

Supervised Classification of Polymers for Lithium Metal Battery Solid-State electrolytes

Simbarashe Makahamadze



A dissertation submitted to the University of Zimbabwe in partial fulfilment for the
requirements of a Bachelor of Science Honours in Chemistry Degree

Faculty of Science, Department of Chemistry and Earth Sciences

December 2022

Supervised by

Dr K Zvarevashe

Head of The Department of Analytics and Informatics

THIS PAGE INTENTIONALLY LEFT BLANK

Abstract

Lithium metal batteries (LMBs) have the potential to pave a path towards the widespread adoption of affordable electric vehicles and robust renewable energy ecosystems. But, due to inherent shortcomings, LMBs have not been able to fulfil their promise. Developing practical solid-state lithium-ion conductors is crucial to realizing practical LMBs with improved safety, performance, and cost.

To serve as an electrolyte within a cell, an ideal solid-state lithium-ion conductor must simultaneously satisfy several technical and non-technical requirements - high ionic conductivity at ambient temperature, a wide electrochemical stability window, good thermal and mechanical stability, excellent interfacial contact with both electrodes, and a high lithium transference number. Finding a material that satisfies all these requirements has proven to be a non-trivial task. And decades of empirical research are yet to yield a stable material capable of rapidly conducting lithium ions at ambient temperature.

The work described in this dissertation highlights the successful application of a data-driven approach to identifying potential polymers for meeting the goal of a lithium battery solid-state electrolyte host material. In this study, data on dry, additive-free solid polymer electrolytes is adopted and used to develop five statistical learning classification models. The best-performing model, a small neural network with one hidden layer, is used to identify many new polymers from a large database of polymers; the PIIM database. The models developed here, performed fairly well as measured by the recall, precision, accuracy, and F1 score and better than random search as indicated by the area under the receiver operating curve.

Ultimately, the results produced in this study demonstrate the utility of statistical learning models on small data sets and reveal hundreds of new variants of well-known polymers and hundreds more that hitherto, have not been investigated.

Contents

Table of Contents

Abstract	3
Contents	4
List of Figures.....	6
List of Tables	6
Glossary	6
Chapter 1	9
<i>Introduction.....</i>	<i>9</i>
<i>Pre-Literature Review</i>	<i>10</i>
<i>Problem Statement</i>	<i>11</i>
<i>Objectives and Outline</i>	<i>13</i>
Chapter 2	15
<i>Solid-State Electrolytes</i>	<i>15</i>
<i>Polyinformatics.....</i>	<i>16</i>
Chapter 3	20
<i>Motivation.....</i>	<i>20</i>
<i>Dataset Development.....</i>	<i>20</i>
<i>Computer Representation of Molecules</i>	<i>21</i>
SMILES.....	21
Molecular Fingerprints.....	22
<i>Supervised Statistical Learning.....</i>	<i>22</i>
<i>High-throughput Virtual Screening</i>	<i>23</i>
Chapter 4	24
Chapter 5	30
<i>Overview.....</i>	<i>30</i>
<i>Oxyethylene.....</i>	<i>31</i>
<i>Polyphosphazene.....</i>	<i>31</i>
<i>Polysiloxane.....</i>	<i>32</i>
<i>Aromatic.....</i>	<i>33</i>

Chapter 634

Chapter 736

References38

List of Figures

1. 1. A plot of the atmospheric carbon dioxide concentration since 1750.
1. 2. Schematic of a lithium-ion battery.
1. 3 Schematic of a lithium-metal battery.
- 4.1 A plot of 5 evaluation metrics for each classification model
- 4.2 A sample of 8 repeat unit structures

List of Tables

1. 1 Table 1. Main requirements for lithium batteries.
- 4.1 A classification report for each model
- 4.2 A sample of 8 SMILES strings for each repeat unit in table 4. 1
- 4.3 A sample of SMILES strings obtained from a recent study

