Predictive Modeling of Ionic Conductivity in Solid Polymer Electrolytes

Yunsong Pan

A thesis

submitted in partial fulfilment of the

requirements for the degree of

MASTER OF SCIENCE IN APPLIED CHEMICAL SCIENCE AND TECHNOLOGY

University of Washington

2024

Committee:

Ting Cao

Luna Huang

Robert Synovec

Program Authorized to Offer Degree:

Department of Chemistry

University of Washington

Abstract

Predictive Modeling of Ionic Conductivity in Solid Polymer Electrolytes

Yunsong Pan

Chair of the Supervisory Committee:

Robert Synovec

Department of Chemistry

Solid polymer electrolytes (SPEs) hold significant potential for energy storage systems, particularly in next-generation solid-state lithium-ion batteries. However, their poor ionic conductivity remains a critical challenge. This thesis investigates the relationship between the microstructural characteristics of SPEs and their ionic conductivity using machine learning techniques. Our goal is to utilize existing literature data to develop predictive models that forecast ionic conductivity based on structural attributes.

We begin with a review of the fundamental properties of SPEs, focusing on ionic transport mechanisms from both theoretical and experimental perspectives. Our methodology involves extracting and synthesizing data from numerous studies to create a comprehensive dataset. Machine learning models, including neural networks, are then trained on this dataset to predict ionic conductivity.

The results highlight significant correlations between specific structural features and ionic conductivity. However, the performance of our predictive models was limited by the small dataset size, leading to potential overfitting and reduced generalizability. Despite these limitations, the study demonstrates the feasibility of using machine learning to gain insights into SPE design.

This research advances our understanding of the factors influencing ionic conductivity in SPEs and underscores the need for larger datasets. It also highlights the potential of combining traditional scientific approaches with modern data science techniques to drive materials innovation.

Acknowledgements

I would like to express my profound gratitude to all those who have made the completion of this thesis possible.

Firstly, I extend my deepest appreciation to my thesis advisors, Dr. Ting Cao and Dr. Luna Huang, whose expertise, understanding, and patience, added considerably to my graduate experience. Their willingness to give time generously has been very much appreciated. I am also grateful to my thesis committee member, Dr. Robert Synovec, for his insightful comments and encouragement.

A special word of appreciation goes to my project members, Aashish and Jonathan. Discussing my ideas with you and receiving your feedback was immensely helpful and provided a sense of community and intellectual camaraderie during my research journey.

Lastly, I would like to thank my parents for their unwavering support, both emotionally and financially throughout my studies. This accomplishment would not have been possible without them.

Table of Contents

List of Figures……...…….……6

List of Tables……...…...……...…7

Chapter 1: Introduction……...…..…8

Chapter 2: Literature Review……...……...……10

2.1 Evolution and Optimization of Solid Polymer Electrolytes…….……10

2.2 Ionic Conductivity of Solid Polymer Electrolytes……...……14

2.2.1 Physical Models for Ionic Conductivity……...……14

2.2.2 Methods to Improve Ionic Conductivity……...16

2.3 Computational Methods and Machine Learning for Polymer Research…..18

2.3.1 Traditional Computational Methods… ……...…….18

2.3.2 Machine Learning for Polymer Research…….……..…..19

Chapter 3: Methodology……...……...……...……23

3.1 Introduction……..23

3.2 Data Collection and Preparation……...…….24

3.2.1 Data Sources and Acquisition……...25

3.2.2 Data Cleaning…….…..28

3.2.3 Feature Engineering……..……29

3.2.4 Molecular Representation……...……..30

3.3 Model Development…….……33

Chapter 4: Results and Discussion……..35

4.1 Model Performance and Validation……..……35

4.2 Limitations and Implications of the Study……..…..39

Chapter 5: Conclusions…… ...41

5.1 Summary of Findings…….……..41

5.2 Contributions to the Field…….…41

5.3 Future Directions…….….42

References……...…….44

Appendix…….…….50

List of Figures

Figure 2.1. Schematic illustration of (a) the morphology of a polymer consisting of amorphous and crystalline regions; (b) the structure of a polymer chain……...…….…..11

Figure 2.2. Schematic illustration of the lithium-ion's segmental motion in SPEs…...13

Figure 2.3. Approaches to improve the ionic conductivity of SPE…….…..13

Figure 2.4. Arrhenius plot for the PVDF-HFP/SG electrolyte…….…….16

Figure 2.5. An example of branched polymer: star-shaped siloxane acrylate (D4A)…17

Figure 2.6. An example of interpenetrating network polymer: LiBAMB -PETMP…..17

Figure 2.7. Overview of ML algorithms……..…….20

Figure 2.8. Three types of feature-extracting methods for SMILES……...…….22

Figure 3.1. Ionic conductivity for SPEs with different backbone chemical structures..27

Figure 3.2. Schematic illustration for the classification for SPEs…….……28

Figure 3.3. Example of molecular structure conversion to SMILES using MolScribe.31

Figure 3.4. Example of polymer structure recreated in ChemDraw…….…….32

Figure 4.1. R^2 plot for training and testing dataset using CNN with Morgan fingerprint features……. .35

Figure 4.2. Comparison of actual and predicted ionic conductivity using CNN with Morgan fingerprint features…….……..36

Figure 4.3. Model training and validation loss for CNN with Morgan fingerprint features…….…….37

Figure 4.4. R^2 plot for training and testing dataset using CNN with Mol2vec features.38

Figure 4.5. Model training and validation loss for CNN with Mol2vec features…….39

List of Tables

Table 3.1. Publicly Accessible Polymer Databases Examined…….……...….27

Table 3.2. Categorization of Polymer Properties and Their Correlation with Ionic Conductivity…….….28

Table 3.3. Tools Explored for Molecular Structure Conversion to SMILES……..…33

Chapter 1: Introduction

Developing renewable energy to replace nonrenewable energy such as fossil fuels has been regarded as one of the most important global challenges. Lithium batteries, such as lithium-ion batteries (LIB) have been deemed as next-generation energy storage system due to their high energy density (energy per unit volume) and specific energy (energy per unit mass), low maintenance cost, and good electrochemical potential.[1,2] However, the liquid electrolytes used in current LIB contain highly volatile and flammable solvents, such as ethylene carbonate, which cause fire and explosion concerns.[3] In addition, liquid electrolytes are unable to withstand high operational voltages, which would lead to unstable solid electrolyte interphase (SEI), causing harmful lithium dendrites.[4] This limitation is particularly problematic in next-generation lithium-metal anode batteries (LMB), which require effective dendrite suppression. Lithium-metal anodes offer exceptionally high capacity (3860 mAh g^{-1}), over ten times that of current graphite anodes (372 mAh g^{-1}),[5] underscoring the need for advanced electrolyte materials.

During the past decades, several types of Li-ion conducting electrolytes have been suggested. Generally, they can be classified into two categories: inorganic ceramic electrolytes and solid polymer electrolytes (SPEs). Compared with ceramic electrolytes, SPEs have lower cost of manufacturing and better interfacial compatibility. SPEs enhance battery safety and durability through their solid-state properties, reducing risks associated with dendrite formation, flammability, and chemical instability.[6] The integration of SPEs aligns with industry trends that prioritize safety and performance, especially in sectors like consumer electronics and electric vehicles, where battery reliability is crucial. Furthermore, SPEs' ability to suppress dendrite formation makes them suitable for next-generation LMB, thereby improving the energy density of LIBs.

The core of SPEs lies in their unique ability to conduct lithium ions through a solid polymer matrix. This conduction is typically achieved through the incorporation of a lithium salt into a polymer host, such as polyethylene oxide (PEO). The ionic conductivity in these systems occurs via the movement of lithium ions along the

polymer chains, facilitated by their interaction with the electronegative (e.g. oxygen) atoms in the polymer chain, known as segmental motion.[7] By optimizing these interactions and the polymer's molecular structure, researchers aim to enhance the ionic conductivity, which is key to improving the overall battery performance.

Despite these advantages, the adoption of SPEs in LIBs faces challenges, primarily related to their low ionic conductivity at ambient temperatures.[8] Traditionally, the relationship between the molecular structure of polymers and their properties has been explored through experimental methods and classical computational chemistry techniques such as density functional theory (DFT).[9,10] However, these methods are often time-consuming and computationally intensive, limiting their utility in rapidly screening and optimizing large datasets of polymer structures.

Recent advances in machine learning (ML) provide powerful tools for predicting material properties based on structural data. ML algorithms excel at uncovering complex patterns in data that are not readily apparent through conventional analysis.[11] For SPEs, ML can be applied to predict ionic conductivities from the SMILES (Simplified Molecular Input Line Entry System) notations of polymers, offering a rapid and cost-effective method for material design and selection. By training models on datasets that comprise SMILES notations and known conductivities, ML facilitates the discovery of novel polymer electrolytes with tailored properties.

Chapter 2: Literature Review

2.1 Evolution and Optimization of Solid Polymer Electrolytes

In 1973, Fenton et al.[12] reported the semicrystalline complex of polyethylene oxide (PEO) and certain sodium and potassium salts. The electrical properties were studied later, and they highlighted the ionic conduction in the amorphous phase of the polymer (**Fig. 2.1**).[13] Armand later proposed using these Li-ion conducting systems as SPEs for solid-state LMBs, and he particularly recommended weak coordinating anions like bis(trifluoromethanesulfonyl)imide (TFSI) as the ideal lithium salt anions. This proposition has since been validated, especially in guiding the selection of anions to develop practical SPEs for LMBs, sparking extensive global research over the past four decades. Despite initial skepticism, SPEs are now recognized for potentially resolving safety concerns, particularly by mitigating Li dendrite growth common in traditional liquid electrolytes and enhancing energy density due to their advantages including cost-effectiveness and non-flammability.

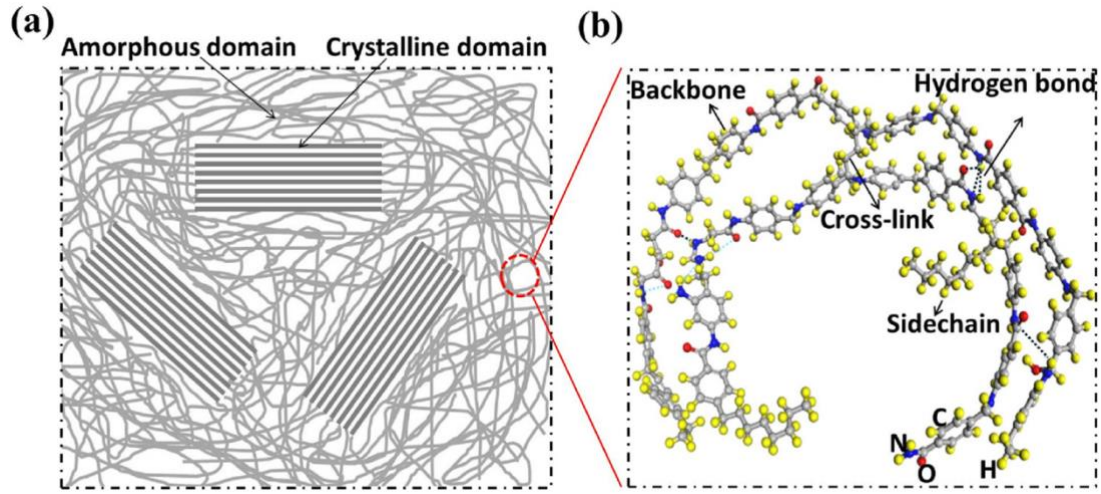

Figure 2.1. Schematic illustration of (a) the morphology of a polymer consisting of amorphous and crystalline regions; (b) the structure of a polymer chain.[14]

A significant advancement in this area is exemplified by the electric vehicle Autolib, which utilizes an LMB comprising a SPE of lithium bis(trifluoromethanesulfonyl)imide (LiTFSI) and PEO.[15] Despite the widespread adoption of SPE, especially PEO based material in such applications, the material

faces several significant challenges, especially their low mechanical strength and low ionic conductivity.

The mechanical strength of SPEs is crucial in suppressing dendrite growth, a common failure mechanism in LIBs. Dendrites form when lithium deposits unevenly during the charging process, eventually piercing the electrolyte layer and leading to short circuits or even battery fires.[16] SPEs often suffer from insufficient mechanical robustness to withstand the mechanical stresses caused by dendrite formation.[17] Monroe and Newman suggested that dendrite growth can be successfully suppressed if the shear modulus of SPE is at least twice of that of Li metal (4.8 GPa at 298 K).[18] Enhancing the mechanical strength of these electrolytes is therefore essential, not only to prevent dendrite growth but also to maintain the integrity and longevity of the battery under typical operating conditions. Recent research has focused on developing composite SPEs by incorporating rigid fillers (e.g. silica nanoparticles), organic fillers (crosslinking) or robust supporting framework,[19–21] which can significantly improve their mechanical properties and their dendrite suppression capabilities.

