}.[167]

Beyond solid-state electrolytes, MLIPs have also provided valuable insights into ionic transport mechanisms in battery electrode materials. Ha et al. (2022) demonstrated the application of SGPR-accelerated molecular dynamics to investigate the effect of aluminium doping on Li-ion transport in Li-excess layered oxide cathodes.[168] Their nanosecond-timescale simulations of $Li_{1.22}Ru_{0.61}Ni_{0.11}Al_{0.06}O_2$ revealed that Al-doping reduces the Li-ion diffusion activation energy from 0.48 eV to 0.40 eV, demonstrating enhanced ionic transport alongside improved structural stability. This reduction in activation energy resulted in approximately twice the Li-ion diffusion coefficient at elevated temperatures. The study showed how strategic dopant selection can simultaneously optimize both transport properties and electrochemical stability, with strengthened Al-O bonding suppressing oxygen oxidation while facilitating Li-ion mobility.

Despite their transformative potential, MLIP-based MD simulations require careful validation to ensure reliable predictions, particularly given inherent uncertainties in force predictions and energy errors.[169] Best-practice validation strategies extend beyond simple energy and force comparisons to include systematic benchmarking against AIMD for key properties such as diffusion coefficients, phase stability, and thermal transport.[170] Uncertainty quantification through ensemble methods, gradient-based approaches, or committee models provides essential error estimates during simulations, enabling active learning protocols that iteratively improve MLIP reliability.[171,172] Furthermore, domain-specific validation tests, including rare event prediction and long-timescale dynamical properties, are crucial for establishing confidence in MLIP extrapolation beyond training domains.[173] As the field matures, standardized validation protocols and uncertainty reporting will be essential for establishing MLIP credibility in high-stakes materials discovery applications.

Table 2 summarizes these seminal contributions, illustrating how MLIPs have advanced our understanding of ion dynamics in SSEs.

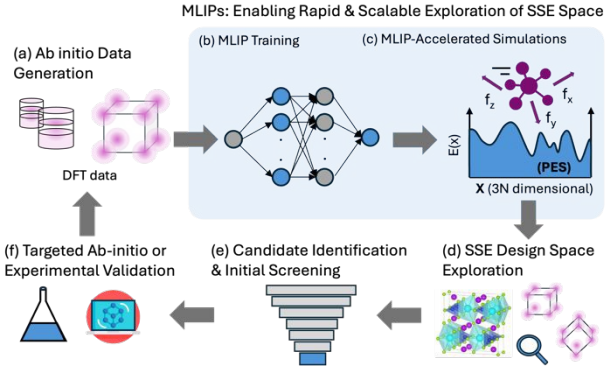

Figure 4: A schematic of the machine learning interatomic potential (MLIP) driven workflow for accelerated discovery of solid-state electrolytes (SSEs). (a) The process begins with generating a dataset of energies and forces from ab-initio calculations (e.g., DFT). (b) This data is used to train a machine learning model, such as a neural network, to create an MLIP. (c) The trained MLIP rapidly predicts the potential energy surface (PES), enabling large-scale and long-timescale molecular dynamics simulations. These simulations allow for (d) the systematic exploration of the vast SSE design space, which is constructed by varying elemental compositions, introducing dopants, and considering diverse crystalline and amorphous structures. (e) From these simulations, promising candidates are identified through a screening funnel. (f) The most promising materials are then validated with targeted, high-fidelity DFT calculations or experimental synthesis. This framework can operate as a closed loop, where new data from the validation step is used to further refine the MLIP.

View Article Online
DOI: 10.1039/D5MH01525A

5. Navigating the Frontiers of Solid-State Electrolyte Discovery: Addressing Key Challenges

Despite the considerable enthusiasm and initial successes, the application of ML in SSE research is confronted by several deeply ingrained challenges that currently limit its full potential. These research gaps, which form the central motivation for this review, include pervasive data scarcity, particularly for emerging material systems; the complex demands of multi-objective optimization for practical applications; the often-opaque nature of ML models, which hinders scientific understanding and trust; issues with the transferability and generalization of models to new chemical domains; and the need to move beyond simple screening towards generative design frameworks capable of proposing entirely novel materials. These challenges are not merely isolated obstacles but are often interconnected, where, for instance, a lack of sufficient high-quality data directly impedes the development of generalizable models capable of robust multi-objective optimization. Addressing these interconnected hurdles is paramount for ML to truly catalyze a paradigm shift in materials discovery, transitioning from serendipitous discovery to a more predictive, efficient, and accelerated design cycle for SSEs and, by extension, other advanced functional materials. Figure 5 provides a schematic overview of the key machine learning methodologies that have emerged to address these core challenges.

5.1. Challenge 1: Navigating Data Deficiencies in ML-Driven SSE Discovery

The fundamental challenge limiting ML-driven SSE discovery is the pervasive scarcity of high-quality training data, particularly for multivalent ion conductors. This data deficit manifests in three critical dimensions: insufficient quantity, poor quality heterogeneity, and severe chemical imbalance across ion types.

The Non-Lithium Data Crisis

While Li^+ systems benefit from decades of intensive research generating relatively substantial datasets, non-lithium ion conductors including Na^+ and multivalent systems (Mg^{2+}, Ca^{2+}, Zn^{2+}, Al^{3+}) remain critically underrepresented.[174–176] This disparity is not merely quantitative. Non-lithium ions exhibit fundamentally different transport mechanisms characterized by varying ionic radii, coordination preferences, and in the case of multivalent systems, stronger Coulombic lattice interactions and sluggish diffusion kinetics.[177] Consequently, ML models trained on Li^+ data cannot reliably extrapolate to these alternate systems, as evidenced by uMLIPs failing to generalize beyond their chemical training space.[178] The fundamental differences in transport mechanisms, optimization priorities, and critical descriptors across Li, Na, Mg, and Al systems (summarized in Table S3) necessitate system-specific ML framework design.

The data quality problem compounds this scarcity. SSE datasets aggregate information from disparate experimental protocols, computational methods with varying theoretical rigor, and literature reports lacking standardized metrics.[179] This heterogeneity introduces systematic noise, missing values, and conflicting measurements that undermine model reliability. The absence of centralized, standardized databases for multivalent SSE properties forces fragmented, redundant curation efforts across research groups,[87] impeding collaborative progress.

Solution 1: Leveraging Existing Scarce Data Through Advanced Learning Paradigms

Unsupervised Learning for Pattern Discovery

When labeled data is scarce, unsupervised learning methods like clustering, dimensionality reduction, and representation learning can extract meaningful structural patterns from abundant unlabeled datasets. This approach is particularly useful for hypothesizing which features might transfer from data-rich systems (e.g., Li^+, Na^+) to data-scarce ones (e.g., multivalents). For example, Park et al. successfully applied clustering to over 12,000 Na-containing materials, revealing that high-conductivity candidates consistently shared specific structural characteristics, such as the abundance of certain polyhedral motifs (XO_4 tetrahedra) and the presence of spacious ion channels.[106] This finding suggests a path for methodological transfer to beyond-lithium systems. While the optimal structural features for a Mg^{2+} conductor will differ from those for Na^+, the types of descriptors identified as critical such as coordination environments, polyhedral packing, and framework connectivity, are likely to be fundamentally important across different ion systems. An effective strategy, therefore, involves using unsupervised learning on large Li^+ or Na^+ datasets to identify these critical feature classes, which can then guide the engineering of more targeted descriptors for the subsequent supervised modeling of multivalent systems.

Transfer Learning for Cross-Domain Knowledge

Transfer learning offers a strategic pathway to leverage knowledge from data-rich domains (e.g., Li^+ systems, general materials databases) for data-scarce targets (multivalent conductors). A compelling demonstration showed successful cross-domain ionic conductivity classification, where models trained exclusively on Na^+-based NASICON compounds accurately predicted Li^+-based materials.[47] However, the chemical similarity between Na^+ and Li^+ likely enabled this success. Extending transfer learning to multivalent systems with fundamentally different coordination preferences and transport mechanisms may require sophisticated domain adaptation techniques or physics-informed constraints to bridge the mechanistic gap.

Semi-Supervised Learning for Hybrid Data Exploitation

Semi-supervised learning provides a middle ground between fully supervised and unsupervised approaches by leveraging both labeled and unlabeled data simultaneously. This paradigm is particularly valuable for SSE discovery where experimental conductivity measurements are sparse but structural databases are abundant. The methodology typically involves clustering a

View Article Online
DOI: 10.1039/D5MH01525A

large, unlabeled dataset based on descriptor similarity and then labeling the resulting clusters with the few available experimental data points to identify promising regions of the materials space. This strategy was exemplified by Laskowski et al., who applied unsupervised agglomerative clustering to approximately 26,000 lithium-containing compounds and subsequently annotated the resulting clusters using a limited set of experimental conductivity measurements.[95]

Table 2: Seminal Contributions of ML Interatomic Potentials to Understanding Ion Dynamics in SSEs

Study/MLIP Development (Primary Citation)	MLIP Type/Focus	SSE System(s) Investigated	Key Insights into Ion Dynamics/Mechanisms	Significance/Impact
Behler and Parrinello (2007)[162]	HDNNPs using atom-centered symmetry functions	Bulk Silicon (as proof-of-concept for general condensed matter systems).	Decomposes total energy into local atomic contributions, enabling simulations of arbitrarily sized systems with DFT accuracy by learning the potential energy surface (PES)	Foundational theoretical and methodological work that established the modern framework for atomistic MLIPs, making large-scale, long-timescale simulations of SSEs feasible
Guo et al. (2022)[167]	ANN potential combined with a Genetic Algorithm (GA) for AI-aided sampling	Glass-ceramic Lithium Thiophosphate (LPS) systems: $(Li_2S)_x(P_2S_5)_{1-x}$	Discovered that local Li environments similar to the superionic β-Li_3PS_4 phase are energetically favored around composition x ≈ 0.725. Mapped the amorphous phase diagram and identified miscibility gaps	Demonstrated a powerful workflow combining MLIP-accelerated sampling and structural analysis to design novel, high-conductivity amorphous SSE compositions
Gigli et al. (2024)[163]	GAPs trained on multiple DFT levels (PBEsol, r²SCAN, and PBE0)	All known polymorphs (α, β, γ) of Lithium Thiophosphate (Li_3PS_4)	Showed superionic behavior is driven by a structural transition activated by correlated PS_4 flips, not a "paddle-wheel" effect. The Nernst-Einstein approximation underestimates conductivity by over a factor of 2 due to strong ionic correlations	Resolved a long-standing controversy over the transport mechanism in Li_3PS_4 and highlighted the necessity of using higher-accuracy functionals (PBE0) and correlation-aware analysis for predictive simulations.
Dai et al. (2022)[164]	Artificial Neural Network (SIMPLE-NN) using atom-centered symmetry functions.	Lithium garnet oxides: $Li_xLa_3Zr_{x-5}Ta_{7-x}O_{12}$	Revealed that ionic conduction in garnets follows a non-Arrhenius temperature dependence, better described by the VTF equation. Calculated Haven ratio of 0.1–0.4 indicates strong concerted motion of Li-ions	Provided a highly accurate potential for the garnet family, resolving ambiguity around the optimal composition for conductivity (x = 6.6 to 6.8) by combining simulations with experimental data
Seth et al. (2024)[165]	NequIP, an E(3)-equivariant GNN	Amorphous Lithium Phosphorus Oxynitride (LiPON) and Li	LiPON interface	Accurately modelled the amorphous LiPON structure and bulk Li^+ conductivity. Found that Li^+ transport across the Li
Yang et al. (2025)[166]	DeePMD	Amorphous oxychloride electrolytes: $Li_xAlO_yCl_{3+x-2y}$	Uncovered that Li^+ transport is facilitated by the rotation of Cl atoms within a structural skeleton of Al-chains. Found that O-doping enhances amorphization (enabling Cl rotation) but reduces mobile Cl atoms, creating an optimal O/Cl ratio for conductivity	Elucidated a novel transport mechanism in an emerging class of amorphous oxychloride SSEs and provided a clear design principle based on balancing glass-forming ability with mobile anion concentration
Ha et al. (2022)[168]	SGPR with on-the-fly training	Al-doped Li-excess layered oxide cathodes: $Li_{1.22}Ru_{0.61}Ni_{0.11}Al_{0.06}O_2$	Demonstrated that Al-doping reduces Li-ion diffusion activation energy from 0.48 eV to 0.40 eV, enhancing ionic transport while strengthened Al-O bonding suppresses oxygen oxidation and improves structural stability	Demonstrated how dopant-induced electronic structure modifications can simultaneously enhance ionic transport and suppress degradation mechanisms, providing design principles for stable high-energy-density electrode materials with improved Li-ion mobility