Glossary

AI Artificial Intelligence

AUC Area Under the Receiver Operating Curve

CSV Comma Separated Values

KNN K-Nearest Neighbours

IUPAC International Union of Pure and Applied Chemistry

ISE Inorganic Solid Electrolyte

LIB Lithium-ion Battery

LMB Lithium Metal Battery

ROC Receiver Operating Curve

SMILES Simplified Molecular Input Line Entry System

SPE Solid Polymer Electrolyte

SVM Support Vector Machine

SOC State of Charge

ML Machine Learning

MEEP poly [bis (methoxy-ethoxy-ethoxy) phosphazene

PEM Proton Exchange Membrane

PEO Polyethylene Oxide

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

Before the Industrial Revolution, the earth's carbon cycle was roughly in equilibrium. The average temperature on earth was ideal, and the concentration of carbon dioxide was around 280 ppm^{1,2}. Then, human activities during the course of the Industrial Revolution began to introduce a great deal of extra carbon dioxide into the atmosphere. For decades this seemingly innocuous additional carbon dioxide gradually accumulated in the lower atmosphere. Up until it started perilously enhancing the greenhouse effect that warms the planet, and without which we could not survive on this planet. The amplified greenhouse effect has been slowly warming up the whole world and noticeably altering our climate system in irreversible and unprecedented ways^{1,3}, bringing about global warming and climate change. Today, global warming and climate change are among the biggest challenges we have ever faced, and the effects of climate change are already apparent as record droughts, heatwaves, wildfires, floods, and extreme weather events devastate communities worldwide⁴⁻⁶. But, things are poised to get worse if greenhouse gas emissions continue to soar unabated^{7,8}.

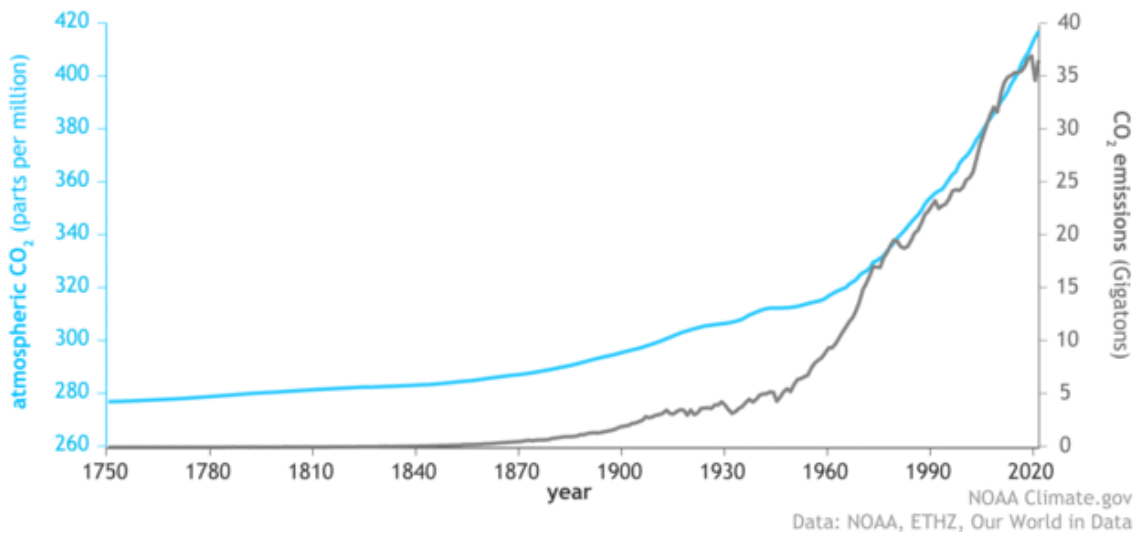


Fig 1. 1. A plot of the atmospheric concentration of carbon dioxide since 1750, shows a steady increase of CO₂ that became much steeper during the 20th century.