Ionic conductivity of the electrolyte is fundamental to the efficiency, energy density, and stability of LIB. The low ionic conductivity of SPEs represents a significant barrier to their widespread adoption in next-generation LMBs. This limitation is primarily due to the semi-crystalline nature of the polymer matrix, which impedes the free movement of lithium ions.[22] Within the polymer, lithium ions migrate mainly through a process known as segmental motion (**Fig. 2.2)**.[23] This involves the coordination of lithium ions with electronegative functional groups, followed by their migration to new sites where they coordinate with other functional groups. This movement is intrinsically linked to the local motion of the polymer chain. Therefore, the efficiency of segmental motion is closely linked to the flexibility of the polymer chain and the coordinating environment. A high degree of crystallinity results in a more ordered matrix structure, which restricts the ability of the polymer chains to deform and effectively coordinate lithium ions and vice versa. This challenge becomes more pronounced at lower temperatures, where decreased polymer flexibility leads to further reductions in ion mobility and, consequently, ionic

conductivity. Moreover, the coordinating properties of the functional groups play a crucial role, as ion migration along or between polymer chains to new coordination environments involves the exchange of coordinating groups within the ion's solvation shell.[24]

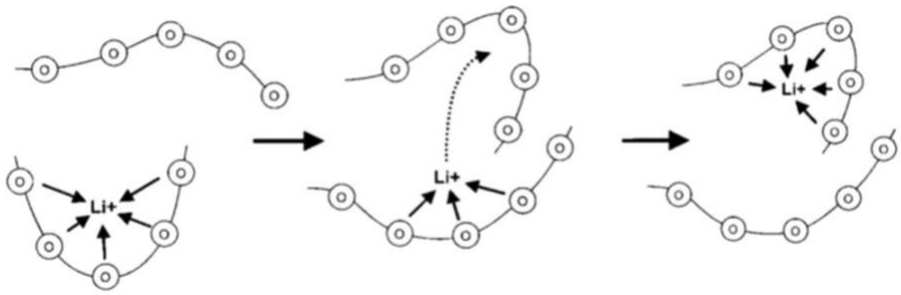

Figure 2.2. Schematic illustration of the lithium-ion's segmental motion in SPEs.[22]

To overcome the challenges associated with the low ionic conductivity of SPEs, extensive research has focused on modifying the polymer structure to enhance ion transport. Several strategies have proven effective in improving the ionic conductivity of SPEs, which can be summarized to three ways: increase ion-pair dissociation, create ionic conduction pathway, and decrease crystallinity (**Fig. 2.3**).[25]

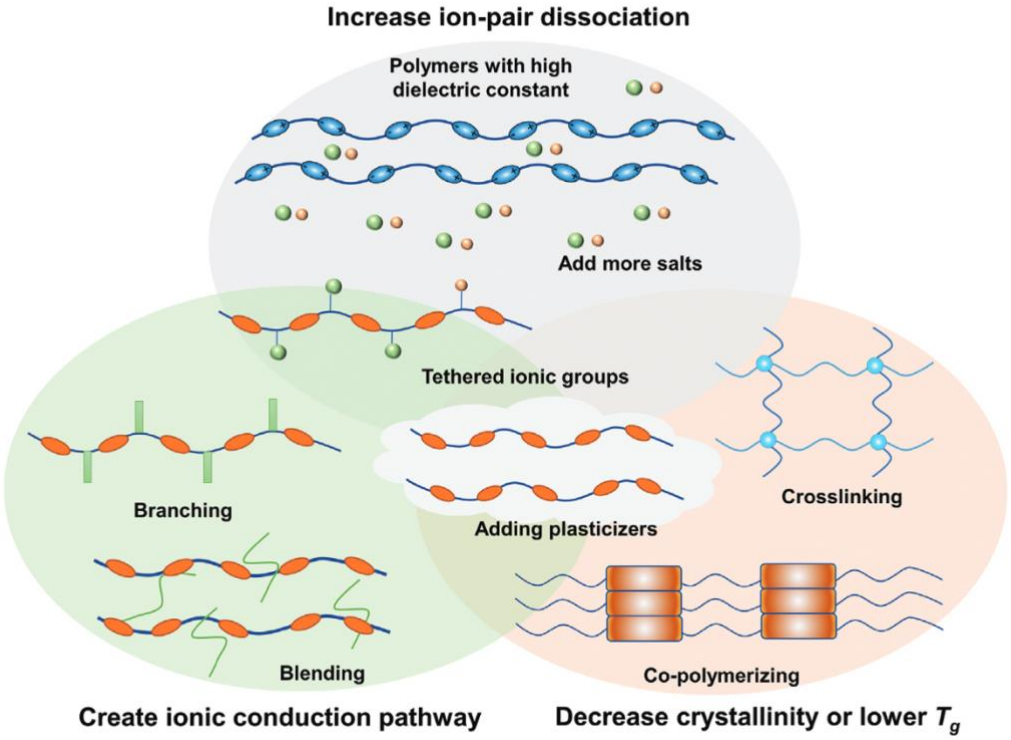

Figure 2.3. Approaches to improve the ionic conductivity of SPE.[25]

Besides ionic conductivity, SPEs formed by dissolving lithium salts with small anions in a polymer host (mostly PEO) have low lithium transference number (LTN) because they are dual-ion conductors, where both Li^+ cations and their counter anions are mobile. Li^+ cations are typically less mobile than their anionic counterparts, as their movement is hindered by the Lewis basic atoms in the polymer matrix. This results in a low LTN, generally less than 0.5.[26] During battery charge-discharge cycles, Li^+ cations and counter anions migrate in opposite directions within the polymer matrix, often leading to anion accumulation at the anode, creating concentration gradients that cause polarization, voltage losses, increased internal impedance, and undesirable reactions, ultimately leading to dendrite formation and cell failure.[27] Conversely, research has shown that when the LTN of SPE approaches 1, the electrolyte prevents the formation of concentration gradients even under high charge and discharge rates.[28] Additionally, the risk of Li dendrite formation is significantly reduced.[29] Moreover, studies have found that an electrolyte with a LTN of 1 can deliver performance on par with dual-ion electrolytes with ionic conductivity that is ten times higher.[28]

Therefore, the pursuit of single-ion solid polymer electrolytes (SISPEs) emerges as a promising direction to address these challenges. SISPEs are designed to facilitate the exclusive transport of lithium ions while immobilizing the anions, thereby potentially achieving a LTN close to 1. In the early 1980s, Bannister et al. proposed two approaches for achieving single Li-ion conduction in SISPEs.[30] The first method involves anchoring the anion to a polymer side chain, which immobilizes the anion and only allow the Li^+ cations to contribute to charge flow. The second method restrict the mobility of the anion by using a lithium salt of a dibasic acid, specifically dilithium hexafluoroglutarate, thereby enhancing the LTN.[30] The selective ion transport in SISPEs minimizes the migration of anions towards the anode, substantially reducing the formation of concentration gradients, voltage losses, and the associated internal impedance. SISPEs not only enhance the uniformity of ion distribution throughout the electrolyte but also decrease the likelihood of dendrite formation and the resultant cell failure. The successful development of SISPEs could

thus revolutionize the performance and safety of LIB systems, positioning them as a viable candidate for high-efficiency, high-safety energy storage solutions.

2.2 Ionic Conductivity of Solid Polymer Electrolytes

SPEs have garnered significant attention in the field of electrochemical devices due to their potential applications in high-performance LIBs. As introduced in ***Chapter 2.1***, ionic conductivity is a pivotal property of SPEs, influencing their efficiency and applicability in LIBs. This section provides an overview of the ionic conductivities of SPEs, focusing on their mechanisms, influencing factors, and comparative performance with traditional liquid electrolytes.

2.2.1 Physical Models for Ionic Conductivity

A distinctive characteristic of ion conducting SPEs compared to other ionic conductors is their composition, which involves dissolving salts with low lattice energy into a polar polymer matrix. Despite the fact that the conduction mechanism has not been fully understood yet, it is widely accepted that SPEs consist of both amorphous and crystalline phases at room temperature, with ion transport predominantly occurring within the amorphous regions; long-range transport of cations occurs through coordination-decoordination steps. Under an electrical field, these steps involve the movement of cations between adjacent coordination sites located either on the same polymer chain or from one chain to another chain (**Fig. 2.2)**.[31] Some reviews suggest that the polymer chains form cylindrical tunnels that act as channels for cation movement, with cations positioned and coordinated within these tunnels by the functional groups.[32,33] Therefore, the amount of Li^+ and the mobility of polymer chain is critical on the ion transport capability in SPEs. This model is known as the segmental motion model, and the ionic conductivity can be denoted as (**Eq. 2.1**):[30]

$$\sigma = \sum n_i q_i \mu_i \qquad \textbf{Equation 2.1}$$

where n_i is the concentration of carriers, q_i is the charge number of mobile ions, and μ_i is the mobility of carriers. Incorporation of polar functional groups such as -O-, C=O, -N-, and -S- into the polymer matrix is crucial as it enhances the dissociation of lithium salts, thus increasing the concentration of mobile Li^+ ions.[34] This is further supported by the choice of lithium salts, which should ideally possess low lattice energy to ease dissociation, and polymers with a high dielectric constant to increase ion mobility.[35] It can also be seen from the equation that if the charge q_i is fixed, the ionic conductivity σ is proportional to the number and mobility of mobile ions. Therefore, decreasing the polymer crystallinity can increase the concentration and mobility of the carriers (both cations and anions) and thereby increase ionic conductivity. However, while strong coordination between Li^+ ions and the polar groups promotes the dissociation of lithium salts, it can also impede the mobility of Li^+ ions.[36] Therefore, optimal ionic conductivity requires a balanced interaction between the polymer and the Li^+ ions.

Several other theoretical models, such as Arrhenius, Vogel-Tamman-Fulcher (VTF) and free volume have been proposed to interpret the mechanism of ion transport. These models provide a framework for analyzing the behavior of ion transport, which is critical for design and preparation of high-performance SPEs. However, only the Arrhenius model will be introduced further in this research due to its versatility in handling the data.

The Arrhenius model, originally developed to describe chemical reaction rates, has been effectively adapted to explain how ionic conductivity in SPEs varies with temperature. According to the Arrhenius equation, the ionic conductivity is expressed as (**Eq. 2.2**):[37]

$$\sigma(T) = \sigma_0 \exp\left(\frac{-E_a}{k_B T}\right) \qquad \textbf{Equation 2.2}$$

where σ_0, E_a and k_B are pre-exponential factor, activation energy and Boltzmann constant, respectively. This equation illustrates that an increase in temperature or a decrease in activation energy significantly enhances the ionic conductivity (**Fig. 2.4**). As temperature rises, the amplitude and frequency of motion of the polymer backbone

and side chains also increase, facilitating tighter coordination.[37] This enhanced mobility results in the segmental motion of the amorphous polymer becoming increasingly independent from the motion of the ions, thereby reducing the energy required for ions to transition from one site to another.

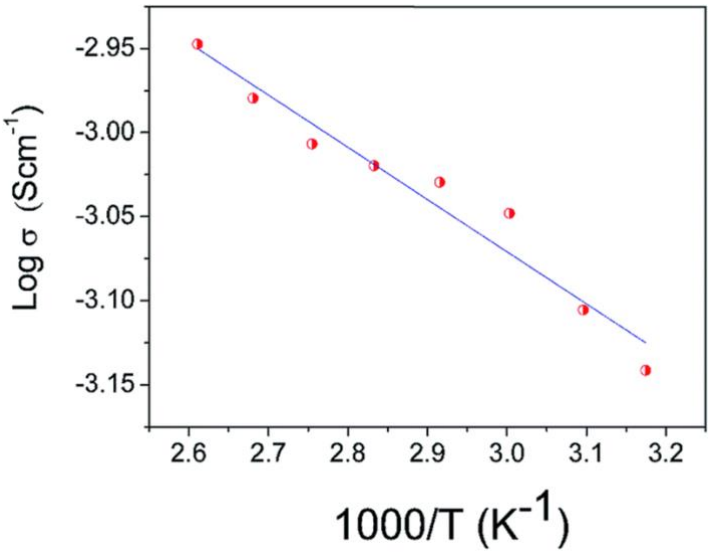

Figure 2.4. Arrhenius plot for the PVDF-HFP/SG electrolyte.[38]

2.2.2 Methods for Structural Modification to Improve Ionic Conductivity

Chapter 2.1 outlines three mechanisms to improve ionic conductivity: increase ion-pair dissociation, create ionic conduction pathway, and decrease crystallinity. Research in these areas has identified several strategies to improve the ionic conductivity of SPEs, including blending, copolymerization, branching, and interpenetrating networks:[7]

1. Blending: This technique involves physically mixing two or more polymers to create the SPE, such as adding a plasticizer. Interactions between the polymers reduce their crystallinity, thereby enhancing ionic conductivity. This method also allows the electrolyte to inherit properties from various polymer hosts, providing a simple yet effective means to tailor electrolyte characteristics.[39]
2. Copolymerization: SPEs created through the covalent bonding of polymers with distinct properties exhibit varied functional segments. For instance, SPEs combining vinyl ethylene carbonate (VEC) and hydroxyethyl methacrylate