This methodology successfully identified a cluster exhibiting high probability for superionic conduction, which led to the experimental confirmation of Li_3BS_3 as a novel ionic conductor. The success of this approach provides a template for a targeted discovery pipeline in underexplored chemical spaces, such as those for multivalent conductors. Such a workflow would involve first clustering the vast space of hypothetical multivalent host structures using reliable structural descriptors. Following this, a small and diverse set of compounds from different clusters could be strategically synthesized to serve as initial "seed" labels. Subsequent experimental efforts could then be prioritized on the unlabeled materials within or adjacent to clusters containing the most promising initial results, thereby maximizing the value of each experiment and accelerating the identification of novel beyond-lithium SSEs.

Solution 2: Targeted Data Generation Through Computational High-Throughput Screening

High-throughput density functional theory (HTP-DFT) calculations provide a systematic approach to generate large, internally consistent datasets for intrinsic material properties.[180] This computational pipeline can systematically evaluate thousands of candidate materials, creating valuable training data while maintaining theoretical consistency. Furthermore, ML models can be trained to predict expensive DFT results, enabling large-scale screening by circumventing first-principles calculations for every candidate.[98]

Successful liquid electrolyte platforms like the Electrolyte Genome[181] demonstrate the value of systematic property correlation mapping and automated screening workflows beyond simple high-throughput calculation. These liquid-phase

systems also offer opportunities for cross-domain learning: ion transport patterns in liquid and polymer electrolytes including solvation dynamics, coordination environment effects, and structure-transport correlations can inform descriptor engineering and mechanistic understanding for solid electrolytes, particularly for data-scarce multivalent systems where liquid-phase computational studies are more prevalent. Adapting these methodologies to solid-state systems could establish not only standardized data specifications but also automated multi-property optimization pipelines that integrate atomic-scale MLIP predictions with mesoscale grain boundary and interface modelling.

The synergy between HTP-DFT and ML creates a self-reinforcing cycle: computational data trains ML models, which subsequently accelerate screening by reducing computational bottlenecks.

Solution 3: Active Learning for Intelligent Data Acquisition

Active learning addresses the resource constraints of both experimental synthesis and computational simulations by strategically selecting the most informative data points for generation.[182] In this iterative framework, ML models identify candidates where they exhibit maximum uncertainty or where new data would optimally improve performance. These selections are then prioritized for experimental characterization or DFT calculation.

This approach has demonstrated practical success in optimizing doping strategies for LLZO electrolytes.[57] By combining ML models with uncertainty quantification, the active learning framework efficiently navigated the vast compositional space, identifying promising dopant combinations while minimizing required simulations and experiments.[57]

However, the effectiveness of these data-centric approaches depends critically on establishing clear prioritization criteria for data collection efforts. Future experimental and computational campaigns should prioritize: (1) multivalent systems with intermediate ionic radii (Mg^{2+}, Zn^{2+}) that bridge the gap between monovalent and highly charged species, (2) materials exhibiting mixed ionic-electronic conductivity where transport mechanisms remain poorly understood, and (3) interfacial properties and degradation pathways that are systematically underrepresented in current databases. Computationally, emphasis should be placed on generating temperature-dependent transport data and correlated ionic motion descriptors, as these are essential for capturing the non-Arrhenius behavior observed in many superionic conductors yet remain scarce in existing datasets. The choice among these strategies or, more likely, a combination thereof will depend critically on the specific SSE system under investigation, the target property, and the nature of the available data. For instance, while transfer learning might be effective for predicting properties of Na-ion conductors based on Li-ion data due to their chemical similarities, discovering novel multivalent conductors might necessitate more extensive de novo data generation via HTP-DFT, guided by active learning, to capture their unique physics. A universal solution to data scarcity is improbable; instead, a versatile toolkit of these data-centric approaches is essential for continued progress. Table 3 summarizes the key data challenges encountered in the application of ML to SSE discovery and outlines potential mitigation strategies.

Table 3: Summary of Data Challenges in ML for SSEs and Mitigation Strategies.

Data Challenge	Impact on ML Model Development	Key Mitigation Strategies & Supporting Evidence
Overall Scarcity for SSEs	Poor generalization, difficulty modelling complex phenomena, bias towards well-studied systems.	HTP-DFT Data Generation[180], Development of Curated Databases[87], Active Learning[57], Semi-supervised learning[95]
Specific Scarcity for Multivalent Ion Conductors	Inability to model distinct physics (e.g., stronger Coulombic interactions, sluggish diffusion) accurately, poor extrapolation from Li-ion systems.	Targeted HTP-DFT for Multivalents, Transfer Learning[47], Physics-Informed ML[183,184], Unsupervised Learning for feature discovery[106,185]
Data Heterogeneity/Quality (Multi-source, noise, missing values)	Reduced model reliability, inconsistent predictions, difficulty in training robust models.	Rigorous Data Curation & Preprocessing[179], Standardized Data Reporting Protocols, Robust ML algorithms tolerant to noise.
Small Sample Sizes for Truly Novel Chemistries	High risk of overfitting, poor predictive power for unexplored chemical spaces.	Generative Models for candidate proposal[186], Transfer Learning from broader chemical domains[187], LOGO-CV for realistic performance assessment[188]

5.2. Challenge 2: Multi-Objective Optimization: Balancing Performance Metrics in SSE Design

Commercially viable SSEs require concurrent optimization of multiple, often conflicting properties rather than maximizing a single parameter. Practical SSEs must satisfy stringent requirements including:

- **High Ionic Conductivity (σ):** Typically targeted to be ≥ 10^{-4} S cm−1 at room temperature, approaching or exceeding that of liquid electrolytes, to enable high power densities.
- **Wide Electrochemical Stability Window (ESW):** The electrolyte must remain stable against both highly reducing (anode) and highly oxidizing (cathode) potentials, ideally > 5.5 V vs. Li/Li+ for high-voltage applications.

- **Good Electrode Compatibility:** Minimal chemical and electrochemical reactivity with both anode (especially Li metal) and cathode materials to prevent detrimental interfacial layer growth and impedance rise.
- **Sufficient Mechanical Strength and Appropriate Moduli:** The SSE should possess adequate mechanical robustness to suppress lithium dendrite penetration and withstand the stresses induced by electrode volume changes during cycling, while also maintaining good interfacial contact.
- **High Li^+ Transference Number (tLi+):** Ideally close to unity, indicating that Li^+ ions are the primary charge carriers, which minimizes concentration polarization and improves rate capability.
- **Other Considerations:** Factors such as ease of processing, scalability, low cost, and environmental impact also play crucial roles in practical viability.

These requirements, however, must be contextualized within the distinct challenges posed by different battery chemistries. Li-ion systems prioritize dendrite suppression and require stable solid electrolyte interphases (SEI) compatible with graphite anodes, necessitating optimization for both mechanical strength and interfacial stability.[189] Na-ion systems face fundamentally different constraints, requiring compatibility with hard carbon anodes due to graphite's incompatibility with Na^+ ions, which shifts the optimization focus toward different voltage windows and interfacial chemistries.[190] Mg-ion systems naturally avoid dendrite formation due to the divalent nature of Mg^{2+}, but face critical challenges from sluggish ion transport kinetics caused by strong solvation effects and higher activation energies, requiring optimization strategies that prioritize conductivity enhancement over mechanical dendrite suppression.[191] Al-ion systems present additional complexity, demanding electrolytes compatible with limited cathode options while managing the high charge density effects of trivalent Al^{3+} ions.[192] Silicon-based Li systems introduce further complications through large volume changes (>300%) that destabilize conventional SEIs, requiring electrolytes optimized for mechanical flexibility and stable interfacial reformation rather than static interfacial stability.[193]

The interplay between these properties is complex; materials with very high ionic conductivity might exhibit poor mechanical properties or a narrow electrochemical stability window. Traditional single-objective ML approaches, predominantly focused on maximizing ionic conductivity,[101,109,194] fail to capture these trade-offs and produce materials unsuitable for practical applications. A critical limitation lies in the lack of frameworks that account for the distinct physics governing different ionic species and their corresponding electrode compatibility requirements. Additionally, the computational expense of evaluating multiple properties for every candidate material during multi-objective optimization searches can be substantial, even when using ML-based surrogate models for property prediction.

View Article Online
DOI: 10.1039/D5MH01525A

Solution 1: Bayesian Optimization for Multi-Objective Materials Discovery

Bayesian Optimization (BO) frameworks efficiently navigate high-dimensional design spaces by constructing probabilistic surrogate models (typically Gaussian Processes) for each objective property. Specialized acquisition functions can incorporate system-specific constraints and property weightings that reflect the distinct requirements of different battery systems. Harada et al. demonstrated this approach by optimizing NASICON-type $LiZr_2(PO_4)_3$ composition co-doped with Ca and Y, simultaneously enhancing Li-ion conductivity, phase stability, and densification.[195] Similarly, BO has been applied to maximize lithium diffusivity while incorporating computational checks for electronic bandgap and stability at lithium metal interfaces, effectively handling multiple criteria through sequential, guided evaluation.[58] Future implementations should incorporate tailored objective weightings—prioritizing mechanical properties for Li systems prone to dendrite formation while emphasizing transport kinetics for Mg systems where sluggish diffusion dominates performance.

Solution 2: Evolutionary Algorithms for Pareto-Optimal Solutions

Evolutionary Algorithms (EAs), including Genetic Algorithms (GAs), inherently support multi-objective optimization through population-based approaches. These algorithms can be enhanced with tailored fitness functions that reflect the distinct physical constraints and performance priorities of different ionic systems. These algorithms apply bio-inspired operators (selection, crossover, mutation) to iteratively improve candidate populations against multiple fitness criteria, generating Pareto-optimal solution sets representing optimal trade-offs where no objective can be improved without degrading others. While direct applications to comprehensive inorganic SSE discovery remain limited, frameworks like evolutionary variational autoencoders (EVAPD) developed for perovskite discovery[196] demonstrate adaptability to SSE applications through suitable multi-objective fitness function definitions.