While a range of human activity produces greenhouse gases, the lion's share of anthropogenic emissions comes from the burning of fossil fuels for energy^{9,10}. This alone accounts for approximately three-quarters of global greenhouse gas emissions^{1,2,9}. Therefore, to cut greenhouse gas emissions and stand a chance in stopping global warming and staving off the worst effects of climate change, the whole world needs to transition to sustainable

energy, avoid burning fossil fuels, and sort out alternative ways of acquiring energy without emitting greenhouse gases.^{6,11}

Renewable energy holds tremendous potential in this regard. A virtually inexhaustible and environmentally friendly source of energy. More importantly, renewable energy has become more and more economically viable. Today some of the cheapest electricity generated is from solar and wind¹². The biggest drawback to renewables, however, is their intermittency. The wind does not always blow, and the sun does not always shine, and yet our energy needs are incessant. Therefore, energy storage is undoubtedly required to complement renewables, to offer a steady supply of energy. That energy storage solution could very well be a lithium battery¹³⁻¹⁵.

Pre-Literature Review

Modern-day lithium-ion battery cells¹⁶ are made up of two host structures or electrodes that accommodate lithium ions - an anode and a cathode. These electrodes are connected to an electric circuit through a copper and an aluminium current collector respectively. To prevent them from coming into contact with each other, and the resultant short circuit, the two electrodes are physically separated with a microporous plastic called a separator. This separator electronically isolates the electrodes whilst allowing lithium ions to pass through. These three components are integrated and encased into a tightly sealed steel casing. Lastly, a fourth component, an organic liquid electrolyte is injected into the cell. The all-encompassing electrolyte is the medium through which lithium ions move as they get shuttled from one host structure to the other, during recurrent charge and discharge cycles.¹⁷⁻¹⁹

In the as-manufactured state, all of the lithium ions are intercalated, within the cathode^{17,20,21}. During the charge cycle, electrons are driven from the cathode and moved through the external circuit to the anode. As the cathode material is being oxidised, lithium ions are de-intercalated from the structure of the cathode and dissolved into the ubiquitous liquid electrolyte, to maintain electrical neutrality. As this process transpires, lithium ions begin to slowly diffuse from the cathode side of the electrolyte through the porous separator to the less concentrated anode side, where they intercalate into the anode structure^{17,20-22}. This series of events is reversed during the discharge cycle; electrons are released from the anode via a spontaneous chemical reaction, and they flow freely through the external circuit into the load - where they do useful work - and eventually return to the cathode. Meanwhile, internally, lithium ions redissolve into the electrolyte and diffuse back to the cathode through the separator. Lithium ions are repeatedly shuttled from one electrode to the other during successive charge and discharge cycles throughout the cell's lifetime. This is known as the 'rocking chair' mechanism.

The earliest, working lithium batteries were developed by The Sony Group Corporation in the early 1990s. They used a 'rocking chair' mechanism, a graphite anode, a lithium cobalt oxide, Li_2CoO_2 cathode, and an organic liquid electrolyte; lithium hexafluorophosphate, LiPF_6 , dissolved in a mixture of propylene and ethylene carbonate.^{25,27,30,28}

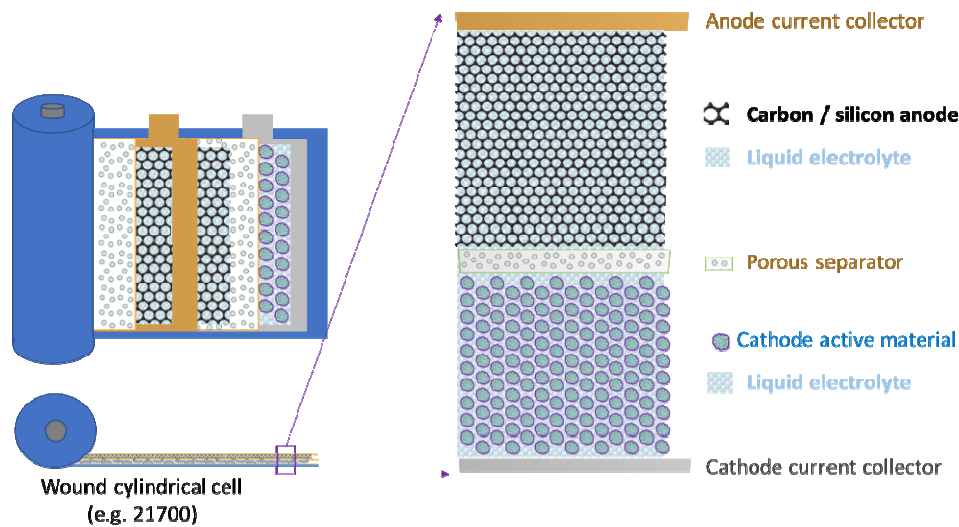


Fig 1. 2. Schematic of a lithium-ion battery showing how the internal structure is wrapped into a steel housing.