(HEMA) have been developed. Huang et al. reported an SPE comprising VEC-HEMA and LiTFSI, which demonstrated a high Young's modulus (0.5 GPa) from HEMA and increased electron conductivity (8×10^{-4} S cm^{-1}) from VEC. Such SPEs are also effective in suppressing dendrite growth in polymer lithium-ion batteries.[40]

3. Branching (**Fig. 2.6**): Incorporating a large number of branched chains within the polymer can significantly reduce crystallinity, thus enhancing ionic conductivity. Kang et al. synthesized a star-shaped siloxane acrylate SPE with various oligo(ethylene oxide) repeating units, achieving ionic conductivities in the range of 6.3 to 7.8×10^{-4} S cm^{-1} with 80 wt% poly(ethylene oxide) dimethyl ether at 30°C. The ionic conductivity of these branched SPEs is comparable to that of SPEs based on amorphous linear PEO.[41]

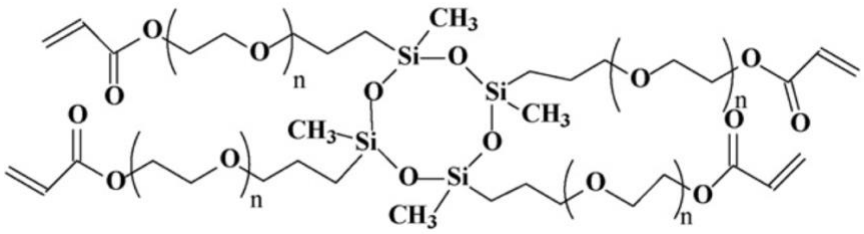

Figure 2.5. An example of branched polymer: star-shaped siloxane acrylate (D4A).[41]

4. Interpenetrating network (**Fig. 2.7**): This method SPEs with cross-linked network structures can partially inhibit polymer crystallization, improving both the mechanical and conductive properties of the electrolyte. Deng et al. prepared a new network boron-based single-ion SPE that has an ionic conductivity of 1.47×10^{-3} S cm^{-1} at room temperature.[26]

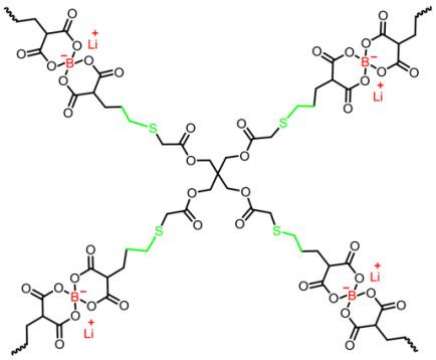

Figure 2.6. An example of interpenetrating network polymer: lithium bis(allylmalonato)borate (LiBAMB) - pentaerythritol tetrakis(2-mercaptoacetate) (PETMP).[26]

2.3 Computational Methods and Machine Learning for Polymer Research

The ionic conductivity of SPEs is heavily influenced by their molecular composition and structural arrangement. Accurately predicting and improving the ionic conductivity through structural modifications not only optimizes the performance of existing SPEs but also aids in the development of new formulations for SPEs.

2.3.1 Traditional Computational Methods

In the field of polymer science, computational methods such as Molecular Dynamics (MD), Monte Carlo (MC) simulations and Density Functional Theory (DFT) have traditionally played pivotal roles in advancing the understanding of polymer structures and properties.[42] For example, MD offers insights into polymer dynamics, such as ionic conductivity in stretched Nafion and molecular dynamics of polymer chain in solution by simulating the physical movements of molecules.[43,44] Meanwhile, DFT has been utilized in analyzing electronic structures and understanding the bonding mechanisms within polymers. DFT helps predict properties such as electronic, optical and mechanical properties of polymers, making it crucial for the design of new polymer materials for electronic, photonic and energy applications.[45,46] Both methods have significantly contributed to the foundational knowledge in polymer science, paving the way for the development and optimization of innovative polymer materials.

However, these traditional computational methods face several challenges. One significant challenge is the computational cost associated with these techniques, which increases exponentially with the size of the polymer system and the duration of the simulated processes. This makes it difficult to model the long-time scales and large spatial scales that are often necessary to capture the full range of polymer

behaviors, particularly for complex or highly entangled polymers.[47] Furthermore, MD and MC simulations rely on accurate force fields, which can be difficult to parameterize for polymers, especially when dealing with novel or poorly understood materials. These methods also generally assume equilibrium conditions, which might not be representative of all real-world scenarios where non-equilibrium dynamics play a critical role.[48] Consequently, while MD and MC provide valuable insights into the microscopic properties and behaviors of polymers, their applicability is sometimes limited by scale, computational resources, and the accuracy of the underlying physical models.

2.3.2 Machine Learning for Polymer Research

In contrast to biomolecular research, which benefits from extensive, homogeneous databases like the Protein Data Bank, polymer research faces challenges due to the lack of similar comprehensive databases.[49] These databases are often sparse, heterogeneous, and sometimes unavailable, compounded by the absence of rapid and scalable techniques for measuring polymer properties.[50] This makes constructing new databases for specific design problems resource-intensive and difficult.

However, machine learning (ML) offers potential solutions to these challenges and accelerates the design of polymers. ML, a key component of artificial intelligence (AI), evolves from traditional AI that relied on hard-coded algorithms. Unlike its predecessor, ML allows computers to derive rules directly from data. This data typically includes input values paired with corresponding outputs. The ML model learns by adjusting its internal parameters to map inputs to outputs effectively. Once trained, the model can predict outputs for new data. ML encompasses three main types: supervised learning, unsupervised learning, and reinforcement learning (**Fig. 2.8**).[51] Supervised learning involves training a model on a labeled dataset (input and desired output values) to achieve data classification or regression, such as support vector machine. Meanwhile, unsupervised learning learns patterns exclusively from unlabeled data to achieve clustering or dimension reduction. Reinforcement learning

involves an agent taking actions to alter its state and interact with the environment, aiming to maximize a target reward, such as Markov decision process.

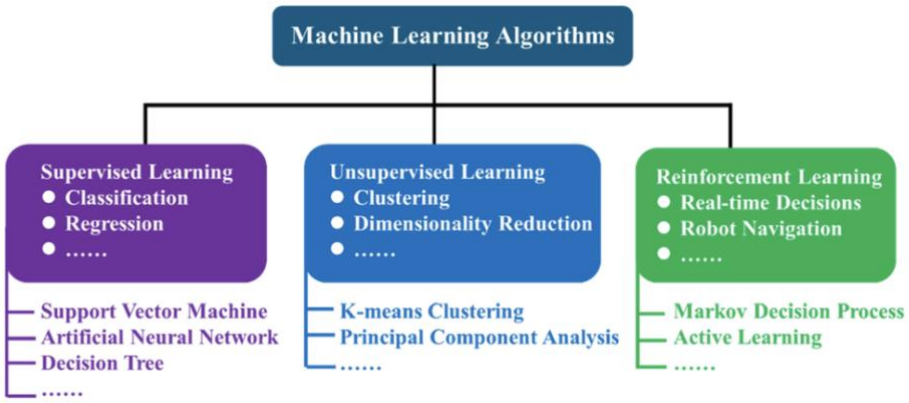

Figure 2.7. Overview of ML algorithms.[51]

The rise of ML based data-driven materials design strategies has proven successful across various materials science domains, leading to the creation of large materials databases that facilitate this new research paradigm.[11,52] Initially focused on inorganic materials and small organic molecules, these ML approaches have gained traction in polymer and soft matter research.[51] ML has been successfully applied to predict the structure-property relationship of polymers, with several databases now available that predominantly contain computationally estimated structure-property data. [51,53–55]

ML of polymers involves polymer featurization, which is representing polymers in a machine-readable numeral representation. In this regard, one of the most popular descriptors is the SMILES (Simplified Molecular Input Line Entry System) string representation, which is frequently used to explore chemical spaces in polymer research.[56] The SMILES string can be further processed to extract features, which mainly has three feature extraction methods: molecular descriptors (MD), molecular fingerprint (MF), and molecular graph (MG) (**Fig. 2.9**).[55,57]

MD are numerical values that quantify specific properties or features of a molecule. These properties can be derived from various aspects of the molecule, such as its constitutional, geometrical, or topological characteristics. Constitutional descriptors represent basic atom-based information, such as atom type, atomic weight,

number of atoms, and types of bonds within a molecule. These descriptors provide fundamental insights into the molecule's composition and structure. Geometrical descriptors capture 3D structural information like shape, volume, and surface area, derived from the atomic coordinates in three-dimensional space. These descriptors are particularly useful for differentiating monomers with the same chemical composition but different 3D structures, which is crucial for understanding how shape and structural changes affect polymer properties. However, calculating geometrical descriptors can be computationally intensive, especially for large molecules or complex structures. Topological descriptors focus on the 2D connectivity and arrangement of atoms, representing the molecule as a graph where vertices correspond to atoms and edges to bonds. These descriptors encode the molecule's shape, size, connection types, and bonds, providing a comprehensive view of the molecule's 2D structure.[58]

A subset of molecular descriptors, molecular embedding (ME), involves generating continuous-value vectors from a molecule's structure. Pretrained machine learning models, such as Mol2vec, are used to create these embeddings, which avoid issues like hash collisions that can occur with other descriptor types.[59] In our studies, we chose Mol2vec because it efficiently captures complex molecular features in a continuous vector space, enhancing the model's ability to learn and generalize from the data.

Fingerprints are simplified binary or numerical vectors that encode the presence or absence, and sometimes the count, of predefined structural features or substructures within a molecule. Morgan fingerprint (MF) is one of the most widely used mathematical representations for polymer molecules. In the MF method, substructures surrounding each non-hydrogen atom are identified within a specified radius and converted into unique identifiers. These identifiers are then hashed into a fixed-length vector format, where each element indicates the presence (as a binary value) or frequency (as a count) of a particular substructure. The advantage of MF lies in its ability to capture detailed substructural information in a compact form, though it is prone to hash collisions. In our studies, we also chose MF because of its compactness.

The molecular graph (MG) representation treats the molecule as a graph where atoms are vertices and bonds are edges. This method captures the 2D connectivity and arrangement of atoms, providing a detailed depiction of the molecule's structure. MG representations are particularly effective in modeling the properties of polymers, as they consider the entire connectivity pattern within the molecule.

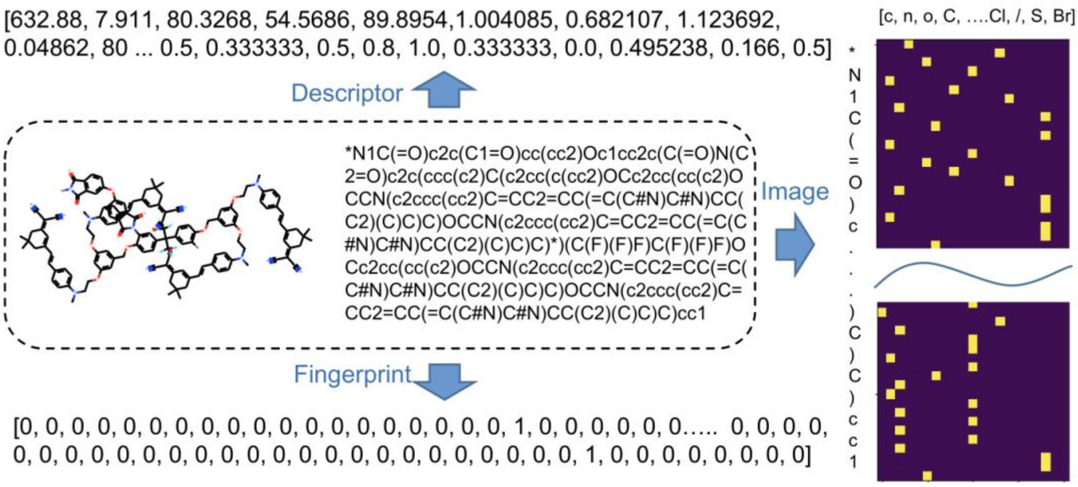

Figure 2.8. Three types of feature extracting methods for SMILES.[55]

Chapter 3: Methodology

3.1 Introduction

The investigation into the relationship between the microstructural characteristics of SPEs and their ionic conductivity is pivotal for advancing energy storage technologies. This chapter details the methodological framework employed to explore this relationship using machine learning (ML) techniques. Our approach encompasses several key stages: data acquisition and preprocessing, model selection and development, and the subsequent training, validation, and evaluation of the models. By adopting these techniques, we aim to predict ionic conductivity and elucidate the underlying factors that influence it, ultimately contributing to the optimization of SPE design and performance.

Data acquisition forms the bedrock of our analysis. We aggregated a comprehensive dataset from various sources, including peer-reviewed publications and public databases such as PoLyInfo. This dataset encompasses a wide range of features, including polymer names, SMILES strings, activation energy, and temperature, all of which are known to impact ionic conductivity. However, integrating data from disparate sources posed significant challenges, particularly due to inconsistencies in data formats and missing information. To address these issues, we implemented a multi-step preprocessing strategy. This involved standardizing data formats to ensure uniformity, managing missing values through imputation or exclusion, and scaling features to a common range to facilitate accurate model training.