Solution 3: Collaborative Framework Development and Physics-Informed Search Strategies

Effective multi-objective optimization requires enhanced collaboration between ML specialists and battery application experts to define meaningful optimization targets with application-specific weighting schemes and constraint hierarchies. Electric vehicle batteries might prioritize safety-related mechanical strength and electrochemical stability alongside cycle life, accepting reduced peak ionic conductivity, while high-power portable devices might emphasize maximizing ionic conductivity above other metrics. For Na-ion systems, optimization frameworks should prioritize compatibility with hard carbon anodes and appropriate voltage windows, while Mg-ion systems require frameworks emphasizing transport enhancement strategies such as optimized coordination

DOI: 10.1039/D5MH01525A

environments. Advanced strategies can leverage physical understanding to focus searches on design space regions where multiple desirable properties are more likely to be co-optimized, reducing computational requirements while maintaining search effectiveness by incorporating fundamental materials science principles into the optimization process. This approach addresses both the challenge of defining quantitative targets and minimizing expensive multi-property evaluations through intelligent, system-specific search space reduction.

5.3. Challenge 3: Illuminating the "Black Box": Enhancing Interpretability in ML for SSEs

Complex ML models, particularly deep neural networks, achieve remarkable predictive accuracy but function as "black boxes" that obscure the reasoning behind their predictions. For materials scientists, this lack of transparency presents a significant barrier to trust and adoption, limiting the potential for extracting new scientific understanding. Simply predicting high-performing SSE candidates is insufficient; scientists require insights into why particular materials exhibit desirable properties and what underlying structural features or chemical principles drive performance. Black-box predictions, devoid of such explanations, offer limited utility for advancing fundamental knowledge or formulating new design hypotheses. This interpretability challenge is particularly acute for multivalent systems, where the distinct physics governing Mg^{2+}, Zn^{2+}, and Al^{3+} transport requires understanding of system-specific structure-property relationships that may differ fundamentally from well-studied Li-ion systems.

Solution 1: Model-Agnostic Explainability Methods

Model-agnostic explainability techniques provide insights into ML model behavior without requiring modifications to the underlying algorithms. SHAP (SHapley Additive exPlanations) values, based on game theory, quantify each feature's contribution to specific predictions, while LIME (Local Interpretable Model-agnostic Explanations) explains individual predictions by learning simpler, interpretable models locally around the prediction.[197] The XpertAI framework exemplifies advanced implementation by integrating XAI methods with Large Language Models (LLMs) to automatically generate human-understandable natural language explanations of structure-property relationships.[198] This framework identifies crucial features using XAI and draws upon scientific literature to articulate connections, providing a methodology highly pertinent to understanding ML models for SSEs. A key application is the direct comparison of feature importance, for example, using SHAP values to contrast the governing principles in lithium-based SSEs with those in multivalent systems, identifying which structural motifs (e.g., tetrahedral coordination environments, face-centered lattice arrangements, radial distribution patterns between mobile ions and framework anions) and compositional parameters (e.g., cation electronegativity differences, framework atom ionization energies, anion polarizability) are universal versus cation-specific descriptors.

Solution 2: Interpretable Algorithm Design and Physics-Informed Architectures

Interpretable tree-based ensemble learning methods and graph neural network approaches specifically designed for SSE applications focus on learning and explaining relationships between crystal structures and their corresponding thermodynamic and kinetic properties.[199,200] For multivalent systems, this approach is particularly powerful for extracting explicit design rules from classification models; for instance, a decision tree trained to identify stable hosts could yield a human-readable rule like, 'IF the cation coordination number is > 6 AND the anion framework has a specific void volume, THEN the material is likely to be stable,' directly guiding experimental efforts**.** Chemistry-informed ML models enhance interpretability by incorporating known physical relationships directly into model architecture. For example, a model for solid polymer electrolytes explicitly encoded the Arrhenius equation in its readout layer, enabling prediction of physically meaningful parameters like activation energy (Ea) and pre-exponential factor (A), making temperature-dependent conductivity predictions directly interpretable in terms of fundamental parameters.[201]

Solution 3: Extraction of Scientific Insights and Design Principles

XAI applications in SSE research have successfully extracted human-understandable insights that translate into actionable design principles. Studies examining factors affecting dendrite suppression revealed that material stiffness increases with mass density and the ratio of Li to sublattice bond ionicity while decreasing with increasing volume per atom and sublattice electronegativity.[202] Universal machine learning interatomic potentials have uncovered how crystal structure, anion disorder levels, and mobile ion arrangement influence ionic transport. Simulations demonstrated that appropriate S/Cl disorder in Li_6PS_5Cl enhances diffusion pathway connectivity, improving ionic conductivity.[178] Extending this approach, XAI can help answer related questions in multivalent systems by revealing how system-specific descriptors, such as the migrating cation's ionic radius and charge density, supplant or interact with the framework properties that are dominant in lithium conductors. Heuristic structure descriptors derived from universal interatomic potential analysis rank materials by expected ionic mobility, reflecting potential energy surface properties that correlate with ion hopping.[203]

Solution 4: Iterative Model Refinement Through Explainability

XAI insights create a virtuous cycle by informing future feature engineering efforts and model development. When XAI consistently highlights specific structural motifs (coordination environments for Li^+ ions, framework topologies) or chemical attributes as critical for high performance across diverse SSE candidates, this leads to formulation of new, generalizable scientific knowledge and design principles. For example, if XAI consistently identifies cation coordination environments as

View Article Online
DOI: 10.1039/D5MH01525A

critical for multivalent ion mobility, this insight can be used to engineer more sophisticated features for the next generation of models, thereby accelerating the discovery cycle for novel battery chemistries. Complex derived features identified as highly predictive (combinations of bond angles and lengths defining specific local environments) can be explicitly calculated and incorporated into subsequent, potentially simpler and more robust ML models.

Emerging approaches promise to advance beyond feature importance quantification toward mechanistic discovery. Causal machine learning methods can distinguish genuine causal relationships from spurious correlations in structure-property data, revealing which structural modifications directly influence ionic conductivity versus those that merely correlate.[204] Symbolic regression techniques, which search for explicit mathematical equations governing material properties, offer an alternative path to interpretability by automatically discovering closed-form expressions that relate compositional and structural descriptors to transport properties or rediscover interatomic potentials.[205] These physics-discovering approaches could uncover governing equations analogous to how the Arrhenius relation describes temperature-dependent conductivity, potentially revealing universal scaling laws across different ionic systems.

This iterative refinement, guided by explainability, produces models that are both accurate and grounded in scientifically meaningful parameters, representing a shift from ML merely predicting outcomes to actively contributing to fundamental understanding of solid-state ionics.

Despite these promising developments, successful implementation of XAI in SSE research requires awareness of key methodological limitations. SHAP values exhibit instability in highly correlated feature spaces typical of materials datasets, where structural descriptors often show strong interdependencies.[206] LIME's local approximations may inadequately represent global model behavior, particularly problematic for complex structure-property relationships.[207] Both approaches assume feature independence, which conflicts with the intrinsically coupled nature of atomic positions, coordination environments, and bonding in crystalline materials. Best practices include validating XAI outputs through multiple complementary methods, examining feature correlation matrices before interpretation, and systematically cross-checking computational insights against experimental observations and established physical principles.

5.4. Challenge 4: Bridging Chemical Spaces: Enhancing Model Transferability and Generalization

A significant hurdle for the practical application of ML in SSE discovery is the ability of models to generalize from known materials to novel chemical compositions and crystal structures. Models trained on specific datasets, often limited to well-explored Li-based compounds, frequently exhibit poor performance when tasked with extrapolating to uncharted territories, such as Na^+-based systems or, more drastically, multivalent ion conductors.

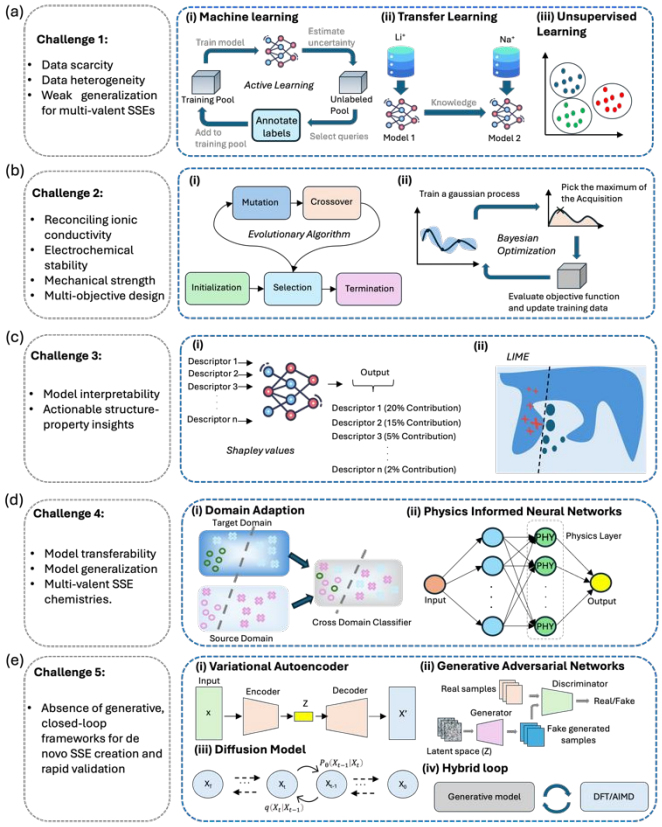

Figure 5: Key Machine Learning Strategies to Accelerate Solid-State Electrolyte (SSE) Discovery. This figure illustrates five classes of ML methods used to address critical challenges in SSE research, from data scarcity to *de novo* design. **(a) Data scarcity:** To combat data scarcity, (i) active learning based iterative loops are used to intelligently guide expensive data acquisition, (ii) transfer learning mitigates the need for a large dataset in a target domain by leveraging knowledge gained from a related, data-rich source domain and (iii) unsupervised learning to identify patterns and promising candidates in unlabeled data. **(b) Multi-Objective Optimization (MOO):** To reconcile competing material properties, techniques like (i) Evolutionary Algorithms and (ii) Bayesian Optimization navigate design trade-offs (e.g., ionic conductivity vs. stability) to identify Pareto-optimal materials. **(c) Explainable AI (XAI):** To overcome the "black-box" nature of ML models, methods like (i) SHAP (Shapley values) and (ii) LIME are applied to quantify feature importance, providing human-understandable insights into structure-property relationships. **(d) Transfer Learning:** To improve model generalization across different chemical families, knowledge from a data-rich source (e.g., Li-ion systems) is transferred to a data-scarce target (e.g., multivalent conductors) using methods like (i) domain adaptation or (ii) Physics informed neural networks. **(e) Generative & Hybrid Frameworks:** For de novo material design, generative models like (i) VAEs, (ii) GANs, and (iii) Diffusion Models propose novel compositions and crystal structures, which are then validated in a (iv) closed-loop with DFT/AIMD simulations to enable rapid, autonomous discovery.