Problem Statement

Through its high potential, energy density, and capacity, the LIB has since, helped to improve our lives, and will arguably continue to do so in the coming years^{21,25,26}. But, despite the remarkable success and progress of LIBs over the years, they still suffer from many shortcomings.^{15,25,27-34}

Table 1. Main requirements for lithium batteries.^{22,26,30}

Battery Attributes	Main Requirements
Energy Density	>750 Wh/L and >350 Wh/kg for cells
Cost	< \$100/kWh for cells
Fast Charge	80% SOC in 15 minutes
Life	15 years
Performance	Minimum impart by environment
Safety	No fire, flame, rupture or explosion

Most of the safety problems, and limitations that have stymied lithium-ion batteries primarily stem from the electrolyte^{20,35-37} - lithium salt dissolved in an organic solvent. The organic solvents used to make liquid electrolytes are not particularly stable around the very low voltages at which ideal anode materials operate. This situation consigns the electrolyte to

inevitable decomposition where the two come into contact. A very complex passivating solid electrolyte interface is formed as a result.^{20,35,38} Although this passivation layer plays an important role in thermodynamically stabilizing the cell by protecting the electrolyte from reacting any further with the anode, it also irreversibly consumes lithium ions, which contributes to capacity fade, a curtailed lifetime, and a safety hazard^{34,39} when it eventually decomposes via an exothermic reaction.

Organic liquid electrolytes are also volatile, and very flammable; easily catching fire under thermal runaway conditions. A combination of these factors, as well as others, is responsible for a litany of safety issues, and constraints that handicap modern lithium-ion batteries.

Although modern-day lithium-ion batteries work well enough for most purposes, among battery experts, the consensus is that the immediate predecessor of the LIB, the lithium metal battery LMB, is the key to unlocking the full potential of lithium-based batteries and enabling the electrification of transportation and industrial processes, and sustainable energy ecosystems. The main advantage of this cell over a LIB cell is that in this type of cell the electromotive force, and capacity is considerably larger. However, this configuration is currently deemed too impractical because, as in the early days of battery research, it is still considered unsafe.^{21,40}

Despite their earlier demise,²¹ LMBs remain the Holy Grail of battery chemistry^{34,35,40,41}, with an anode of pure lithium metal,³⁴ LMBs have an inherently high energy density, a more elegant architecture, and are possibly cheaper and easier to manufacture.^{13,40,41} All things considered, LMBs could pave the way for grid-energy storage^{13,14,43,44} and affordable electric vehicles⁴⁵⁻⁴⁷; that travel much farther on a single charge, and are capable of fully charging within minutes. This is the promise of LMBs.^{13,35,41,42}