In developing our models, we focused on selecting machine learning algorithms that balance complexity and interpretability. We evaluated a spectrum of algorithms, from traditional linear regression models to advanced ensemble methods and neural networks. The selection process was informed by the specific characteristics of our dataset and the nature of the predictive task. A systematic hyperparameter tuning process, employing grid search and cross-validation techniques, was used to optimize model performance. This rigorous approach allowed us to fine-tune the models,

striking an optimal balance between underfitting and overfitting, and ensuring robust predictive accuracy and generalization capability.

The training phase involved feeding the prepared data into the chosen models and iteratively adjusting model parameters to minimize prediction errors. A portion of the dataset was reserved for training, while the remainder was used for validation. Throughout the training process, model performance was closely monitored to ensure that the learned patterns generalized well to unseen data. Validation is a critical component, providing an assessment of model performance in conditions that approximate real-world scenarios. This step is crucial for fine-tuning the models and determining their predictive accuracy on new data.

Model evaluation focused on understanding how well the developed models predicted ionic conductivity and what these predictions revealed about the relationship between structural attributes and conductivity. We employed a range of performance metrics, including mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R^2) scores, to gauge model accuracy. Comparative analysis was conducted to identify the strengths and weaknesses of each model, providing insights into the suitability of different machine learning techniques for this research.

In summary, this chapter outlines the structured approach we adopted to investigate and predict the structure-conductivity relationship of SPEs using machine learning. The subsequent sections will delve deeper into each aspect of our methodology, detailing the specific strategies employed and the results obtained. By rigorously applying these techniques, we aim to contribute to the scientific understanding and practical optimization of SPEs, enhancing their performance for advanced energy storage applications.

3.2 Data Collection and Preparation

Accurate data collection and preparation are essential for building robust predictive models. For this study, the goal was to collect comprehensive datasets capturing the structure-ionic conductivity relationship of SPEs. These datasets were sourced from both proprietary polymer databases and scientific literature. The

following sections outline the challenges encountered in data acquisition and the methods used to resolve them, as well as the strategies adopted to preprocess the data and ensure the resulting dataset is suitable for machine learning.

3.2.1 Data Sources and Acquisition

An initial exploration of databases identified several sources relevant to the study, as shown in **Table 3.1**. However, most of these databases were extremely limited in size and lacked sufficient information for polymers. PoLyInfo, the largest databases for polymers available, contains data on polymer properties for more than 20,000 polymers, which could be potentially correlated with ionic conductivity.[60] While PoLyInfo includes extensive data on various properties of polymers, it lacks specific information on ionic conductivity. Therefore, supplementary data from literature were required to obtain conductivity measurements along with other relevant properties such as ionic conductivity and activation energy.

Table 3.1. Publicly Accessible Polymer Databases Examined

Name	**Description**	**URL**
Materials Project[61]	Computed properties for mainly inorganic materials	https://materialsproject.org
Polymer Genome[62]	Polymer property prediction	https://www.polymergenome.org
PoLyInfo[60]	Comprehensive data for polymers	https://polymer.nims.go.jp
Polymer Property Predictor and Database	Flory–Huggins χ parameters and glass transition temperatures for polymers	https://pppdb.uchicago.edu
Physical Properties of Polymers	Physical properties and characterization techniques of polymers	ISBN: 9780387690025, 0387690026

Nevertheless, an attempt to combine non-ionic conductivity data from PoLyInfo with literature data on ionic conductivity revealed that the polymers referenced in

literature often did not match those available in PoLyInfo. This discrepancy made direct integration challenging.

Despite the limitations, PoLyInfo contained valuable information on polymer characteristics, such as thermal properties and mechanical properties. Therefore, these data were retained and analyzed to understand their correlation with ionic conductivity. To facilitate understanding and interpretation, the correlations between these properties and ionic conductivity were categorized into five distinct groups, each represented by a different color code. These categorizations are detailed in ***Appendix*** - **Table 1** as follows:

Table 3.2. Categorization of Polymer Properties and Their Correlation with Ionic Conductivity

Color Code	Description	Relationship with Ionic Conductivity
Green	Directly supported by literature	Strong and clear relationship
Yellow	Supported by scientific reasoning and/or indirect literature	Significant relationship, albeit less direct
Orange	Supported by scientific reasoning and/or indirect literature	Weaker and less direct relationship
Red	Little relationship or influenced by too many factors	Not suitable for straightforward ML applications
Pink	Requires further investigation	Relationship unclear, needs more research

Given the limitations of existing polymer databases such as PoLyInfo, which primarily provides non-ionic conductivity related data and often lacks entries found in newer research publications, our data collection was predominantly literature-based. We meticulously reviewed recent academic papers and theses focusing on SPEs, extracting data on ionic conductivity, activation energy, temperature, lithium transference number and other relevant physical properties of polymers such as glass transition temperature. To ensure the purity of the dataset in terms of chemical composition, we specifically excluded studies involving polymers enhanced with additives like plasticizers, solvents, and inorganic or organic fillings. The rationale was to isolate the effect of the polymer backbone's chemical structure on ionic

conductivity, thereby avoiding the confounding influences of external enhancement agents.

During the data collection phase, we classified SPEs into three categories based on their backbone structures: linear, branched, and network (**Fig. 3.1**, **Fig. 3.2**). This classification allowed us to investigate the impact of structural complexity on ionic conductivity. As shown in **Fig. 3.1**, there is a clear trend indicating that network structures generally exhibit higher ionic conductivities compared to linear and branched structures. Linear polymers, represented by the light blue points, predominantly occupy the lower conductivity range, while branched and network polymers, depicted in darker blue shades, display higher conductivities. This preliminary finding aligns with the hypothesis in academia that increased structural complexity and the presence of more interconnected pathways facilitate better ionic transport. The data also reveals that network polymers maintain relatively high ionic conductivity across a broader temperature range, suggesting their potential advantage in varied operational conditions. These insights from the preliminary classification highlight the importance of backbone structure in designing SPEs with enhanced ionic conductivity.

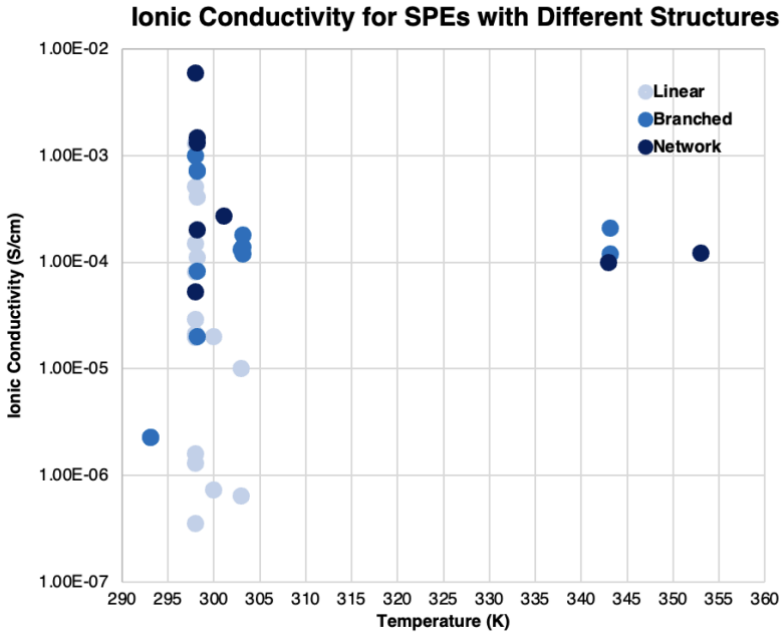

Figure 3.1. Ionic conductivity for SPEs with different backbone chemical structures.

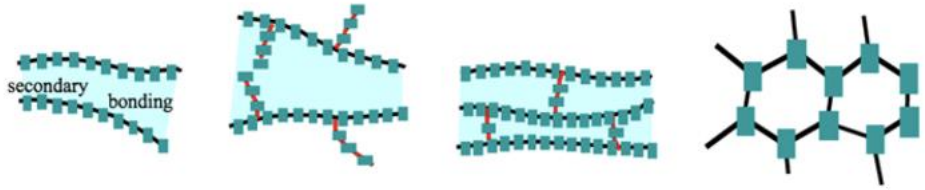

Figure 3.2. Schematic illustration for the classification for SPEs.

3.2.2 Data Cleaning

Data cleaning is a crucial step in preparing data for ML, as it ensures that the data fed into the model is accurate and uniform. For this study on the ionic conductivity of SPEs, several cleaning tasks were performed, with a key focus on standardizing ionic conductivity measurements to a common temperature base (298 K).

Ionic conductivity in polymers is significantly influenced by temperature. To ensure comparability across different samples and studies, it was necessary to standardize all ionic conductivity measurements to a consistent temperature of 298 K. The standardization process began with the identification of temperature-dependent data. Ionic conductivity data was gathered from various sources, where each potentially reported under different experimental conditions, particularly varying temperatures. We cataloged the temperature at which each conductivity measurement was reported, as stated in the source materials.

The temperature dependence of ionic conductivity in polymers is commonly modeled by the Arrhenius equation (**Eq. 2.2)**. While ***Chapter 2.2*** discusses additional models that incorporate variables such as the glass transition temperature, these models are not utilized in this study. The necessary parameters, like glass transition temperatures, are often not reported in the SPE literature, making these models impractical for our dataset. Therefore, the Arrhenius equation remains the preferred method for adjusting ionic conductivity to a standardized temperature due to its reliance on readily available data.

The adjustments were implemented using a scripted algorithm in Python, which iteratively applied the Arrhenius equation to each conductivity value, normalizing

them to reflect what they would be at 298 K. Following these adjustments, the data underwent a verification process where a visual examination of data plots was conducted to confirm that the adjusted conductivity values fell within expected ranges.

In addition to the temperature standardization, handling missing values for critical properties such as glass transition temperature and lithium transference number was an essential part of the data cleaning process. These properties are significant as they influence the ionic conductivity and overall performance of SPEs.

For each of these properties, the median was calculated from all available non-missing entries. This calculation was performed separately for each property to ensure that the imputed values were representative of the typical behavior observed in similar polymers. The calculated median values were then used to replace missing entries in the dataset. This imputation helped maintain the consistency of the dataset, ensuring that all samples had complete data entries for these critical properties.

3.2.3 Feature Engineering

In the domain of polymer research, the representation of polymers in a machine-readable format is crucial for applying machine learning techniques to predict material properties effectively, as introduced in ***Chapter 2.3.2***. For this thesis, we utilized two primary methods of feature engineering: Morgan Fingerprint (MF) and Mol2vec, which are instrumental in transforming polymer descriptors into numerical representations suitable for machine learning models.

MF is widely adopted in computational chemistry for its efficacy in capturing molecular structure in a compact form. This method involves the enumeration of substructures around each non-hydrogen atom within a defined radius. These substructures are then encoded into a fixed-length, high-dimensional, sparse vector using a hashing technique. The primary advantage of MF is its ability to rapidly process large datasets and its ease of integration with various machine learning algorithms. However, a notable limitation is the potential for hash collisions, which can obscure the uniqueness of molecular representations.

Mol2vec, inspired by the natural language processing model Word2Vec, represents molecules by converting SMILES strings into continuous vector spaces.[59] This method leverages a pre-trained model on 20 million molecular structures to learn vector representations of molecular substructures, effectively capturing the contextual relationships between different chemical environments. The vectors generated by Mol2vec are dense and continuous, which significantly reduces the chances of hash collisions compared to traditional hashing methods used in Morgan Fingerprints. This feature allows for more nuanced similarity measurements between molecules, facilitating the discovery of subtle structural relationships that may influence polymer properties.

Both Morgan Fingerprint and Mol2vec provide robust frameworks for the feature engineering of polymers, each with unique strengths and applications. The choice of feature extraction method can significantly influence the performance of subsequent machine learning models, making it essential to select the method that best aligns with the specific research questions and data characteristics of a given study.

3.2.4 Molecular Representation

In the context of using machine learning to predict ionic conductivity from the structure of SPEs, representing the molecular structures in a computationally accessible format is crucial. The process of converting these structures into a standard notation that machine learning algorithms can process involved several key steps, primarily focused on the creation of SMILES (Simplified Molecular Input Line Entry System) strings.

Molecular representations such as SMILES strings allow for the incorporation of complex chemical structures into data models. SMILES notation provides a way to encode molecular structures in a line of text, which is particularly useful for computational analyses including machine learning. However, the primary challenge was that many publications provide molecular structures only in graphical formats.

Initially, we explored the use of various image-to-SMILES conversion tools (**Table 3.3**). These tools are primarily designed to recognize and interpret graphical

depictions of molecules, converting them into SMILES notation. However, we encountered issues with accuracy and reliability, particularly with the complex polymers typical in SPE research (**Fig. 3.3**). This outcome was somewhat predictable, considering that these tools are generally optimized for recognizing small molecules, such as those found in pharmaceuticals, rather than the large, complex structures of polymer systems.