A significant hurdle for practical ML application in SSE discovery is the ability of models to generalize from known materials to novel chemical compositions and crystal structures. Models trained on specific datasets, often limited to well-explored Li-based compounds, frequently exhibit poor performance when extrapolating to uncharted territories such as Na^+-based systems or multivalent ion conductors. The core issue is that ML models excel at interpolation within their training data domain but struggle with extrapolation to chemically distinct regions. Conventional cross-validation techniques, which randomly split data into training and test sets, often overestimate a model's

View Article Online
DOI: 10.1039/D5MH01525A

true extrapolative power because test sets usually contain materials chemically similar to training data. More rigorous "leave-one-group-out cross-validation" (LOGO-CV), where entire chemical families are held out for testing, has demonstrated that conventional ML methods can fail when predicting properties of completely novel compound classes.[188] This presents a critical concern for SSE discovery, where the goal is often to identify entirely new material families with breakthrough properties. While universal interatomic potentials like M3GNet are trained on vast databases (e.g., the Materials Project) and aim for broad applicability across diverse chemical spaces,[208] achieving reliable extrapolation remains a frontier challenge.

Solution 1: Domain Adaptation Techniques

Domain adaptation encompasses ML techniques designed to leverage knowledge from a "source" domain, where data may be abundant, to improve model performance on a "target" domain that might be data-scarce or have different underlying data distributions. In SSE contexts, this involves adapting models trained on Li-ion conductors to predict properties for Na-ion or K-ion conductors, or transferring knowledge from computational data to guide experimental outcome predictions. A multi-stage ML approach for electrocatalyst discovery successfully integrated domain adaptation to enhance theoretical simulations and align them with experimental findings,[209] demonstrating a concept directly transferable to SSE research. However, the effectiveness of domain adaptation is intrinsically linked to the relevance of incorporated knowledge; if source and target domains are too disparate in their underlying physics or chemistry, transferred knowledge may be of limited value or even detrimental.

Solution 2: Physics-Informed Machine Learning (PIML)

PIML improves model generalization and physical consistency, especially in data-limited scenarios, by embedding known physical laws, constraints, or symmetries directly into ML model architecture, loss functions, or feature representations. By constraining models to adhere to fundamental physics, PIML can lead to more robust and interpretable predictions that extrapolate better to unseen data. Universal ML potentials for liquid electrolytes, trained via iterative DFT calculations, accurately predict physical properties like density, viscosity, and ionic conductivity, implying that models have learned underlying physical consistencies.[210] The DiffMix model, a differentiable geometric deep learning approach for chemical mixtures, explicitly extends thermodynamic and transport laws (e.g., Vogel-Fulcher-Tammann for ionic conductivity) with GDL-learnable physical coefficients, demonstrating improved accuracy and robustness for predicting liquid electrolyte properties.[211] Nevertheless, PIML success hinges on the accuracy and completeness of embedded physical laws; overly simplified or incomplete physical constraints can restrict a model's ability to learn complex phenomena and generalize correctly.

Solution 3: Advanced Universal Representation Learning

Achieving truly "universal" ML models that can reliably extrapolate across vastly different chemical spaces and discover entirely new material classes remains a formidable scientific challenge. This likely necessitates a paradigm shift towards models that can learn or infer fundamental physical laws more directly from data, rather than relying solely on statistical correlations or pre-defined explicit constraints. Promising approaches include more sophisticated PIML frameworks, integration of ML with symbolic regression techniques to discover governing equations, or development of AI systems capable of formulating and testing new physical hypotheses. Such advanced approaches could potentially overcome the limitations of current transferability strategies by learning more fundamental representations of chemical and physical relationships that generalize across diverse material systems.

5.5. Challenge 5: Beyond Screening: Generative and Hybrid Frameworks for Novel SSE Design

Traditional computational materials discovery, even when augmented by ML, relies on screening predefined candidate lists derived from existing databases or combinatorial variations of known crystal structures. While efficient for exploring local chemical space, these methods are less effective at proposing radically new compositions or structural archetypes that lie outside the initial search parameters. They are fundamentally tools for evaluation rather than de novo creation, inherently limiting the scope of discovery to variations of known materials rather than truly novel SSEs with unprecedented properties.

Solution 1: Deep Generative Models for Novel Material Design

Deep generative models offer a paradigm shift by learning underlying patterns and design rules from existing materials data and using this knowledge to propose entirely new candidate compositions or crystal structures from scratch, often guided by desired performance criteria.

- **VAEs** learn a compressed, continuous latent representation of materials, from which new candidates can be generated by sampling points in this latent space and decoding them back into material structures or compositions. Noh et al. applied a VAE-based framework to the inverse design of solid-state materials, efficiently exploring chemical compositional spaces to generate novel candidates with desired properties.[55]
- **GANs** employ a two-network architecture: a generator that creates new material candidates and a discriminator that tries to distinguish these synthetic candidates from real materials in a training dataset. Through this adversarial training, the generator learns to produce increasingly realistic and potentially novel materials.
- **Diffusion Models** are an emerging class of powerful generative models that operate by learning to reverse a gradual noise-adding process. They have shown significant promise for generating high-quality

View Article Online
DOI: 10.1039/D5MH01525A

samples in various domains, like crystal structure exemplifies an open-source EA for crystal structure

Generative Model Type	Core Working Principle	Strengths for SSE Design	Limitations/Challenges in SSE Context	Key Examples/Potential
Variational Autoencoders (VAEs)	Learns a continuous latent representation of data; new samples generated by decoding points from this latent space.	Smooth latent space allows for interpolation and generation of similar but novel structures/compositions; can be conditioned on properties.	Quality of reconstructed/generated materials can be an issue; ensuring chemical validity and stability of generated crystal structures.	Inverse material design[55].
Generative Adversarial Networks (GANs)	A generator network creates candidates, and a discriminator network tries to distinguish them from real data; adversarial training improves generator.	Capable of generating highly novel and diverse candidates; can learn complex data distributions.	Training can be unstable (mode collapse); ensuring generated crystal structures are physically realistic and stable is challenging.	Crystal structure prediction[128]; Inverse design of materials (MatGAN)[54]
Evolutionary Algorithms (EAs) / Genetic Algorithms (GAs)	Population-based search; applies operators (mutation, crossover, selection) guided by a fitness function (target properties).	Robust global search capabilities; can explicitly handle multiple objectives and complex constraints (e.g., stability, synthesizability).	Can be computationally expensive if fitness evaluation (e.g., DFT calculation) is slow for each candidate; defining effective representations and evolutionary operators for crystal structures.	Crystal structure prediction (XtalOpt)[214]; Guiding phase field exploration for Li-ion conductors[215];
Diffusion Models	Learns to reverse a noise-adding process; new samples generated by iterative denoising from a random starting point.	Can generate very high-quality, realistic samples; emerging as state-of-the-art in many generative tasks.	Can be computationally intensive for sampling; developing effective conditioning mechanisms for specific material properties and crystal symmetries.	General crystal structure generation[212]; MatterGen (fine-tuneable generative model)[213].
Hybrid Models (e.g., VAE-GA)	Combines strengths of different generative approaches, e.g., VAE for generation and GA for optimization.	Potential to overcome limitations of individual methods; e.g., VAE explores broadly, GA refines promising candidates.	Increased model complexity; requires careful integration of components.	EVAPD for perovskites[196].
Integrated Closed-Loop Frameworks	ML proposes candidates → computational validation (DFT) → experimental synthesis/characterization → feedback to refine ML models in iterative cycles.	Combines theoretical prediction with experimental validation; continuous model improvement; reduces experimental waste through guided exploration.	Requires substantial infrastructure investment; standardized synthesis protocols needed; complex integration of computational and experimental platforms; slower iteration cycles.	CAMEO system;[216] Electrolytomics;[217] $Na_xLi_{3-x}YCl_6$ discovery;[218] DiffMix for electrolyte optimization[211]

Table 4: Comparison of Generative Model Approaches for Novel SSE Discovery.

generation.[212] The MatterGen model, for example, can generate stable, diverse inorganic materials and can be fine-tuned to steer generation towards specific property constraints, including chemistry, symmetry, and various physical properties, with one generated structure successfully synthesized and validated.[213]

Solution 2: Evolutionary Algorithms and Hybrid Generative Approaches

Evolutionary Algorithms serve as powerful generative tools, particularly for crystal structure prediction and compositional optimization. EAs maintain a population of candidate solutions (materials) and iteratively apply evolutionary operators like mutation (small changes to composition or structure) and crossover (combining features of good candidates) to generate new candidates. A fitness function incorporating predicted stability, ionic conductivity, and other desired properties guides the selection of candidates for subsequent generations. XtalOpt prediction.[214] Unsupervised ML has also guided the prioritization of elemental phase fields for synthetic exploration, leading to the discovery of a novel quaternary lithium solid electrolyte in a collaborative workflow resembling evolutionary search.[215]

A hybrid approach combining a VAE with a genetic algorithm, termed the Evolutionary Variational Autoencoder for Perovskite Discovery (EVAPD), has been developed to discover new perovskite materials.[196] This framework leverages the VAE's ability to generate diverse candidates from a learned latent space and the GA's strength in optimizing these candidates based on a defined fitness function (e.g., predicted stability). Such hybrid generative approaches hold considerable potential for SSE discovery if adapted with relevant property targets.

The success of these generative models is critically dependent on the quality and relevance of the design rules or property targets they are given. If these targets are ill-defined, incomplete (focusing only on ionic conductivity without considering stability or synthesizability), or do not capture all essential practical constraints, the generated candidates may be theoretically interesting but practically irrelevant or impossible to realize. The ability of models like MatterGen to be fine-tuned for a broad range of property constraints, and its subsequent experimental validation, underscores the importance of multi-faceted and accurate guidance for generative design.[213]

Solution 3: Integrated Closed-Loop Experimental-Computational Frameworks

The true acceleration in SSE discovery is anticipated from hybrid design frameworks that tightly integrate ML predictions with DFT calculations for validation and, crucially, with experimental synthesis and characterization in a closed-loop or active learning fashion. These "predictive synthesis" pipelines allow ML models to propose candidate materials, which are then computationally validated (e.g., by DFT for stability and preliminary property estimates) and/or experimentally synthesized and tested. The results feed back into the ML model, refining its predictions and guiding the next iteration of discovery.