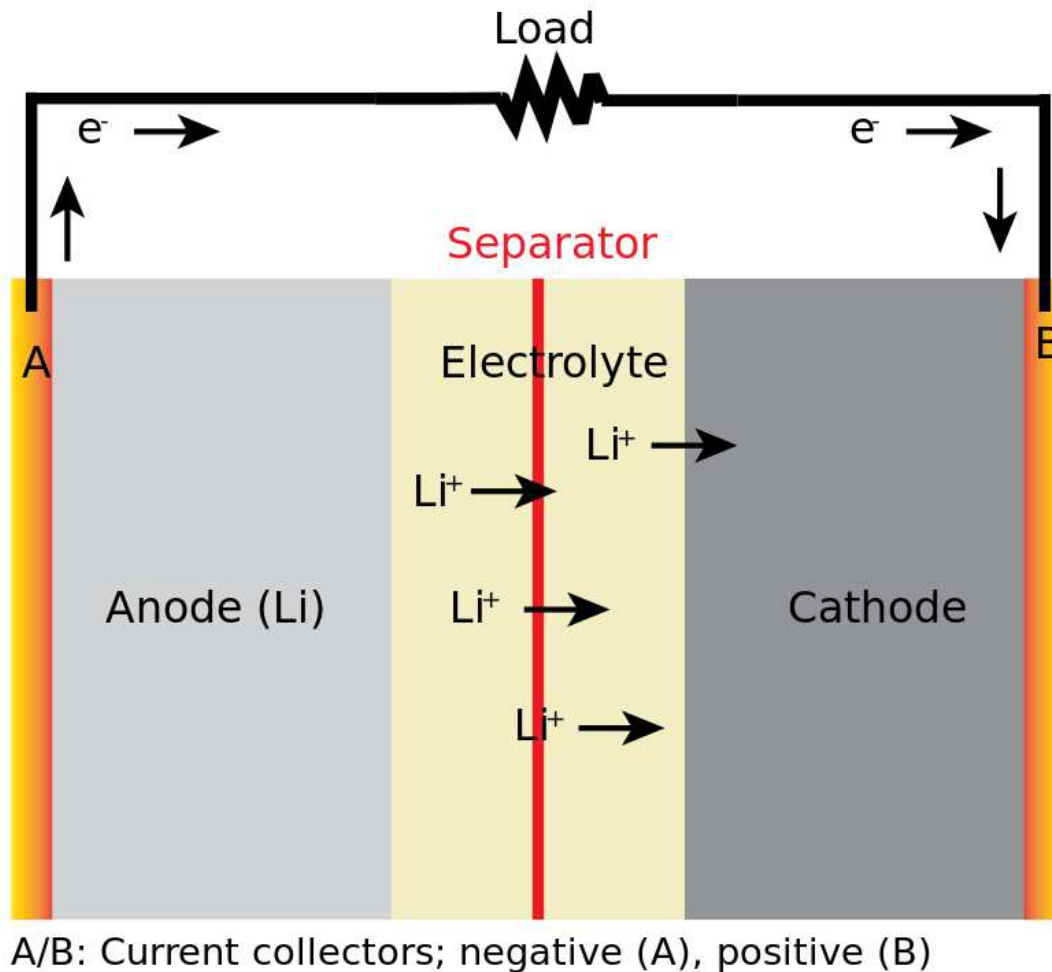


Fig 1. 3. Schematic of a LMB. The left-hand side shows an anode of pure lithium metal as opposed to a host structure that would normally be found here in a LIB. The electrolyte which also serves as a separator in this configuration sits between the two electrodes, preventing electrical contact.

Objectives and Outline

The use of lithium metal in practical rechargeable batteries is acceptable only if the safety of operation is ensured. This essential condition can be achieved by a complete replacement of a conventional liquid electrolyte with a solid-state electrolyte.^{34,48–50}

Yet, finding an ideal solid-state material, organic or inorganic, that satisfies all the requirements needed to enable a practical LMB cell has proven to be a difficult optimization problem. And much like any other material, to a large extent, the development process has been an experimentally driven trial-and-error process that is guided more by intuition, experience, and conceptual insight than by data and rational design. This approach has proven to be inefficient, expensive, lengthy and painstaking.^{51–54}

The primary objective of this study is to use a data-driven approach to identify novel polymers capable of meeting electrolyte requirements, from a large and diverse database of virtually a million, real and hypothetical. This approach involves preparing a small database of solid polymer electrolyte data, and building, evaluating and applying statistical learning models as described in Chapter 3. Chapter 2 provides context for this problem through a literature review on solid polymer electrolytes, statistical learning models and the use of statistical learning models to analyse solid polymer electrolytes. Chapter 4 describes the results obtained from applying the best-performing model on a database of previously untested polymers. Chapter 5 discusses the results in a high level of detail and contrasts these results with those obtained from closely related research. Chapters 6 and 7 provide a broad outlook, recommendations and future steps.