Table 3.3. Tools Explored for Molecular Structure Conversion to SMILES

Name	**URL**	**Comment**
ChemSchematicResolver[63]	https://github.com/edbeard/ChemSchematicResolver	Unable to install
MolScribe[64]	https://github.com/thomas0809/MolScribe	Best among all tools, but still has low accuracy
DECIMER[65]	https://github.com/Kohulan/DECIMER-Image_Transformer	Low accuracy
Img2Mol[66]	https://github.com/bayer-science-for-a-better-life/Img2Mol	Low accuracy
SwinOCSR[67]	https://github.com/suanfaxiaohuo/SwinOCSR/tree/main	Low accuracy
ChemScraper[68]	https://gitlab.com/dprl/graphics-extraction/-/tree/icdar2024	Low accuracy
OSRA[69]	https://cactus.nci.nih.gov/cgi-bin/osra/index.cgi	Low accuracy
ChatGPT 4	https://chatgpt.com	Unable to convert

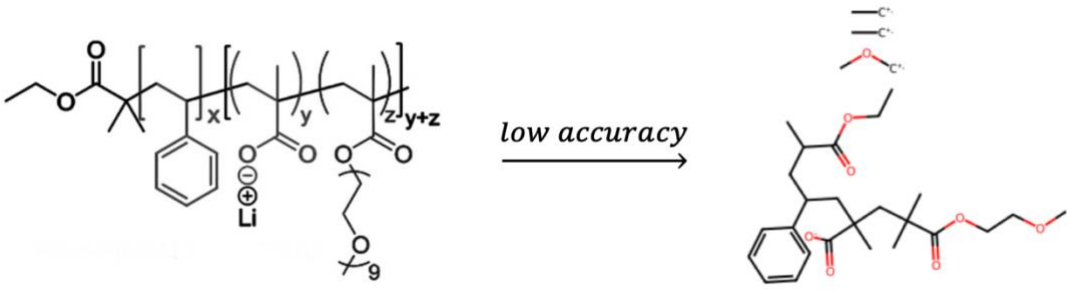

Figure 3.3. Example of molecular structure conversion to SMILES using MolScribe. Although MolScribe proved to be the most accurate tool among those listed in **Table**

3.3, its accuracy for converting complex polymer structures to SMILES notation remains relatively low.

Given the limitations of automated tools, we opted for the manual conversion of molecular images into SMILES strings using ChemDraw software. This approach required a meticulous process where each molecular structure was carefully recreated in ChemDraw by closely examining the graphical depictions in the source literature. This precise attention to detail ensured that all aspects of the molecular structure were accurately represented.

Once the structures were recreated, ChemDraw's built-in functionality was used to convert these structures into SMILES strings. This conversion mechanism provided by ChemDraw is known for its accuracy, ensuring that the generated SMILES strings are chemically valid and truly representative of the original structures. After generating the SMILES strings, we conducted a validation step to ensure their accuracy by utilizing RDKit, a widely used cheminformatics software. The SMILES strings were input into RDKit to regenerate the molecular structure images. These images were then meticulously compared to the original molecular figures from the literature. This comparison was essential to confirm that the regenerated structures precisely matched the intended molecular designs as depicted in the source materials. Discrepancies between the RDKit-generated images and the original figures prompted a re-evaluation and potential re-drawing of the structures to ensure that our SMILES strings accurately represented the chemical structures of the polymers. This careful validation process was crucial to maintaining the integrity of our molecular representations, providing a solid foundation for subsequent predictive modeling in our study.

For polymers consisting of several monomers, each monomer was drawn individually. The "A", or "*" symbol was used to represent the repeating unit of a polymer, which is a common notation to indicate polymerization in chemical structures (**Fig. 3.4**). The validated SMILES strings were then integrated into the main dataset, where each string was associated with its respective polymer's properties,

such as ionic conductivity and glass transition temperature, thus preparing the dataset for further machine learning analysis.

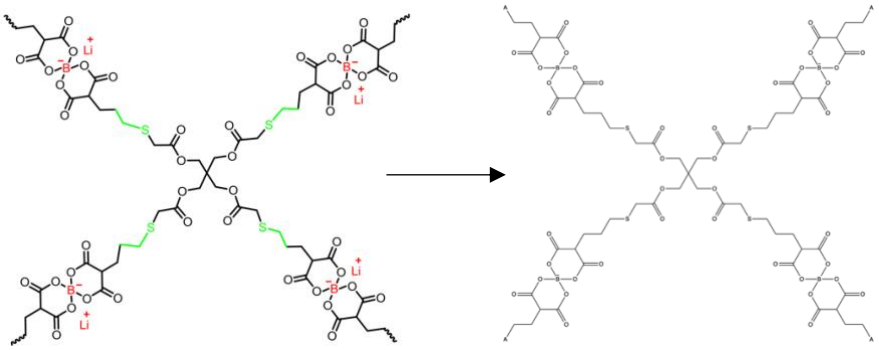

Figure 3.4. Example of polymer structure recreated in ChemDraw.

Despite being more time-consuming than automated methods, manual conversion provided significantly higher accuracy and reliability. This manual approach ensured that the molecular representations used in our machine learning models were precise and reflective of the actual chemical structures. This comprehensive approach to molecular representation through manual drawing and conversion to SMILES strings was crucial for the subsequent steps of our study, allowing us to effectively include molecular structure as a feature in our machine learning models to predict ionic conductivity.

3.3 Model Development

In this study, we aimed to predict the ionic conductivity of SPEs using machine learning techniques, specifically a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM). The development process for both models involved careful selection, architecture design, parameter tuning, and rigorous training procedures to ensure accuracy and reliability.

The CNN was chosen for its strengths in handling image data, which is beneficial for analyzing the 2D molecular graphs derived from SMILES strings. The architecture of the CNN included multiple convolutional layers to extract features from the images, followed by pooling layers to reduce dimensionality and prevent overfitting. Fully connected layers were used to integrate these features and make the final prediction. Training the CNN involved using the Adam optimizer with a learning rate

of 0.001, optimizing the Mean Squared Error (MSE) as the loss function. We trained the model over 100 epochs, incorporating early stopping to avoid overfitting, monitored through validation loss.

On the other hand, the SVM was selected for its effectiveness in high-dimensional spaces, suitable for the molecular descriptors generated from the data. Using a Radial Basis Function (RBF) kernel, the SVM captured non-linear relationships within the data. We employed grid search with cross-validation to fine-tune the regularization parameter (C) and the gamma parameter, ensuring the model was neither too complex nor too simple. After selecting the optimal hyperparameters, the model was retrained on the entire training dataset to finalize its setup.

Both models were evaluated using a test dataset, which was composed of 20% of the size of training dataset and kept separate from the training and validation phases to provide an unbiased assessment of their performance. We used metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to measure predictive accuracy comprehensively.

The CNN's ability to process and learn from image data provided valuable insights, particularly in extracting spatial features from molecular graphs. Meanwhile, the SVM's performance in high-dimensional feature spaces complemented this by leveraging the detailed molecular descriptors. The combination of these approaches allowed for a robust evaluation and comparison of their predictive capabilities.

Chapter 4: Results

4.1 Model Performance and Validation

In this section, we present the performance and validation results of the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) models using two different feature extraction methods: Morgan fingerprint and Mol2vec. The results highlight the effectiveness of each model and feature extraction combination in predicting the ionic conductivity of SPEs.

1. CNN with Morgan Fingerprint Features

The CNN model, trained on features extracted using the Morgan fingerprint method, demonstrated high performance on both training and testing datasets. For the training set, the model achieved an R^2 score of 0.95, indicating a strong correlation between the predicted and actual values. The Mean Absolute Error (MAE) was 0.23, and the Root Mean Squared Error (RMSE) was 0.29, reflecting the model's accuracy and precision in the training phase.

On the test set, the CNN model maintained robust performance with an R^2 score of 0.93. The MAE increased slightly to 0.28, and the RMSE remained consistent at 0.29. These results suggest that the CNN model generalized well to the unseen data, making it a potential predictor for this application (**Fig. 4.1**).

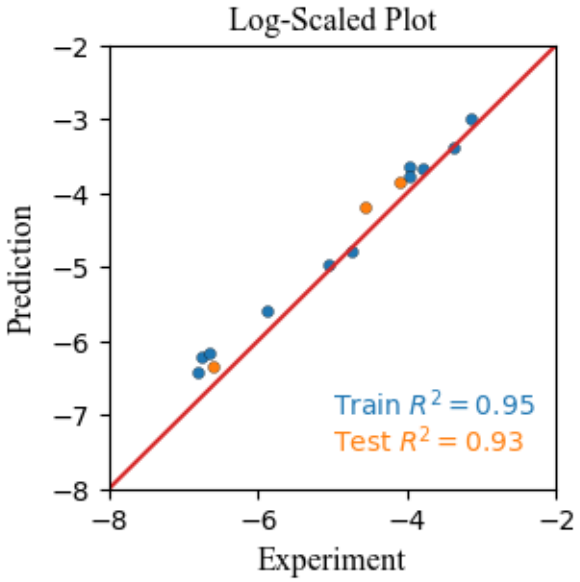

Figure 4.1. R^2 plot for training and testing dataset using CNN with Morgan fingerprint features.

Fig. 4.2 shows a comparison of actual and predicted ionic conductivity values (log scaled) for the test set. The blue points represent the actual values, while the red points represent the predicted values. The close alignment between the blue and red points indicates the model's strong predictive capability, with most predictions falling near the actual values.

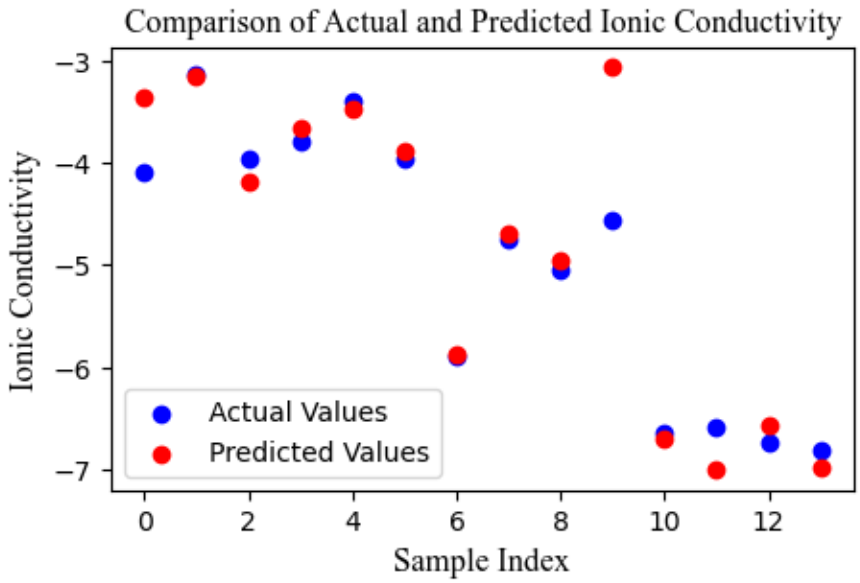

Figure 4.2. Comparison of actual and predicted ionic conductivity using CNN with Morgan fingerprint features.

Fig. 4.3 illustrates the training and validation loss over 100 epochs. The training loss (blue line) decreases steadily, indicating the model's learning progress. However, the validation loss (orange line) exhibits significant fluctuations, suggesting potential overfitting. Overfitting occurs when the model learns the training data too well, including noise, and fails to generalize to new data. This is evident from the inconsistent and sometimes spiking validation loss.

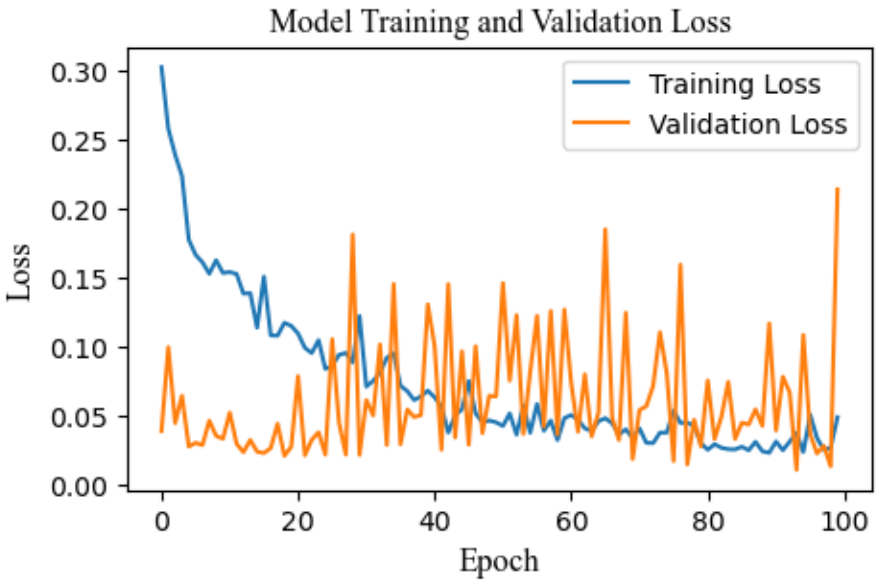

Figure 4.3. Model training and validation loss for CNN with Morgan fingerprint features.