Several pioneering efforts exemplify this approach:

- The CAMEO system is a real-time, closed-loop autonomous materials exploration platform that uses Bayesian active learning integrated with synchrotron beamline experiments for on-the-fly phase mapping and property optimization.[216]
- The "Electrolytomics" initiative describes an AI-guided approach that combines data science, robotic experimentation for validation, and computation, leading to the discovery and experimental confirmation of novel high-performance liquid electrolytes.[217]
- A computational-experimental pipeline successfully combined AI models, physics-based simulations on cloud HPC for large-scale screening, and subsequent experimental synthesis and characterization to discover promising new SSE compositions like $Na_xLi_{3-x}YCl_6$.[218]
- The DiffMix model, a differentiable GDL model, has been used to guide robotic experimentation for optimizing fast-charging liquid battery electrolytes, achieving significant conductivity improvements in few experimental steps.[211]
- An integrated high-throughput robotic platform combined with active learning has been developed to accelerate the discovery of optimal liquid electrolyte formulations. This approach efficiently identifies high-solubility redox-active molecules by evaluating a small fraction of candidates, demonstrating the effectiveness of closed-loop frameworks in materials discovery.[219]
- Iterative training of universal MLPs, where DFT calculations are performed on structures where the MLP shows high uncertainty, also represents a form of closed-loop learning to refine the potential across a wide chemical space.[210]

Fully autonomous closed-loop systems, often termed "self-driving laboratories," represent the apex of accelerated materials discovery. However, their widespread adoption for SSE research faces significant hurdles. Beyond the continued advancement of ML algorithms and robotic platforms, a major challenge lies in the development of standardized, automatable, and rapid synthesis and characterization protocols suitable for the diverse range of solid-state chemistries. The synthesis of inorganic solids often involves high temperatures, controlled atmospheres, and multi-step processes that are not as easily automated as liquid-phase formulations. Furthermore, critical to the success of these frameworks is the implementation of robust validation workflows that prevent costly experimental efforts on unfeasible materials. Effective validation protocols should include thermodynamic stability screening via DFT hull distance calculations, with chemistry-dependent thresholds based on the metastability scales established for different material classes,[220] kinetic accessibility assessment through thermodynamic upper bounds such as the amorphous limit for polymorph synthesizability,[221] and rapid experimental validation using automated characterization techniques[222] such as XRD phase identification and impedance spectroscopy.[223,224] These multi-tier filters ensure that generative models guide experimental efforts toward genuinely promising candidates rather than thermodynamically unstable or synthetically inaccessible compositions.

The cost and complexity of establishing and maintaining such highly integrated experimental and computational platforms, combined with the need for standardized validation protocols, require substantial investment and interdisciplinary expertise.

Table 4 provides a comparative overview of different generative model approaches and their potential in the context of novel SSE discovery.

6. Conclusion: Charting the Path Forward for AI-Accelerated SSE Innovation

The journey towards high-performance, safe, and commercially viable solid-state electrolytes is complex, yet the integration of machine learning offers unprecedented opportunities to accelerate progress. This review has highlighted several critical research gaps and challenges that currently temper the full impact of ML in this domain: the persistent scarcity of diverse, high-quality data, especially for multivalent ion systems and interfacial phenomena; the necessity for multi-objective optimization to balance competing performance metrics; the demand for interpretable ML models that provide scientific insights rather than just black-box predictions; the crucial need for models that can generalize and transfer knowledge across

View Article Online
DOI: 10.1039/D5MH01535A

diverse chemical spaces and novel material classes; and the imperative to move beyond screening predefined candidates towards generative design of entirely new materials within hybrid, closed-loop discovery frameworks.

The data scarcity challenge is particularly acute for multivalent systems (Mg^{2+}, Ca^{2+}, Zn^{2+}, Al^{3+}), where solid-state battery research remains in its early stages both experimentally and computationally. Beyond the stark quantitative disparity with Li-ion SSE databases containing thousands of compounds while Mg^{2+}, Ca^{2+}, Zn^{2+}, and Al^{3+} conductors each number in the tens to low hundreds[225], these systems exhibit fundamentally different physics that cannot be addressed through simple data augmentation. Multivalent ions face stronger Coulombic interactions with the host lattice due to their higher charge densities, leading to sluggish diffusion kinetics and significantly higher activation energies compared to monovalent systems.[177] The migration mechanisms differ qualitatively: while Li^+ transport often proceeds via direct hopping between tetrahedral sites, Mg^{2+} migration typically requires concerted structural relaxation or even temporary coordination changes to overcome the strong cation-anion binding. Additionally, defect chemistry and strain accommodation mechanisms vary substantially—multivalent dopants introduce different charge compensation schemes and elastic distortions that alter migration pathways in ways not captured by Li-based training data. These mechanistic distinctions mean that ML models trained predominantly on Li-ion data lack the physical descriptors and feature representations necessary to capture the governing principles in multivalent systems, creating a critical bottleneck for advancing beyond lithium-ion technologies that cannot be resolved by transfer learning alone without substantial new data generation and physics-informed constraints.

Encouragingly, the research landscape is actively addressing these challenges. Strategies such as transfer learning, unsupervised learning, and advanced data augmentation techniques are being developed to combat data limitations. Physics-informed machine learning and the pursuit of universal descriptors and interatomic potentials aim to enhance model transferability and generalization. Explainable AI methods are beginning to shed light on the complex structure-property relationships learned by ML models, fostering trust and guiding scientific intuition. Furthermore, generative models, including VAEs, GANs, EAs, and diffusion models, are showing increasing promise in proposing novel SSE candidates from scratch, while sophisticated multi-objective optimization algorithms are helping to navigate the intricate trade-offs inherent in materials design. The most transformative advances, however, are emerging from hybrid frameworks that tightly integrate ML predictions with high-fidelity computations (like DFT) and, crucially, experimental validation, often within automated, closed-loop "predictive synthesis" pipelines.

This review provides several distinctive contributions that advance the field beyond existing literature. We present the first systematic mapping of five interconnected challenges with corresponding emerging solutions, providing a strategic roadmap for practitioners. Unlike previous reviews that predominantly focus on Li-ion systems, we emphasize the critical data gap for multivalent systems and provide specific strategies for addressing this limitation through transfer learning and physics-informed approaches. We uniquely bridge conventional computational methods with cutting-edge ML techniques, demonstrating how hybrid workflows can overcome individual limitations while leveraging complementary strengths. Rather than merely surveying techniques, we provide actionable recommendations for data collection priorities, validation strategies, and best practices for applying explainable AI methods to materials discovery.

To further propel AI-accelerated SSE innovation, future research should prioritize several key directions. The development of next-generation multi-objective optimization algorithms that can simultaneously optimize ionic conductivity, electrochemical stability, mechanical properties, and synthesizability while incorporating real-world constraints represents a critical need. Physics-informed universal models that embed fundamental physical laws governing ionic transport and electrochemical stability directly into model architecture require immediate attention. These must learn temperature-dependent behavior, incorporate many-body interactions, and predict interfacial stability through first-principles constraints.

Robust uncertainty quantification methods for ML predictions, particularly when extrapolating to novel chemical spaces, represent another urgent priority. Cross-domain transfer learning protocols must be established to enable knowledge transfer between different ion types and between computational and experimental domains. Several fundamental research questions require immediate investigation: How can we systematically quantify and improve model transferability across different crystal structure families and ionic species? What are optimal strategies for incorporating experimental uncertainty into ML training datasets? How can we develop models that predict long-term degradation and interfacial evolution beyond static property prediction?

The practical implementation of these advances requires immediate action across multiple fronts. A concerted community-wide effort is essential to build FAIR[226] databases that encompass multivalent systems and include comprehensive interfacial property data with standardized metadata. The integration of automated synthesis platforms specifically designed for SSE discovery represents a transformative opportunity, requiring real-time characterization capabilities and automated feedback loops. Comprehensive validation workflows for generative models must include thermodynamic stability screening, kinetic accessibility assessment, and rapid experimental validation using automated characterization techniques.

Future experimental and computational campaigns should prioritize multivalent systems with intermediate ionic radii, materials exhibiting mixed ionic-electronic conductivity, and interfacial properties that remain underrepresented in current databases. The establishment of industry-academic partnerships will be crucial for scaling promising discoveries to commercial applications, while advanced generative models

must be refined to ensure chemical validity, thermodynamic stability, and practical synthesizability of proposed candidates. The path forward for revolutionizing SSE development lies in a deeply synergistic approach where machine learning realizes its transformative potential through intimate integration with fundamental domain knowledge from physics and chemistry, rigorous computational modeling, and iterative experimental validation. As these integrated intelligence frameworks mature, particularly those enabling autonomous closed-loop discovery, the pace of innovation in solid-state electrolytes is poised for significant acceleration, bringing the promise of safer, more energy-dense, and longer-lasting battery technologies closer to reality.

Conflicts of interest

There are no conflicts to declare.

Data Availability

No primary research results, software or code have been included, and no new data were generated or analysed as part of this review.

Acknowledgements

This project is partially supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of Schmidt Sciences, LLC

Keywords

Machine learning, batteries, solid-state electrolytes, energy materials, computational science