2. SVM with Morgan Fingerprint Features

The SVM model using Morgan fingerprint features showed contrasting performance. The training Mean Squared Error (MSE) was 0.1802, indicating a reasonable fit on the training data. However, the testing MSE increased substantially to 0.8297, and the R^2 score on the test set was 0.293, highlighting a significant drop in predictive accuracy on unseen data.

Cross-validation revealed further inconsistencies with scores ranging from -3.703 to -298.468, and a mean cross-validation score of -65.236. These negative values indicate poor generalization and potential overfitting issues, suggesting that the SVM model struggled to capture the underlying patterns in the data using Morgan fingerprints.

3. SVM with Mol2vec Features

Switching to Mol2vec features improved the SVM model's performance. The training MSE dropped to 0.0095, showcasing a near-perfect fit to the training data. However, the test MSE was 0.4867, and the R^2 score was 0.586, indicating moderate improvement in predictive performance compared to using Morgan fingerprints.

The cross-validation scores varied from -2.154 to -734.090, with a mean score of -148.047. Although the performance improved compared to Morgan fingerprints, the

wide range in scores indicates potential overfitting and variability in model performance across different data splits.

4. CNN with Mol2vec Features

The CNN model using Mol2vec features exhibited exceptional performance on both the training and test datasets. For the training set, the model achieved an R^2 score of 0.97, with an MAE of 0.14 and an RMSE of 0.22. The test set results were even more remarkable, with an R^2 score of 1.00, an MAE of 0.05, and an RMSE of 0.07. These results suggest near-perfect prediction accuracy (**Fig. 4.4**).

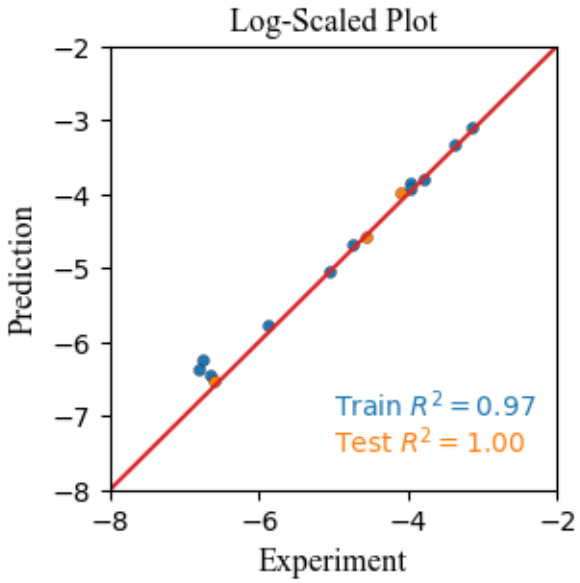

Figure 4.4. R^2 plot for training and testing dataset using CNN with Mol2vec features.

Fig. 4.5 illustrates the training and validation loss over 100 epochs for the CNN model with Mol2vec features. The training loss consistently decreases and remains low, indicating effective learning. However, the high and fluctuating validation loss

suggests overfitting, as the model's performance on the validation set is not as stable.

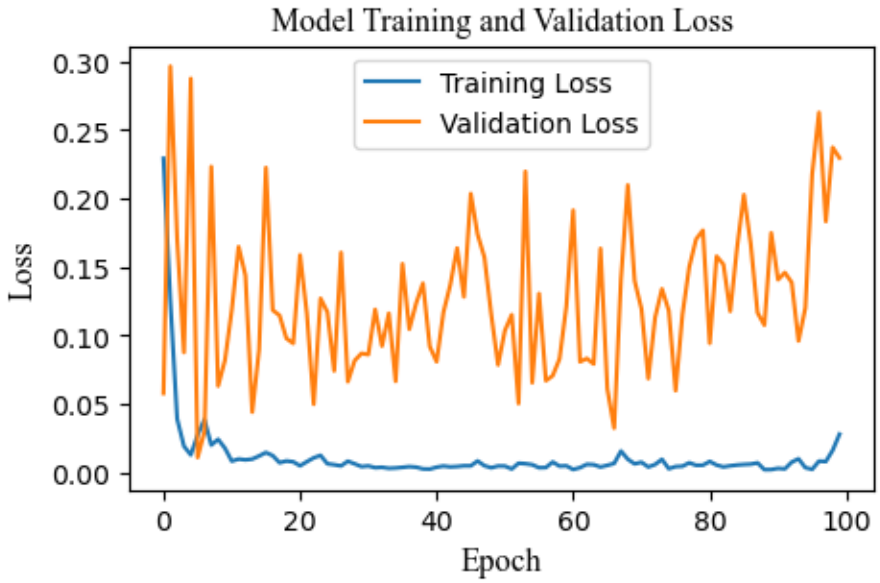

Figure 4.5. Model training and validation loss for CNN with Mol2vec features.

Given the small dataset size of only 14 samples, this high performance likely indicates overfitting. The model's near-perfect fit on such a limited dataset suggests that it may not generalize well to larger or more diverse datasets. Despite this, the CNN with Mol2vec features shows promise and could be highly effective when applied to a more extensive dataset in future studies.

In summary, the CNN model, particularly with Mol2vec features, outperformed the SVM models and demonstrated high predictive accuracy. However, the potential overfitting issue due to the small dataset size needs to be addressed. As the dataset grows, the robustness and generalizability of these models should be re-evaluated. The results underscore the importance of feature selection and model choice in predicting the ionic conductivity of SPEs, providing valuable insights for future research in this domain.

4.2 Limitations and Implications of the Study

Despite the promising results obtained from our machine learning models, several limitations must be acknowledged. A primary limitation of this study is the small dataset size, comprising only 14 samples. This limited dataset raises concerns about the generalizability of our models, particularly the Convolutional Neural Network

(CNN) trained with Mol2vec features, which exhibited near-perfect performance on both training and test sets. Such high accuracy on a small dataset often indicates overfitting, where the model captures noise and specificities in the training data rather than general patterns that would be applicable to new, unseen data.

The high variability in the validation loss observed during the training of the CNN models further underscores the overfitting issue. While the training loss consistently decreased, the validation loss fluctuated significantly, suggesting that the model’s performance on unseen data is unstable. This instability is particularly concerning given the small sample size, which does not provide a robust foundation for validating model performance.

Another limitation is the feature extraction methods used. Although both Morgan fingerprints and Mol2vec features were effective, each has inherent biases that might influence the model's performance. Morgan fingerprints, for instance, may not capture all relevant structural information about complex polymer chains, while Mol2vec embeddings, although more comprehensive, could introduce noise from irrelevant molecular substructures. The reliance on these methods means that our models might miss important nuances of the polymer structures that could affect ionic conductivity.

The implications of these limitations are significant for the broader applicability of our findings. While our models show potential, particularly the CNN with Mol2vec features, their utility in real-world applications remains uncertain until they can be validated on larger, more diverse datasets. The overfitting observed suggests that caution must be exercised when interpreting the model's predictions, and any practical use of these models should be preceded by extensive validation on new data.

Furthermore, the study highlights the importance of comprehensive data collection and the need for larger datasets in polymer research. The challenges encountered in integrating and validating data from various sources underscore the necessity for standardized reporting in the field, which would facilitate the creation of more robust and generalizable machine learning models.

Chapter 5: Conclusions

5.1 Summary of Findings

This study explored the use of machine learning models to predict the ionic conductivity of SPEs based on their molecular structures. Employing two feature extraction methods, Morgan fingerprint and Mol2vec, we developed and evaluated Convolutional Neural Network (CNN) and Support Vector Machine (SVM) models. Our findings indicate that the CNN model, particularly when combined with Mol2vec features, demonstrated superior predictive accuracy, achieving an R^2 score of 1.00 on the test set, albeit with signs of overfitting due to the small dataset size.

The SVM models showed variable performance, with the model using Mol2vec features outperforming the one using Morgan fingerprints, yet still falling short of the CNN's performance. The variability in the SVM results, reflected in wide-ranging cross-validation scores, highlighted potential issues with model stability and generalizability. In contrast, the CNN's ability to closely align predicted values with actual measurements suggests a robust learning capability, although the significant fluctuations in validation loss signal overfitting concerns.

Additionally, the study revealed that polymers with flexible single bond long chain backbones, more complex backbone structures, and polar functional groups like sulfonate and borate groups, could enhance ionic conductivity, corroborating existing literature. Furthermore, the evaluation of molecular image-to-SMILES software highlighted inefficiencies in conversion accuracy, emphasizing the need for manual verification. Finally, we found that many polymers reported in the literature are not available in the PoLyInfo database, limiting the scope of readily accessible data.

5.2 Contributions to the Field

This research contributes to the burgeoning field of computational materials science by demonstrating the potential of machine learning models to predict key properties of polymer electrolytes, which are critical for the development of advanced energy storage systems. The use of Mol2vec features, in particular, represents a novel

approach in the context of SPEs, providing a more nuanced understanding of molecular structures and their influence on ionic conductivity.

Our study underscores the efficacy of deep learning models, such as CNNs, in handling complex molecular data, outperforming traditional machine learning approaches like SVMs in this context. Furthermore, the detailed examination of model performance and the identification of overfitting issues provide valuable insights into the limitations and potential pitfalls of applying machine learning to small datasets in materials science.

By integrating cheminformatics tools like RDKit for SMILES string generation and validation, this research also highlights the importance of accurate molecular representation in predictive modeling. The meticulous process of data preparation and validation outlined in this study sets a precedent for future research, emphasizing the need for rigorous data handling practices.

Additionally, the findings regarding the structural characteristics of polymers that enhance ionic conductivity offer practical insights for the design of new materials. The inefficiency of current molecular image-to-SMILES software underscores the need for improved tools in this area, while the lack of comprehensive polymer data in existing databases like PoLyInfo highlights the necessity for more extensive and accessible data repositories.

5.3 Future Directions

Building on the findings and limitations identified in this study, several avenues for future research are recommended. A primary objective should be the expansion of the dataset to include a broader range of polymer structures and ionic conductivity measurements. Larger and more diverse datasets will mitigate overfitting and enhance the generalizability of the models.

Future studies should also explore advanced feature extraction techniques, such as graph neural networks (GNNs), which can provide more comprehensive representations of complex polymer structures. Integrating domain-specific

knowledge into the feature engineering process could lead to more interpretable and accurate models.

Regularization techniques, such as dropout and L2 regularization, should be systematically applied and tested to address the overfitting issues observed in this study. Additionally, the use of cross-validation methods, particularly k-fold cross-validation, can provide a more robust evaluation framework for model performance.

Collaboration between experimentalists and computational researchers is crucial for generating new data and validating model predictions. Developing hybrid models that combine machine learning approaches with mechanistic insights from polymer science could yield more reliable and actionable predictions.

Lastly, creating user-friendly tools and platforms that integrate machine learning workflows for polymer research will democratize access to these advanced techniques, fostering broader adoption and innovation in the field. Efforts should also be directed towards improving molecular image-to-SMILES conversion software and expanding databases like PoLyInfo to include a wider array of polymers documented in the literature.

In conclusion, while this study demonstrates the promise of machine learning in predicting the ionic conductivity of SPEs, addressing the limitations through larger datasets, advanced feature extraction, regularization, and collaborative efforts will be key to advancing this field and unlocking its full potential for materials discovery and optimization.

References

(1) Assat, G.; Tarascon, J.-M. Fundamental Understanding and Practical Challenges of Anionic Redox Activity in Li-Ion Batteries. *Nat. Energy* **2018**, *3* (5), 373–386. https://doi.org/10.1038/s41560-018-0097-0.
(2) Zhang, H.; Li, C.; Piszcz, M.; Coya, E.; Rojo, T.; Rodriguez-Martinez, L. M.; Armand, M.; Zhou, Z. Single Lithium-Ion Conducting Solid Polymer Electrolytes: Advances and Perspectives. *Chem. Soc. Rev.* **2017**, *46* (3), 797–815. https://doi.org/10.1039/C6CS00491A.
(3) Zhao, R.; Liu, J.; Gu, J. Simulation and Experimental Study on Lithium Ion Battery Short Circuit. *Appl. Energy* **2016**, *173*, 29–39. https://doi.org/10.1016/j.apenergy.2016.04.016.
(4) Zhang, S. S. Liquid Electrolyte Lithium/Sulfur Battery: Fundamental Chemistry, Problems, and Solutions. *J. Power Sources* **2013**, *231*, 153–162. https://doi.org/10.1016/j.jpowsour.2012.12.102.
(5) Cheng, X.-B.; Zhang, R.; Zhao, C.-Z.; Zhang, Q. Toward Safe Lithium Metal Anode in Rechargeable Batteries: A Review. *Chem. Rev.* **2017**, *117* (15), 10403–10473. https://doi.org/10.1021/acs.chemrev.7b00115.
(6) Castillo, J.; Santiago, A.; Judez, X.; Garbayo, I.; Coca Clemente, J. A.; Morant-Miñana, M. C.; Villaverde, A.; González-Marcos, J. A.; Zhang, H.; Armand, M.; Li, C. Safe, Flexible, and High-Performing Gel-Polymer Electrolyte for Rechargeable Lithium Metal Batteries. *Chem. Mater.* **2021**, *33* (22), 8812–8821. https://doi.org/10.1021/acs.chemmater.1c02952.
(7) An, Y.; Han, X.; Liu, Y.; Azhar, A.; Na, J.; Nanjundan, A. K.; Wang, S.; Yu, J.; Yamauchi, Y. Progress in Solid Polymer Electrolytes for Lithium-Ion Batteries and Beyond. *Small* **2022**, *18* (3), 2103617. https://doi.org/10.1002/smll.202103617.
(8) Kerman, K.; Luntz, A.; Viswanathan, V.; Chiang, Y.-M.; Chen, Z. Review—Practical Challenges Hindering the Development of Solid State Li Ion Batteries. *J. Electrochem. Soc.* **2017**, *164* (7), A1731–A1744. https://doi.org/10.1149/2.1571707jes.
(9) Ao, X.; Wang, X.; Tan, J.; Zhang, S.; Su, C.; Dong, L.; Tang, M.; Wang, Z.; Tian, B.; Wang, H. Nanocomposite with Fast Li+ Conducting Percolation Network: Solid Polymer Electrolyte with Li+ Non-Conducting Filler. *Nano Energy* **2021**, *79*, 105475. https://doi.org/10.1016/j.nanoen.2020.105475.
(10) Marchiori, C. F. N.; Carvalho, R. P.; Ebadi, M.; Brandell, D.; Araujo, C. M. Understanding the Electrochemical Stability Window of Polymer Electrolytes in Solid-State Batteries from Atomic-Scale Modeling: The Role of Li-Ion Salts. *Chem. Mater.* **2020**, *32* (17), 7237–7246. https://doi.org/10.1021/acs.chemmater.0c01489.
(11) Ahmad, Z.; Xie, T.; Maheshwari, C.; Grossman, J. C.; Viswanathan, V. Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in Lithium Metal Anodes. *ACS Cent. Sci.* **2018**, *4* (8), 996–1006. https://doi.org/10.1021/acscentsci.8b00229.
(12) Fenton, D. E.; Parker, J. M.; Wright, P. V. Complexes of Alkali Metal Ions with Poly(Ethylene Oxide). *Polymer* **1973**, *14* (11), 589. https://doi.org/10.1016/0032-3861(73)90146-8.

(13) Wright, P. V. Electrical Conductivity in Ionic Complexes of Poly(Ethylene Oxide). *Br. Polym. J.* **1975**, *7* (5), 319–327. https://doi.org/10.1002/pi.4980070505.
(14) Huang, C.; Qian, X.; Yang, R. Thermal Conductivity of Polymers and Polymer Nanocomposites. *Mater. Sci. Eng. R Rep.* **2018**, *132*, 1–22. https://doi.org/10.1016/j.mser.2018.06.002.
(15) Jolly, D.; Cres, Y.; Dimitriadis, S. Vincent Bolloré's Long Bet on Solid-State Batteries for Electric Cars. *The New York Times*. https://www.nytimes.com/2015/06/13/business/international/vincent-bollores-long-bet-on-solid-state-batteries-for-electric-cars.html.
(16) Xu, X.; Wang, S.; Wang, H.; Hu, C.; Jin, Y.; Liu, J.; Yan, H. Recent Progresses in the Suppression Method Based on the Growth Mechanism of Lithium Dendrite. *J. Energy Chem.* **2018**, *27* (2), 513–527. https://doi.org/10.1016/j.jechem.2017.11.010.
(17) Guo, B.; Fu, Y.; Wang, J.; Gong, Y.; Zhao, Y.; Yang, K.; Zhou, S.; Liu, L.; Yang, S.; Liu, X.; Pan, F. Strategies and Characterization Methods for Achieving High Performance PEO-Based Solid-State Lithium-Ion Batteries. *Chem. Commun.* **2022**, *58* (59), 8182–8193. https://doi.org/10.1039/D2CC02306G.
(18) Monroe, C.; Newman, J. The Impact of Elastic Deformation on Deposition Kinetics at Lithium/Polymer Interfaces. *J. Electrochem. Soc.* **2005**, *152* (2), A396. https://doi.org/10.1149/1.1850854.
(19) Ma, B.; Zhong, L.; Huang, S.; Xiao, M.; Wang, S.; Han, D.; Meng, Y. Covalent Organic Framework Enhanced Solid Polymer Electrolyte for Lithium Metal Batteries. *Molecules* **2024**, *29* (8), 1759. https://doi.org/10.3390/molecules29081759.
(20) Wang, L.; Huang, J.; Shen, Y.; Ma, M.; Ruan, W.; Zhang, M. ARGET-ATRP-Mediated Grafting of Bifunctional Polymers onto Silica Nanoparticles Fillers for Boosting the Performance of High-Capacity All-Solid-State Lithium–Sulfur Batteries with Polymer Solid Electrolytes. *Polymers* **2024**, *16* (8), 1128. https://doi.org/10.3390/polym16081128.
(21) Zhao, T.; Yang, J.; Cai, L.; Zhu, W.; Wu, X.; Shao, L. Enhancing the Performance of Poly(Ethylene Oxide) Solid Electrolytes through Synergistic Interaction of Starch and Butanedinitrile. *J. Appl. Polym. Sci.* **2024**, e55538. https://doi.org/10.1002/app.55538.
(22) Yang, H.; Wu, N. Ionic Conductivity and Ion Transport Mechanisms of Solid-state Lithium-ion Battery Electrolytes: A Review. *Energy Sci. Eng.* **2022**, *10* (5), 1643–1671. https://doi.org/10.1002/ese3.1163.
(23) Ratner, M. A.; Johansson, P.; Shriver, D. F. Polymer Electrolytes: Ionic Transport Mechanisms and Relaxation Coupling. *MRS Bull.* **2000**, *25* (3), 31–37. https://doi.org/10.1557/mrs2000.16.
(24) Eriksson, T.; Mace, A.; Mindemark, J.; Brandell, D. The Role of Coordination Strength in Solid Polymer Electrolytes: Compositional Dependence of Transference Numbers in the Poly(ε-Caprolactone)–Poly(Trimethylene Carbonate) System. *Phys. Chem. Chem. Phys.* **2021**, *23* (45), 25550–25557. https://doi.org/10.1039/D1CP03929F.

(25) Li, Z.; Fu, J.; Zhou, X.; Gui, S.; Wei, L.; Yang, H.; Li, H.; Guo, X. Ionic Conduction in Polymer-Based Solid Electrolytes. *Adv. Sci.* **2023**, *10* (10), 2201718. https://doi.org/10.1002/advs.202201718.
(26) Deng, K.; Wang, S.; Ren, S.; Han, D.; Xiao, M.; Meng, Y. Network Type Sp3 Boron-Based Single-Ion Conducting Polymer Electrolytes for Lithium Ion Batteries. *J. Power Sources* **2017**, *360*, 98–105. https://doi.org/10.1016/j.jpowsour.2017.06.006.
(27) Sun, X.-G.; Kerr, J. B. Synthesis and Characterization of Network Single Ion Conductors Based on Comb-Branched Polyepoxide Ethers and Lithium Bis(Allylmalonato)Borate. *Macromolecules* **2006**, *39* (1), 362–372. https://doi.org/10.1021/ma0507701.
(28) Doyle, M.; Fuller, T. F.; Newman, J. The Importance of the Lithium Ion Transference Number in Lithium/Polymer Cells. *Electrochimica Acta* **1994**, *39* (13), 2073–2081. https://doi.org/10.1016/0013-4686(94)85091-7.
(29) Brissot, C.; Rosso, M.; Chazalviel, J.-N.; Lascaud, S. Dendritic Growth Mechanisms in Lithium/Polymer Cells. *J. Power Sources* **1999**, *81–82*, 925–929. https://doi.org/10.1016/S0378-7753(98)00242-0.
(30) Bannister, D. J.; Davies, G. R.; Ward, I. M.; McIntyre, J. E. Ionic Conductivities for Poly(Ethylene Oxide) Complexes with Lithium Salts of Monobasic and Dibasic Acids and Blends of Poly(Ethylene Oxide) with Lithium Salts of Anionic Polymers. *Polymer* **1984**, *25* (9), 1291–1296. https://doi.org/10.1016/0032-3861(84)90378-1.
(31) Mao, G.; Perea, R. F.; Howells, W. S.; Price, D. L.; Saboungi, M.-L. Relaxation in Polymer Electrolytes on the Nanosecond Timescale. *Nature* **2000**, *405* (6783), 163–165. https://doi.org/10.1038/35012032.
(32) Long, L.; Wang, S.; Xiao, M.; Meng, Y. Polymer Electrolytes for Lithium Polymer Batteries. *J. Mater. Chem. A* **2016**, *4* (26), 10038–10069. https://doi.org/10.1039/C6TA02621D.
(33) Golodnitsky, D.; Strauss, E.; Peled, E.; Greenbaum, S. Review—On Order and Disorder in Polymer Electrolytes. *J. Electrochem. Soc.* **2015**, *162* (14), A2551–A2566. https://doi.org/10.1149/2.0161514jes.
(34) Young, W.; Kuan, W.; Epps, T. H. Block Copolymer Electrolytes for Rechargeable Lithium Batteries. *J. Polym. Sci. Part B Polym. Phys.* **2014**, *52* (1), 1–16. https://doi.org/10.1002/polb.23404.
(35) Zhou, Q.; Ma, J.; Dong, S.; Li, X.; Cui, G. Intermolecular Chemistry in Solid Polymer Electrolytes for High-Energy-Density Lithium Batteries. *Adv. Mater.* **2019**, *31* (50), 1902029. https://doi.org/10.1002/adma.201902029.
(36) Reinoso, D. M.; Frechero, M. A. Strategies for Rational Design of Polymer-Based Solid Electrolytes for Advanced Lithium Energy Storage Applications. *Energy Storage Mater.* **2022**, *52*, 430–464. https://doi.org/10.1016/j.ensm.2022.08.019.
(37) Aziz, S. B.; Woo, T. J.; Kadir, M. F. Z.; Ahmed, H. M. A Conceptual Review on Polymer Electrolytes and Ion Transport Models. *J. Sci. Adv. Mater. Devices* **2018**, *3* (1), 1–17. https://doi.org/10.1016/j.jsamd.2018.01.002.
(38) Rohan, R.; Pareek, K.; Chen, Z.; Cai, W.; Zhang, Y.; Xu, G.; Gao, Z.; Cheng, H. A High Performance Polysiloxane-Based Single Ion Conducting Polymeric

Electrolyte Membrane for Application in Lithium Ion Batteries. *J. Mater. Chem. A* **2015**, *3* (40), 20267–20276. https://doi.org/10.1039/C5TA02628H.
(39) Tsuchida, E.; Ohno, H.; Tsunemi, K.; Kobayashi, N. Lithium Ionic Conduction in Poly (Methacrylic Acid)-Poly (Ethylene Oxide) Complex Containing Lithium Perchlorate. *Solid State Ion.* **1983**, *11* (3), 227–233. https://doi.org/10.1016/0167-2738(83)90028-0.
(40) Huang, Y.; Xie, Z.; Zhu, W.; Mo, C.; Li, W.; Liao, Y. Polycarbonate Copolymer Solid Electrolyte for Stable Cycling of a Li‖LiCoO $_2$ Cell via In Situ Ultraviolet Irradiation. *ACS Appl. Polym. Mater.* **2024**, *6* (7), 3624–3636. https://doi.org/10.1021/acsapm.3c02451.
(41) Kang, Y.; Lee, J.; Lee, J.; Lee, C. Ionic Conductivity and Electrochemical Properties of Cross-Linked Solid Polymer Electrolyte Using Star-Shaped Siloxane Acrylate. *J. Power Sources* **2007**, *165* (1), 92–96. https://doi.org/10.1016/j.jpowsour.2006.11.019.
(42) Bicerano, J. *Computational Modeling of Polymers*; Plastics engineering; Dekker: New York, 1992.
(43) Allahyarov, E.; Taylor, P. L. Simulation Study of the Correlation between Structure and Conductivity in Stretched Nafion. *J. Phys. Chem. B* **2009**, *113* (3), 610–617. https://doi.org/10.1021/jp8047746.
(44) Dünweg, B.; Kremer, K. Molecular Dynamics Simulation of a Polymer Chain in Solution. *J. Chem. Phys.* **1993**, *99* (9), 6983–6997. https://doi.org/10.1063/1.465445.
(45) Mizera, A.; Dubis, A. T.; Łapiński, A. Density Functional Theory Studies of Polypyrrole and Polypyrrole Derivatives; Substituent Effect on the Optical and Electronic Properties. *Polymer* **2022**, *255*, 125127. https://doi.org/10.1016/j.polymer.2022.125127.
(46) Kiely, E.; Zwane, R.; Fox, R.; Reilly, A. M.; Guerin, S. Density Functional Theory Predictions of the Mechanical Properties of Crystalline Materials. *CrystEngComm* **2021**, *23* (34), 5697–5710. https://doi.org/10.1039/D1CE00453K.
(47) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polym. Au* **2022**, *2* (1), 8–26. https://doi.org/10.1021/acspolymersau.1c00035.
(48) Steinhauser, M.; Hiermaier, S. A Review of Computational Methods in Materials Science: Examples from Shock-Wave and Polymer Physics. *Int. J. Mol. Sci.* **2009**, *10* (12), 5135–5216. https://doi.org/10.3390/ijms10125135.
(49) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(50) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112* (5), 2889–2919. https://doi.org/10.1021/cr200066h.
(51) Sha, W.; Li, Y.; Tang, S.; Tian, J.; Zhao, Y.; Guo, Y.; Zhang, W.; Zhang, X.; Lu, S.; Cao, Y.; Cheng, S. Machine Learning in Polymer Informatics. *InfoMat* **2021**, *3* (4), 353–361. https://doi.org/10.1002/inf2.12167.
(52) Potyrailo, R.; Rajan, K.; Stoewe, K.; Takeuchi, I.; Chisholm, B.; Lam, H. Combinatorial and High-Throughput Screening of Materials Libraries: Review of