References

1 V. Etacheri, R. Marom, R. Elazari, G. Salitra and D. Aurbach, *Energy Environ. Sci.*, 2011, **4**, 3243–3262.
2 B. Dunn, H. Kamath and J.-M. Tarascon, *Science*, 2011, **334**, 928–935.
3 J. W. Choi and D. Aurbach, *Nat. Rev. Mater.*, 2016, **1**, 1–16.
4 J. B. Goodenough and Y. Kim, *Chem. Mater.*, 2010, **22**, 587–603.
5 R. Gond, W. van Ekeren, R. Mogensen, A. J. Naylor and R. Younesi, *Mater. Horiz.*, 2021, **8**, 2913–2928.
6 X. Fan, L. Chen, O. Borodin, X. Ji, J. Chen, S. Hou, T. Deng, J. Zheng, C. Yang, S.-C. Liou, K. Amine, K. Xu and C. Wang, *Nat. Nanotechnol.*, 2018, **13**, 715–722.
7 W. Xu, J. Wang, F. Ding, X. Chen, E. Nasybulin, Y. Zhang and J.-G. Zhang, *Energy Environ. Sci.*, 2014, **7**, 513–537.
8 D. Aurbach, E. Zinigrad, Y. Cohen and H. Teller, *Solid State Ion.*, 2002, **148**, 405–416.
9 K. Xu, *Chem. Rev.*, 2004, **104**, 4303–4418.
10 J. Janek and W. G. Zeier, *Nat. Energy*, 2023, **8**, 230–240.
11 Q. Zhao, S. Stalin, C.-Z. Zhao and L. A. Archer, *Nat. Rev. Mater.*, 2020, **5**, 229–252.
12 J. Janek and W. G. Zeier, *Nat. Energy*, 2016, **1**, 1–4.
13 N. Kamaya, K. Homma, Y. Yamakawa, M. Hirayama, R. Kanno, M. Yonemura, T. Kamiyama, Y. Kato, S. Hama, K. Kawamoto and A. Mitsui, *Nat. Mater.*, 2011, **10**, 682–686.
14 D. Lin, Y. Liu and Y. Cui, *Nat. Nanotechnol.*, 2017, **12**, 194–206.
15 M. Pasta, D. Armstrong, Z. L. Brown, J. Bu, M. R. Castell, P. Chen, A. Cocks, S. A. Corr, E. J. Cussen, E. Darnbrough, V. Deshpande, C. Doerrer, M. S. Dyer, H. El-Shinawi, N. Fleck, P. Grant, G. L. Gregory, C. Grovenor, L. J. Hardwick, J. T. S. Irvine, H. J. Lee, G. Li, E. Liberti, I. McClelland, C. Monroe, P. D. Nellist, P. R. Shearing, E. Shoko, W. Song, D. S. Jolly, C. I. Thomas, S. J. Turrell, M. Vestli, C. K. Williams, Y. Zhou and P. G. Bruce, *J. Phys. Energy*, 2020, **2**, 032008.
16 H.-L. Yang, B.-W. Zhang, K. Konstantinov, Y.-X. Wang, H.-K. Liu and S.-X. Dou, *Adv. Energy Sustain. Res.*, 2021, **2**, 2000057.
17 R. Murugan, V. Thangadurai and W. Weppner, *Angew. Chem. Int. Ed.*, 2007, **46**, 7778–7781.
18 N. Kamaya, K. Homma, Y. Yamakawa, M. Hirayama, R. Kanno, M. Yonemura, T. Kamiyama, Y. Kato, S. Hama, K. Kawamoto and A. Mitsui, *Nat. Mater.*, 2011, **10**, 682–686.
19 B. Chen, J. Ju, J. Ma, H. Du, R. Xiao, G. Cui and L. Chen, *Comput. Mater. Sci.*, 2018, **153**, 170–175.
20 F. Gucci, Grasso ,Marzio, Shaw ,Christopher, Leighton ,Glenn, Marchante Rodriguez ,Veronica and J. and Brighton, *Polym.-Plast. Technol. Mater.*, 2023, **62**, 1019–1028.
21 J. H. Cha, P. N. Didwal, J. M. Kim, D. R. Chang and C.-J. Park, *J. Membr. Sci.*, 2020, **595**, 117538.
22 J. Feng, L. Wang, Y. Chen, P. Wang, H. Zhang and X. He, *Nano Converg.*, 2021, **8**, 2.
23 H. Gao, N. S. Grundish, Y. Zhao, A. Zhou and J. B. Goodenough, *Energy Mater. Adv.*, DOI:10.34133/2021/1932952.
24 W. Xie, Z. Deng, Z. Liu, T. Famprikis, K. T. Butler and P. Canepa, *Adv. Energy Mater.*, 2024, **14**, 2304230.
25 S. Wang, A. L. Monaca and G. P. Demopoulos, *Energy Adv.*, 2025, **4**, 11–36.
26 R. Zhao, G. Hu, S. Kmiec, J. Wheaton, V. M. Torres III and S. W. Martin, *Batter. Supercaps*, 2022, **5**, e202100356.
27 T. K. Schwietert, A. Vasileiadis and M. Wagemaker, *JACS Au*, 2021, **1**, 1488–1496.
28 G. Ceder, S. P. Ong and Y. Wang, *MRS Bull.*, 2018, **43**, 746–751.
29 H. Guo, Q. Wang, A. Stuke, A. Urban and N. Artrith, *Front. Energy Res.*, DOI:10.3389/fenrg.2021.695902.
30 N. A. W. Holzwarth, *Phys. Procedia*, 2014, **57**, 29–37.
31 A. Urban, D.-H. Seo and G. Ceder, *Npj Comput. Mater.*, 2016, **2**, 1–13.
32 B. Liu, P. Liao, X. Shi, Y. Wen, Q. Gou, M. Yu, S. Zhou and X. Sun, *RSC Adv.*, 2022, **12**, 34627–34633.
33 K. Sau, S. Takagi, T. Ikeshoji, K. Kisu, R. Sato, E. C. dos Santos, H. Li, R. Mohtadi and S. Orimo, *Commun. Mater.*, 2024, **5**, 1–27.
34 N. J. J. de Klerk, E. van der Maas and M. Wagemaker, *ACS Appl. Energy Mater.*, 2018, **1**, 3230–3242.
35 X. He, Y. Zhu and Y. Mo, *Nat. Commun.*, 2017, **8**, 15893.
36 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
37 R. Xiao, H. Li and L. Chen, *Sci. Rep.*, 2015, **5**, 14227.
38 C. Lv, X. Zhou, L. Zhong, C. Yan, M. Srinivasan, Z. W. Seh, C. Liu, H. Pan, S. Li, Y. Wen and Q. Yan, *Adv. Mater.*, 2022, **34**, 2101474.
39 J. Dai, Y. Jiang and W. Lai, *Phys. Chem. Chem. Phys.*, 2022, **24**, 15025–15033.
40 G. Wang, C. Wang, X. Zhang, Z. Li, J. Zhou and Z. Sun, *iScience*, 2024, **27**, 109673.
41 J. Kim, D. Lee, D. Lee, X. Li, Y.-L. Lee and S. Kim, *J. Phys. Chem. Lett.*, 2024, **15**, 5914–5922.
42 A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui and E. J. Reed, *Chem. Mater.*, 2019, **31**, 342–352.
43 Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo and C. Ling, *Nat. Commun.*, 2019, **10**, 5260.
44 X. Guo, Z. Wang, J.-H. Yang and X.-G. Gong, *J. Mater. Chem. A*, 2024, **12**, 10124–10136.
45 Y. Zhan, W. Zhang, B. Lei, H. Liu and W. Li, *Front. Chem.*, DOI:10.3389/fchem.2020.00125.
46 X. Tang, D. Zhou, B. Zhang, S. Wang, P. Li, H. Liu, X. Guo, P. Jaumaux, X. Gao, Y. Fu, C. Wang, C. Wang and G. Wang, *Nat. Commun.*, 2021, **12**, 2857.
47 Y. Xu, Y. Zong and K. Hippalgaonkar, *J. Phys. Commun.*, DOI:10.1088/2399-6528/ab92d8.
48 Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo and C. Ling, *Nat. Commun.*, 2019, **10**, 5260.

49 Z. Lu, P. Adeli, C.-H. Yim, M. Jiang, J. Rempel, Z. W. Chen, S. Yadav, P. Mercier, Y. Abu-Lebdeh and C. V. Singh, *ACS Appl. Energy Mater.*, 2022, **5**, 8042–8048.
50 J. Sun, S. Kang, J. Kim and K. Min, *ACS Appl. Mater. Interfaces*, 2023, **15**, 5049–5057.
51 W. Chen, J. Zhou, S. Li, C. Lu, H. Li, Y. Li, Y. Cheng, J. Yang and Y. He, *J. Alloys Compd.*, 2025, **1010**, 177981.
52 K. Wang, V. Gupta, C. S. Lee, Y. Mao, M. N. T. Kilic, Y. Li, Z. Huang, W. Liao, A. Choudhary and A. Agrawal, *Sci. Rep.*, 2024, **14**, 25178.
53 X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski and T. Y.-J. Han, *Npj Comput. Mater.*, 2022, **8**, 1–19.
54 Y. Dan, Y. Zhao, X. Li, S. Li, M. Hu and J. Hu, *Npj Comput. Mater.*, 2020, **6**, 1–7.
55 J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik and Y. Jung, *Matter*, 2019, **1**, 1370–1384.
56 C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka and T. Xie, *Nature*, 2025, **639**, 624–632.
57 J. C. Verduzco, E. E. Marinero and A. Strachan, *Integrating Mater. Manuf. Innov.*, 2021, **10**, 299–310.
58 S. A. Tawfik, J. Berk, T. R. Walsh, S. Rana and S. Venkatesh, *J. Phys. Chem. C*, 2025, **129**, 6148–6156.
59 C. Wang, K. Fu, S. P. Kammampata, D. W. McOwen, A. J. Samson, L. Zhang, G. T. Hitz, A. M. Nolan, E. D. Wachsman, Y. Mo, V. Thangadurai and L. Hu, *Chem. Rev.*, 2020, **120**, 4257–4300.
60 G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.
61 M.-S. Lim and S.-H. Jhi, *Curr. Appl. Phys.*, 2018, **18**, 541–545.
62 K. Ueno, K. Ichikawa, K. Sato, D. Sugita, S. Yotsuhashi and I. Takeuchi, *Phys. Rev. Mater.*, 2021, **5**, 033801.
63 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
64 B. H. Sjølin, P. B. Jørgensen, A. Fedrigucci, T. Vegge, A. Bhowmik and I. E. Castelli, *Batter. Supercaps*, 2023, **6**, e202300041.
65 Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood and S. P. Ong, *J. Phys. Chem. A*, 2020, **124**, 731–745.
66 L. K. Béland, *Phys. Rev. E*, DOI:10.1103/PhysRevE.84.046704.
67 H. Xu, Y. N. Osetsky and R. E. Stoller, *J. Phys. Condens. Matter*, 2012, **24**, 375402.
68 K. Ferasat, Y. N. Osetsky, A. V. Barashev, Y. Zhang, Z. Yao and L. K. Béland, *J. Chem. Phys.*, 2020, **153**, 074109.
69 M. Soleymanibrojeni, C. R. C. Rego, M. Esmaeilpour and W. Wenzel, *J. Mater. Chem. A*, 2024, **12**, 2249–2266.
70 Z. Deng, T. P. Mishra, E. Mahayoni, Q. Ma, A. J. K. Tieu, O. Guillon, J.-N. Chotard, V. Seznec, A. K. Cheetham, C. Masquelier, G. S. Gautam and P. Canepa, *Nat. Commun.*, 2022, **13**, 4470.
71 Y. Chen, D. Liang, E. M. Y. Lee, S. Muy, M. Guillaume, M.-D. Braida, A. A. Emery, N. Marzari and J. J. de Pablo, *ACS Appl. Mater. Interfaces*, 2024, **16**, 48223–48234.
72 N. J. J. de Klerk, E. van der Maas and M. Wagemaker, *ACS Appl. Energy Mater.*, 2018, **1**, 3230–3242.
73 A. Muralidharan, M. I. Chaudhari, L. R. Pratt and S. B. Rempe, *Sci. Rep.*, 2018, **8**, 10736.
74 B. Andriyevsky, K. Doll and T. Jacob, *Mater. Chem. Phys.*, 2017, **185**, 210–217.
75 T. Cheng, B. V. Merinov, S. Morozov and W. A. I. Goddard, *ACS Energy Lett.*, 2017, **2**, 1454–1459.
76 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (pymatgen) (version 2022.1.24) , DOI:10.1016/j.commatsci.2012.10.028 2013.
77 matminer (Materials Data Mining) — matminer 0.1.0 documentation, https://matminer.readthedocs.io/en/latest/, (accessed 14 May 2025).
78 S. Rühl, in *Materials Informatics*, John Wiley & Sons, Ltd, 2019, pp. 41–54.
79 R. H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. Buongiorno Nardelli and S. Curtarolo, *Comput. Mater. Sci.*, 2014, **93**, 178–192.
80 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *Npj Comput. Mater.*, 2015, **1**, 1–15.
81 K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe and F. Tavazza, *Npj Comput. Mater.*, 2020, **6**, 1–13.
82 D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J. K. Nørskov and K. W. Jacobsen, *Comput. Sci. Eng.*, 2012, **14**, 51–57.
83 L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, C. S. Adorf, C. W. Andersen, O. Schütt, C. A. Pignedoli, D. Passerone, J. VandeVondele, T. C. Schulthess, B. Smit, G. Pizzi and N. Marzari, *Sci. Data*, 2020, **7**, 299.
84 S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs and A. Le Bail, *Nucleic Acids Res.*, 2012, **40**, D420–D427.
85 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
86 J. Schmidt, T. F. T. Cerqueira, A. H. Romero, A. Loew, F. Jäger, H.-C. Wang, S. Botti and M. A. L. Marques, *Mater. Today Phys.*, 2024, **48**, 101560.
87 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, M. A. Wright, B.-E. Prasad, F. Blanc, C. M. Collins, C. A. Crawford, B. B. Duff, J. Evans, J. Gamon, G. Han, B. T. Leube, H. Niu, A. J. Perez, A. Robinson, O. Rogan, P. M. Sharp, E. Shoko, M. Sonni, W. J. Thomas, A. Vasylenko, L. Wang, M. J. Rosseinsky and M. S. Dyer, *Npj Comput Mater*, 2023, **9**, 1–14.
88 F. Therrien, J. A. Haibeh, D. Sharma, R. Hendley, A. Hernández-García, S. Sun, A. Tchagang, J. Su, S. Huberman, Y. Bengio, H. Guo and H. Shin, *arXiv*, 2025, preprint, arXiv:arXiv:2502.14234, DOI: 10.48550/arXiv.2502.14234.
89 O. Kononova, T. He, H. Huo, A. Trewartha, E. A. Olivetti and G. Ceder, *iScience*, DOI:10.1016/j.isci.2021.102155.
90 Y.-J. Shon and K. Min, *ACS Omega*, 2023, **8**, 18122–18127.
91 A. Nandy, C. Duan and H. J. Kulik, *Curr. Opin. Chem. Eng.*, 2022, **36**, 100778.
92 P. Xu, X. Ji, M. Li and W. Lu, *Npj Comput. Mater.*, 2023, **9**, 42.
93 S. Kang, M. Kim and K. Min, *J. Phys. Chem. C*, 2023, **127**, 19335–19343.
94 C. W. Park and C. Wolverton, *Phys. Rev. Mater.*, 2020, **4**, 063801.
95 F. A. L. Laskowski, D. B. McHaffie and K. A. See, *Energy Env. Sci*, 2023, **16**, 1264–1276.
96 J. Kim, D. Lee, D. Lee, X. Li, Y.-L. Lee and S. Kim, *J. Phys. Chem. Lett.*, 2024, **15**, 5914–5922.
97 B. Ramsundar, P. Eastman, P. Walters and V. Pande, *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*, O'Reilly, 2019.
98 Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman and V. Viswanathan, *ACS Cent. Sci.*, 2018, **4**, 996–1006.
99 Y. Zhao, N. Schiffmann, A. Koeppe, N. Brandt, E. C. Bucharsky, K. G. Schell, M. Selzer and B. Nestler, *Front. Mater.*, DOI:10.3389/fmats.2022.821817.
100 Y. Chen, M. Duquesnoy, D. Tan, J. Doux, H. Yang, G. Deysher, P. Ridley, A. Franco, Y. Meng and Z. Chen, *ACS ENERGY Lett.*, 2021, **6**, 1639–1648.
101 A. Adhyatma, Y. Xu, N. H. Hawari, P. Satria Palar and A. Sumboja, *Mater. Lett.*, 2022, **308**, 131159.
102 S. Pereznieto, R. Jaafreh, J. Kim and K. Hamad, *Mater. Lett.*, 2023, **349**, 134848.
103 J. Kim, S. Kang and K. Min, *ACS Appl. Mater. Interfaces*, 2023, **15**, 41417–41425.
104 L. Tang, G. Zhang and J. Jiang, *Chin. J. Chem. Phys.*, 2024, **37**, 505–512.
105 Y. Zhang, T. Zhan, Y. Sun, L. Lu and B. Chen, *ChemSusChem*, DOI:10.1002/cssc.202301284.
106 D. Park, W. Chung, B. Min, U. Lee, S. Yu and K. Kim, *NPJ Comput. Mater.*, DOI:10.1038/s41524-024-01392-6.
107 A. Gallo-Bueno, M. Reynaud, M. Casas-Cabanas and J. Carrasco, *Energy AI*, DOI:10.1016/j.egyai.2022.100159.
108 Y. Wang, S. Li and M. Chen, *Mater. Today Commun.*, DOI:10.1016/j.mtcomm.2024.108294.
109 Y. Ma, S. Han, Y. Sun, Z. Cui, P. Liu, X. Wang and Y. Wang, *J. Power Sources*, 2024, **604**, 234492.
110 A. K. Mishra, S. Rajput, M. Karamta and I. Mukhopadhyay, *ACS Omega*, 2023, **8**, 16419–16427.

111 M. Kurniawan, M. H. Alfaruqi, A. N. Fahri, S. Lee and J. Kim, *J. Phys. Chem. Solids*, DOI:10.1016/j.jpcs.2025.112752.
112 Z. Wang, X. Lin, Y. Han, J. Cai, S. Wu, X. Yu and J. Li, *Nano Energy*, 2021, **89**, 106337.
113 Z. Wang, J. Gao, K. Tao, Y. Han, A. Chen and J. Li, *Energy Storage Mater.*, 2023, **59**, 102781.
114 J. Dean, M. Scheffler, T. A. R. Purcell, S. V. Barabash, R. Bhowmik and T. Bazhirov, *J. Mater. Res.*, 2023, **38**, 4477–4496.
115 Z. Lu, P. Adeli, C. Yim, M. Jiang, J. Rempel, Z. Chen, S. Yadav, P. Mercier, Y. Abu-Lebdeh and C. Singh, *ACS Appl. ENERGY Mater.*, DOI:10.1021/acsaem.2c00493.
116 C. Pyo, J. Kim, J. Lim, J. Sim and S. Kang, *ECS Meet. Abstr.*, 2024, **MA2024-02**, 360.
117 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
118 C. W. Park and C. Wolverton, *Phys. Rev. Mater.*, 2020, **4**, 063801.
119 C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
120 C. Chen, W. Ye, Y. Zuo, C. Zheng and S. P. Ong, *Chem. Mater.*, 2019, **31**, 3564–3572.
121 K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.
122 K. Choudhary and B. DeCost, *Npj Comput. Mater.*, 2021, **7**, 1–8.
123 B. Zhang, M. Zhou, J. Wu and F. Gao, *IEEE Access*, 2022, **10**, 62440–62449.
124 Y. Ito, T. Taniai, R. Igarashi, Y. Ushiku and K. Ono, *arXiv*, 2025, preprint, arXiv:arXiv:2503.02209, DOI: 10.48550/arXiv.2503.02209.
125 D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton and A. Agrawal, *Sci. Rep.*, 2018, **8**, 17593.
126 A. Y.-T. Wang, S. K. Kauwe, R. J. Murdock and T. D. Sparks, *Npj Comput. Mater.*, 2021, **7**, 1–10.
127 F. Xie, T. Lu, S. Meng and M. Liu, *Sci. Bull.*, 2024, **69**, 3525–3532.
128 S. Kim, J. Noh, G. H. Gu, A. Aspuru-Guzik and Y. Jung, *ACS Cent. Sci.*, 2020, **6**, 1412–1420.
129 T. M. Nguyen, S. A. Tawfik, T. Tran, S. Gupta, S. Rana and S. Venkatesh, *arXiv*, 2024, preprint, arXiv:arXiv:2411.04323, DOI: 10.48550/arXiv.2411.04323.
130 A. Hajibabaei, *Phys. Rev. B*, DOI:10.1103/PhysRevB.103.214102.
131 Y. Qiu, X. Zhang, Y. Tian and Z. Zhou, *Chin. J. Struct. Chem.*, 2023, **42**, 100118.
132 A. Hajibabaei and K. S. Kim, *J. Phys. Chem. Lett.*, 2021, **12**, 8115–8120.
133 C. G. Staacke, T. Huss, J. T. Margraf, K. Reuter and C. Scheurer, *Nanomaterials*, 2022, **12**, 2950.
134 H. Wang, L. Zhang, J. Han and W. E, *Comput. Phys. Commun.*, 2018, **228**, 178–184.
135 Y. Huang, D. Zhao, M. Deng and H. Shen, *Phys. Chem. Chem. Phys.*, 2025, **27**, 3243–3252.
136 Z. Xu, Y. Lin, Y. Xia, Y. Jiang, X. Feng, Z. Liu, L. Shen, M. Zheng and Y. Xia, *J. Power Sources*, 2025, **637**, 236591.
137 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
138 R. Jacobs, D. Morgan, S. Attarian, J. Meng, C. Shen, Z. Wu, C. Y. Xie, J. H. Yang, N. Artrith, B. Blaiszik, G. Ceder, K. Choudhary, G. Csanyi, E. D. Cubuk, B. Deng, R. Drautz, X. Fu, J. Godwin, V. Honavar, O. Isayev, A. Johansson, B. Kozinsky, S. Martiniani, S. P. Ong, I. Poltavsky, K. Schmidt, S. Takamoto, A. P. Thompson, J. Westermayr and B. M. Wood, *Curr. Opin. Solid State Mater. Sci.*, 2025, **35**, 101214.
139 J. Lee and R. Asahi, *Comput. Mater. Sci.*, 2021, **190**, 110314.
140 S. Y. Willow, A. Hajibabaei, M. Ha, D. C. Yang, C. W. Myung, S. K. Min, G. Lee and K. S. Kim, *Chem. Phys. Rev.*, 2024, **5**, 041307.
141 Q. Chen, S. Wang and C. Ling, *J. Energy Chem.*, 2026, **112**, 666–687.
142 H.-J. Lu, N. Zou, R. Jacobs, B. Afflerbach, X.-G. Lu and D. Morgan, *Comput. Mater. Sci.*, 2019, **169**, 109075.
143 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.
144 A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui and E. J. Reed, *Chem. Mater.*, 2019, **31**, 342–352.
145 R. Jaafreh, J.-G. Kim and K. Hamad, *J. Power Sources*, 2024, **606**, 234575.
146 J. Dong, W. Yang, H. Liu, J. Wu and Z. Wang, *ACS Appl. Mater. Interfaces*, 2025, **17**, 41868–41882.
147 Z. Kharbouch, M. Bouchaara, F. Elkouihen, A. Habbal, A. Ratnani and A. Faik, *Solid State Ion.*, 2024, **417**, 116713.
148 A. Maevskiy, A. Carvalho, E. Sataev, V. Turchyna, K. Noori, A. Rodin, A. H. Castro Neto and A. Ustyuzhanin, *Phys. Rev. Res.*, 2025, **7**, 023167.
149 C. Chen, Y. Zuo, W. Ye, X. Li and S. P. Ong, *Nat. Comput. Sci.*, 2021, **1**, 46–53.
150 M. Ataya, E. McCalla and R. Z. Khaliullin, *J. Phys. Chem. C*, 2024, **128**, 14149–14157.
151 N. V. Kireeva, A. Yu. Tsivadze and V. S. Pervov, *Solid State Ion.*, 2023, **399**, 116293.
152 E. Choi, J. Jo, W. Kim and K. Min, *ACS Appl. Mater. Interfaces*, 2021, **13**, 42590–42597.
153 Y. Wang, S. Li, S. Li and M. Chen, *Mater. Today Commun.*, 2024, **38**, 108294.
154 C. Chen, D. T. Nguyen, S. J. Lee, N. A. Baker, A. S. Karakoti, L. Lauw, C. Owen, K. T. Mueller, B. A. Bilodeau, V. Murugesan and M. Troyer, *J. Am. Chem. Soc.*, 2024, **146**, 20009–20018.
155 B. D. Lee, D. S. Gavali, H. Kim, S. Kim, M. Y. Cho, K. Pyo, Y.-K. Lee, W. B. Park and K.-S. Sohn, *J. Mater. Chem. A*, 2025, **13**, 10462–10474.
156 B. D. Lee, J. Shin, S. Kim, M. Y. Cho, Y.-K. Lee, M. Pyo, W. B. Park and K.-S. Sohn, *Energy Storage Mater.*, 2024, **70**, 103535.
157 R. Sewak, V. Sudarsanan and H. Kumar, *Phys. Chem. Chem. Phys.*, 2025, **27**, 3834–3843.
158 S. R. Xie, S. J. Honrao and J. W. Lawson, *Chem. Mater.*, 2024, **36**, 9320–9329.
159 Z. Wan, X. Chen, Z. Zhou, X. Zhong, X. Luo and D. Xu, *J. Energy Chem.*, 2024, **88**, 28–38.
160 Z. Wang, Y. Han, J. Cai, A. Chen and J. Li, *SmartMat*, 2023, **4**, e1183.
161 J. Cai, Z. Wang, S. Wu, Y. Han and J. Li, *Energy Storage Mater.*, 2021, **42**, 277–285.
162 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
163 L. Gigli, D. Tisi, F. Grasselli and M. Ceriotti, *Chem. Mater.*, 2024, **36**, 1482–1496.
164 J. Dai, Y. Jiang and W. Lai, *Phys. Chem. Chem. Phys.*, 2022, **24**, 15025–15033.
165 A. Seth, R. P. Kulkarni and G. Sai Gautam, *ACS Mater. Au*, DOI:10.1021/acsmaterialsau.4c00117.
166 Q. Yang, J. Xu, X. Fu, J. Lian, L. Wang, X. Gong, R. Xiao and H. Li, *J. Mater. Chem. A*, 2024, **13**, 2309–2315.
167 H. Guo, Q. Wang, A. Urban and N. Artrith, *Chem. Mater.*, 2022, **34**, 6702–6712.
168 M. Ha, A. Hajibabaei, D. Y. Kim, A. N. Singh, J. Yun, C. W. Myung and K. S. Kim, *Adv. Energy Mater.*, 2022, **12**, 2201497.
169 L. Kahle and F. Zipoli, *Phys. Rev. E*, 2022, **105**, 015311.
170 J. D. Morrow, J. L. A. Gardner and V. L. Deringer, *J. Chem. Phys.*, 2023, **158**, 121501.
171 R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse and R. Asahi, *J. Phys. Chem. Lett.*, 2020, **11**, 6946–6955.
172 J. Carrete, H. Montes-Campos, R. Wanzenböck, E. Heid and G. K. H. Madsen, *J. Chem. Phys.*, 2023, **158**, 204801.
173 J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak and B. Kozinsky, *Npj Comput. Mater.*, 2020, **6**, 20.
174 Z. Li, O. Fuhr, M. Fichtner and Z. Zhao-Karger, *Energy Environ. Sci.*, 2019, **12**, 3496–3501.
175 B. Park and J. L. Schaefer, *J. Electrochem. Soc.*, 2020, **167**, 070545.
176 Z. Hu, H. Zhang, H. Wang, F. Zhang, Q. Li and H. Li, *ACS Mater. Lett.*, 2020, **2**, 887–904.
177 L. F. O'Donnell and S. G. Greenbaum, *Batteries*, 2021, **7**, 3.
178 H. Du, J. Hui, L. Zhang and H. Wang, *arXiv*, 2025, preprint, arXiv:arXiv:2502.09970, DOI: 10.48550/arXiv.2502.09970.
179 P. Xue, R. Qiu, C. Peng, Z. Peng, K. Ding, R. Long, L. Ma and Q. Zheng, *Adv. Sci.*, 2024, **11**, 2410065.
180 B. He, S. Chi, A. Ye, P. Mi, L. Zhang, B. Pu, Z. Zou, Y. Ran, Q. Zhao, D. Wang, W. Zhang, J. Zhao, S. Adams, M. Avdeev and S. Shi, *Sci. Data*, 2020, **7**, 151.
181 X. Qu, A. Jain, N. N. Rajput, L. Cheng, Y. Zhang, S. P. Ong, M. Brafman, E. Maginn, L. A. Curtiss and K. A. Persson, DOI:10.1016/j.commatsci.2015.02.050.
182 Q. Hu, K. Chen, J. Li, T. Zhao, F. Liang and D. Xue, *Energy*, 2024, **5**, 100159.
183 S. Singh, Y. E. Ebongue, S. Rezaei and K. P. Birke, *Batteries*, 2023, **9**, 301.