State of the Art. *ACS Comb. Sci.* **2011**, *13* (6), 579–633. https://doi.org/10.1021/co200007w.
(53) Tao, L.; Byrnes, J.; Varshney, V.; Li, Y. Machine Learning Strategies for the Structure-Property Relationship of Copolymers. *iScience* **2022**, *25* (7), 104585. https://doi.org/10.1016/j.isci.2022.104585.
(54) Queen, O.; McCarver, G. A.; Thatigotla, S.; Abolins, B. P.; Brown, C. L.; Maroulas, V.; Vogiatzis, K. D. Polymer Graph Neural Networks for Multitask Property Learning. *Npj Comput. Mater.* **2023**, *9* (1), 90. https://doi.org/10.1038/s41524-023-01034-3.
(55) Tao, L.; Chen, G.; Li, Y. Machine Learning Discovery of High-Temperature Polymers. *Patterns* **2021**, *2* (4), 100225. https://doi.org/10.1016/j.patter.2021.100225.
(56) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36. https://doi.org/10.1021/ci00057a005.
(57) Ma, R.; Liu, Z.; Zhang, Q.; Liu, Z.; Luo, T. Evaluating Polymer Representations via Quantifying Structure–Property Relationships. *J. Chem. Inf. Model.* **2019**, *59* (7), 3110–3119. https://doi.org/10.1021/acs.jcim.9b00358.
(58) Zhao, Y.; Mulder, R. J.; Houshyar, S.; Le, T. C. A Review on the Application of Molecular Descriptors and Machine Learning in Polymer Design. *Polym. Chem.* **2023**, *14* (29), 3325–3346. https://doi.org/10.1039/D3PY00395G.
(59) Jaeger, S.; Fulle, S.; Turk, S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* **2018**, *58* (1), 27–35. https://doi.org/10.1021/acs.jcim.7b00616.
(60) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. PoLyInfo: Polymer Database for Polymeric Materials Design. In *2011 International Conference on Emerging Intelligent Data and Web Technologies*; IEEE: Tirana, Albania, 2011; pp 22–29. https://doi.org/10.1109/EIDWT.2011.13.
(61) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002. https://doi.org/10.1063/1.4812323.
(62) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122* (31), 17575–17585. https://doi.org/10.1021/acs.jpcc.8b02913.
(63) Beard, E. J.; Cole, J. M. ChemSchematicResolver: A Toolkit to Decode 2D Chemical Diagrams with Labels and R-Groups into Annotated Chemical Named Entities. *J. Chem. Inf. Model.* **2020**, *60* (4), 2059–2072. https://doi.org/10.1021/acs.jcim.0c00042.
(64) Qian, Y.; Guo, J.; Tu, Z.; Li, Z.; Coley, C. W.; Barzilay, R. MolScribe: Robust Molecular Structure Recognition with Image-to-Graph Generation. *J. Chem. Inf. Model.* **2023**, *63* (7), 1925–1934. https://doi.org/10.1021/acs.jcim.2c01480.

(65) Rajan, K.; Zielesny, A.; Steinbeck, C. DECIMER: Towards Deep Learning for Chemical Image Recognition. *J. Cheminformatics* **2020**, *12* (1), 65. https://doi.org/10.1186/s13321-020-00469-w.
(66) Clevert, D.-A.; Le, T.; Winter, R.; Montanari, F. Img2Mol – Accurate SMILES Recognition from Molecular Graphical Depictions. *Chem. Sci.* **2021**, *12* (42), 14174–14181. https://doi.org/10.1039/D1SC01839F.
(67) Xu, Z.; Li, J.; Yang, Z.; Li, S.; Li, H. SwinOCSR: End-to-End Optical Chemical Structure Recognition Using a Swin Transformer. *J. Cheminformatics* **2022**, *14* (1), 41. https://doi.org/10.1186/s13321-022-00624-5.
(68) Shah, A. K.; Amador, B. M.; Dey, A.; Creekmore, M.; Ocampo, B.; Denmark, S.; Zanibbi, R. ChemScraper: Leveraging PDF Graphics Instructions for Molecular Diagram Parsing. **2023**. https://doi.org/10.48550/ARXIV.2311.12161.
(69) Filippov, I. V.; Nicklaus, M. C. Optical Structure Recognition Software To Recover Chemical Information: OSRA, An Open Source Solution. *J. Chem. Inf. Model.* **2009**, *49* (3), 740–743. https://doi.org/10.1021/ci800067r.

Appendix

Table 1. Polymer Properties on PoLyInfo and Their Correlation with Ionic Conductivity

Property	Relationship	References
Physical properties		
Density	High density usually causes lower ionic conductivity b/c high density is usually associated with high crystallinity (1) which reduces ionic conductivity (2). Also free volume (less dense) in the polymer (PEO) could facilitate ion transfer (3). Density is also proportional to molecular weight (degree of polymerization/chain length) (4).	1. Chapter 4 - Mechanical Properties of Polymer Solids and Liquids 2. Reducing crystallinity in solid polymer electrolytes for lithium-metal batteries via statistical copolymerization | Communications Materials 3. Decoding Polymer Architecture Effect on Ion Clustering, Chain Dynamics, and Ionic Conductivity in Polymer Electrolytes | ACS Applied Energy Materials 4. https://pubs.acs.org/doi/pdf/10.1021/ma00095a001
Specific volume	Reciprocal of Density	Same as Density
Optical properties		
Refractive index	RI is related to molar refraction which is related to the polarizability and density of the material, with higher molar refractivity values obtained with more polarizable, higher density atoms/functional groups. The beam of light entering a medium causes a disruption of electron density, slowing the electromagnetic wave. More polarizable materials slow the wave more, hence increasing the RI.	1. High refractive index polymers: fundamental research and practical applications 2. Intrinsic high refractive index polymers - Macdonald - 2015
Stress optical coeff.	The stress-optical coefficient is related to a configuration-dependent property of the polymer. Thus, SOC	1. Stress Optical Coefficient, Test Methodology, and Glass Standard Evaluation

	is an indirect measure of the glass composition and structure.	2. Refractive Index, Stress-Optical Coefficient, and Optical Configuration Parameter of Polymers - SpringerMaterials
Thermal properties Intro to thermal properties of polymers: Investigation of Polymers with Differential Scanning Calorimetry Contents		
Crystallization kinetics (R, k, and n)	These three properties are used in the Avrami equation which describes the rate of isothermal crystallization of polymers. They are associated with many factors such as molecular weight of the sample, the molecular weight distribution, temperature and time, etc. Therefore the same polymer could have very different kinetics and may not be appropriate for ML.	1. [PoLyInfo HELP] Property 2. Modeling the crystallization kinetics of polymers displaying high levels of secondary crystallization
Crystallization temp.	Generally, crystallization limits ionic conductivity. Therefore, higher crystallization temp. is favored for ionic conductivity.	Effect of Chemical Structure and Salt Concentration on the Crystallization and Ionic Conductivity of Aliphatic Polyethers
Glass transition temp. (Describes the amorphous part of polymers)	Lower GTT (with accompanying low viscosity) improves ion conductivity (1-4).	1. Glass Transition Temperature and Ion Binding Determine Conductivity and Lithium–Ion Transport in Polymer Electrolytes | ACS Macro Letters 2. Relationship between Ionic Conductivity, Glass Transition Temperature, and Dielectric Constant in Poly(vinyl ether) Lithium Electrolytes | ACS Macro Letters 3. Effect of Plasticization on Ionic Conductivity Enhancement in Relation to Glass Transition Temperature of Crosslinked Polymer Electrolyte Membranes 4. Glass transition and ionic conduction in plasticized and doped ionomers - ScienceDirect

Melting temp. (Describes the crystalline part of polymers)	Melting temperature of polymers reflects the structure of the polymer and chain flexibility, which is critical for ion transport. Also, melting temperature and glass transition temperature are linearly correlated so lower MT improves ion conductivity.	1. Relation between glass transition and melting of PEO–salt complexes - ScienceDirect 2. Dynamics of Poly(ethylene oxide) around Its Melting Temperature | Macromolecules 3. Effects of inhomogeneity on ionic conductivity and relaxations in PEO and PEO–salt complexes - ScienceDirect 4. https://www.sciencedirect.com/science/article/pii/S0167273805002183
Heat of fusion	Heat of fusion decreases and ionic conductivity increases with increasing Li salt concentration. Therefore, heat of fusion may be negatively correlated with ionic conductivity.	1. https://www.sciencedirect.com/science/article/pii/S0378775303005147?via%3Dihub
Heat of crystallization	Heat of crystallization (ΔH_{cr}) is the release of heat of fusion (ΔH_{fus}) during crystallization, so it should have the same value. However, I noticed in PoLyInfo some ΔH_{cr} and ΔH_{fus} are quite different so further study is needed.	1. https://www.sciencedirect.com/science/article/pii/S0378775303005147?via%3Dihub
Linear/volume expansion coeff.	These values vary a lot because they are correlated with too many factors such as orientation of polymer, chain length, method of testing, etc. Therefore it may not be appropriate for ML.	Phys. Rev. B 58, 8416 (1998) - Thermal expansion of polymers: Mechanisms in orthorhombic polyethylene
Thermal conductivity	TC is related to the morphology of polymer chains. More amorphous region reduces thermal conductivity.	Thermal conductivity of polymers and polymer nanocomposites - ScienceDirect
Thermal decomposition temp.	TDT is related to the bond energy between atoms in the polymer. Usually polymers with aromatics and more cross links increase TDT.	1. Dependence of Thermal Stability of Polymers on Their Chemical Structure - IOPscience 2. Effect of structure on the thermal decomposition of polymers

Specific heat capacity (constant pressure)	Heat capacity is related to the degree of freedom in the structure. Amorphous polymers (lower crystallinity) generally have higher heat capacity.	Motion in Polyethylene. I. Temperature and Crystallinity Dependence of the Specific Heat | The Journal of Chemical Physics | AIP Publishing
Specific heat capacity (constant volume)	Similar to heat capacity at constant pressure because volume expansion coefficient of polymer is relatively small.	Motion in Polyethylene. I. Temperature and Crystallinity Dependence of the Specific Heat | The Journal of Chemical Physics | AIP Publishing
Electric properties		
Dielectric constant	High DC is essential for the dissociation of Li salt and high DC generally increases ion conductivity.	Effects of Ion Size and Dielectric Constant on Ion Transport and Transference Number in Polymer Electrolytes
Surface resistivity	Should be similar to volume resistivity b/c p=RA/t, where t is thickness. So surface resistivity and volume resistivity only have differences in thickness.	Surface vs. Volume Resistivity
Volume resistivity	Overall, higher molecular mass, lower density, lower crystallinity (which is critical for ion conductivity), and higher branching degree enhances volume resistivity. Therefore, volume resistivity may be associated with ion conductivity.	1. The Influence of Morphology on the Electric Strength of Polymer Insulation | IEEE Journals & Magazine 2. Influence of different test conditions on volume resistivity of polymeric insulated cables and polyethylene samples
Physicochemical properties		
Gas diffusion coeff.	Mobility of ions in polymers is somewhat analogous to ion movement (1). Also, the substitution of a bulkier or more rigid functional group in a polymer generally inhibits the rotational mobility of the polymer chains, which is commonly reflected by increased T_g (critical for ion conductivity), which reduces the diffusivity of gasses (2).	1. Increasing the conductivity of crystalline polymer electrolytes | Nature 2. Diffusion of Gases in Silicone Polymers: Molecular Dynamics Simulations

Gas permeability coeff.	Permeability partially depends on diffusion: P=KD/x.	A formula for permeability
Gas solubility coeff.	Solubility is strongly related to free volume which is essential for ion transfer.	1. Theory of solubility of gases in polymers - Xie - 1997 2. Decoding Polymer Architecture Effect on Ion Clustering, Chain Dynamics, and Ionic Conductivity in Polymer Electrolytes | ACS Applied Energy Materials
Interfacial tension	Interfacial tension is measured for polymer blends/mixtures so it is not a physical property for a single polymer.	
Surface tension	Surface tension is related with the chemical composition of polymers which can be estimated by the atomic composition and molar volume.	Surface Tension of Polymers