184 Y. Wang, *Npj Comput. Mater.*, 2025, **11**, 1–18.
185 Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo and C. Ling, *Nat Commun*, 2019, **10**, 5260.
186 C.-T. Chen and G. X. Gu, *Adv. Sci.*, 2020, **7**, 1902607.
187 V. Gupta, K. Choudhary, F. Tavazza, C. Campbell, W. Liao, A. Choudhary and A. Agrawal, *Nat. Commun.*, 2021, **12**, 6595.
188 Z.-W. Zhao, M. del Cueto and A. Troisi, *Digit. Discov.*, **1**, 266–276.
189 X.-B. Cheng, R. Zhang, C.-Z. Zhao and Q. Zhang, *Chem. Rev.*, 2017, **117**, 10403–10473.
190 K. Kubota and S. Komaba, DOI:10.1149/2.0151514jes.
191 J. Muldoon, C. B. Bucur and T. Gregory, *Chem. Rev.*, 2014, **114**, 11683–11720.
192 S. K. Das, S. Mahapatra and H. Lahan, DOI:10.1039/C7TA00228A.
193 X. H. Liu, L. Zhong, S. Huang, S. X. Mao, T. Zhu and J. Y. Huang, *ACS Nano*, 2012, **6**, 1522–1531.
194 S. Xiang, S. Lu, J. Li, K. Xie, R. Zhu, H. Wang, K. Huang, C. Li, J. Wu, S. Chen, Y. Shen, Y. Chen and Z. Wen, *ACS Appl. Energy Mater.*, 2025, **8**, 1620–1628.
195 M. Harada, H. Takeda, S. Suzuki, K. Nakano, N. Tanibata, M. Nakayama, M. Karasuyama and I. Takeuchi, *J. Mater. Chem. A*, 2020, **8**, 15103–15109.
196 E. T. Chenebuah, M. Nganbe and A. B. Tchagang, *Front. Mater.*, DOI:10.3389/fmats.2023.1233961.
197 A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir and G. Menegaz, *Adv. Intell. Syst.*, 2025, **7**, 2400304.
198 G. P. Wellawatte and P. Schwaller, *arXiv*, 2023, preprint, arXiv:arXiv:2311.04047, DOI: 10.48550/arXiv.2311.04047.
199 S. J. Honrao, X. Yang, B. Radhakrishnan, S. Kuwata, H. Komatsu, A. Ohma, M. Sierhuis and J. W. Lawson, *Sci. Rep.*, 2021, **11**, 16484.
200 S. J. Honrao, S. R. Xie and J. W. Lawson, Boston, MA.
201 G. Bradford, J. Lopez, J. Ruza, M. A. Stolberg, R. Osterude, J. A. Johnson, R. Gomez-Bombarelli and Y. Shao-Horn, *ACS Cent. Sci.*, 2023, **9**, 206–216.
202 Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman and V. Viswanathan, *ACS Cent. Sci.*, 2018, **4**, 996–1006.
203 A. Maevskiy, A. Carvalho, E. Sataev, V. Turchyna, K. Noori, A. Rodin, A. H. C. Neto and A. Ustyuzhanin, *arXiv*, 2025, preprint, arXiv:arXiv:2411.06804, DOI: 10.48550/arXiv.2411.06804.
204 A. Ghosh, *Comput. Mater. Sci.*, 2024, **233**, 112740.
205 A. Hernandez, A. Balasubramanian, F. Yuan, S. A. M. Mason and T. Mueller, *Npj Comput. Mater.*, 2019, **5**, 112.
206 I. E. Kumar, S. Venkatasubramanian, C. Scheidegger and S. Friedler, in *Proceedings of the 37th International Conference on Machine Learning*, 2020.
207 D. Alvarez-Melis and T. S. Jaakkola, *arXiv*, 2018, preprint, arXiv:arXiv:1806.08049, DOI: 10.48550/arXiv.1806.08049.
208 C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
209 R. Ding, J. Liu, K. Hua, X. Wang, X. Zhang, M. Shao, Y. Chen and J. Chen, *Sci. Adv.*, 2025, **11**, eadr9038.
210 F. Wang, Y.-H. Tang, Z.-B. Ma, Y.-C. Jin and J. Cheng, *ChemRxiv*, 2025, preprint, DOI: 10.26434/chemrxiv-2025-fnw1w.
211 S. Zhu, B. Ramsundar, E. Annevelink, H. Lin, A. Dave, P.-W. Guan, K. Gering and V. Viswanathan, *Nat. Commun.*, 2024, **15**, 8649.
212 S. Yang, K. Cho, A. Merchant, P. Abbeel, D. Schuurmans, I. Mordatch and E. D. Cubuk, *arXiv*, 2024, preprint, arXiv:arXiv:2311.09235, DOI: 10.48550/arXiv.2311.09235.
213 C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka and T. Xie, *Nature*, 2025, **639**, 624–632.
214 D. C. Lonie and E. Zurek, *Comput. Phys. Commun.*, 2011, **182**, 372–387.
215 A. Vasylenko, J. Gamon, B. B. Duff, V. V. Gusev, L. M. Daniels, M. Zanella, J. F. Shin, P. M. Sharp, A. Morscher, R. Chen, A. R. Neale, L. J. Hardwick, J. B. Claridge, F. Blanc, M. W. Gaultois, M. S. Dyer and M. J. Rosseinsky, *Nat. Commun.*, 2021, **12**, 5561.
216 A. G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hattrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A. V. Davydov, R. Agarwal, L. A. Bendersky, M. Li, A. Mehta and I. Takeuchi, *Nat. Commun.*, 2020, **11**, 5966.
217 R. Kumar, M. C. Vu, P. Ma and C. V. Amanchukwu, *Chem. Mater.*, 2025, **37**, 2720–2734.
218 C. Chen, D. T. Nguyen, S. J. Lee, N. A. Baker, A. S. Karakoti, L. Lauw, C. Owen, K. T. Mueller, B. A. Bilodeau, V. Murugesan and M. Troyer, *J. Am. Chem. Soc.*, 2024, **146**, 20009–20018.
219 J. Noh, H. A. Doan, H. Job, L. A. Robertson, L. Zhang, R. S. Assary, K. Mueller, V. Murugesan and Y. Liang, *Nat. Commun.*, 2024, **15**, 2757.
220 W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson and G. Ceder, *Sci. Adv.*, 2016, **2**, e1600225.
221 A. M, D. Ss, S. W and P. Ka, *PubMed*.
222 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
223 N. Meddings, M. Heinrich, F. Overney, J.-S. Lee, V. Ruiz, E. Napolitano, S. Seitz, G. Hinds, R. Raccichini, M. Gaberšček and J. Park, *J. Power Sources*, 2020, **480**, 228742.
224 N. J. Szymanski, C. J. Bartel, Y. Zeng, M. Diallo, H. Kim and G. Ceder, .
225 F. Yang, E. Campos dos Santos, X. Jia, R. Sato, K. Kisu, Y. Hashimoto, S. Orimo and H. Li, *Nano Mater. Sci.*, 2024, **6**, 256–262.
226 M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, *Sci. Data*, 2016, **3**, 160018.