Chapter 2

Solid-State Electrolytes

Numerous types of solid-state electrolytes have been investigated for lithium metal battery applications in the past. Broadly, solid-state electrolytes fall into three classes; inorganic ceramics, organic polymer electrolytes, and composites.^{34,42,49,55,56} The first two types of materials may be combined to form composites. Composites attempt to leverage the advantages of both classes while reducing their shortcomings.^{57,58} Despite facing significant challenges,^{35,59} many researchers from academia, government, and several industries are motivated and devoted to developing and commercializing solid-state electrolytes because of the appealing opportunities^{35,56,60} would open up, as discussed in Chapter 1.

The most popular class of solid-state electrolytes is the inorganic class of materials, which consists of oxide and sulphide compounds.^{42,49} But more recently halides as well as borohydrides.⁴⁹ Inorganic materials typically exhibit two or three excellent properties. For instance, the ionic conductivity of some inorganic electrolytes is comparable to conventional liquid electrolytes⁶¹. In the case of sulphide compounds $\text{Li}_{10}\text{GeP}_2\text{S}_{12}$ (LGPS)^{50,62} and $\text{Li}_{6.75}\text{La}_3\text{Zr}_{1.75}\text{Te}_{0.25}\text{O}_{12}$ (LLZTO),⁴² the ionic conductivity equals or exceeds⁴⁹ that of standard commercial non-aqueous liquid organic electrolytes, but lack considerably in the other important metrics. For instance, LGPS is unstable against both high-voltage cathodes and low-voltage anodes, succumbing to both reduction and oxidation reactions respectively while in contact with both electrodes.

Other typical examples of inorganic solid-state electrolytes (ISEs) include NASICON, $\text{Li}_{1+x}\text{Al}_y\text{Ge}_{2-y}(\text{PO}_4)_3$ (LAGP),⁵⁰ and $\text{Li}_{1+x}\text{Al}_x\text{Ti}_{2-y}(\text{PO}_4)_3$ (LATP)⁴²; borohydride type LiBH_4 ³⁵; LISICON type $\text{Li}_{14}\text{Zn}(\text{GeO}_4)_4$; garnet type $\text{Li}_7\text{La}_3\text{Zr}_2\text{O}_{12}$ (LLZO)^{35,40,42}; perovskite type $\text{Li}_{3x}\text{La}_{2/3-x}\text{TiO}_3$ (LLTO)^{13,42,50}, anti-perovskite type $\text{Li}_{2.95}\text{OCl}_{0.95}$ ⁵⁰, argyrodite type $\text{Li}_6\text{PS}_5\text{X}$ where $\text{X}=\text{Cl}, \text{Br}, \text{I}$ ⁵⁰. Unfortunately, the effective use of ISEs is still limited primarily by low ionic mobility, a limited electrochemical window of stability or both depending on the class of the material.

On the other hand, polymers were first proposed for electrochemical device applications in the late 1970s⁴⁸ by Armand⁶³ and Fenton et al.^{63,64} following the discovery of ionic conduction within poly(ethylene oxide)-alkali-metal salt complexes in 1973 by Wright et al.^{42,65}. Since 1978, PEO derivatives have been well studied and are widely considered the benchmark for SPEs that have been investigated, due to a combination of their relatively high ionic conductivity, desirable electrochemical properties, ability to form stable dry complexes, and facile processability inherent to most polyolefins.

Other polymer hosts that have been investigated include polyvinyl chloride, polyacrylonitrile, poly dimethyl siloxane, poly methyl methacrylate, polyvinylidene fluoride, polyvinyl acetate, and polyethene imine⁶⁶. Despite their diversity and different chemistries, in general, dry SPEs are solid solutions that host alkali metal salts within their polymeric matrix^{42,50}. In such

double ion-conducting electrolyte systems, it is widely accepted that the lithium ions are dissolved, complex, and transported throughout the polymer matrix under the influence of highly electronegative heteroatoms and polar groups⁶⁷.

To date, no single-ion nor double-ion conducting polymer possesses sufficient conductivity to rival conventional liquid electrolytes^{48,63,68}. But, due to renewed interest in high-performance solid-state LMBs, driven by demand for more affordable modern electric vehicles^{54,69–71}, tremendous research efforts have been, and are presently underway in academia and industry to develop viable alternatives to PEO.

In practice, finding a polymer to host and transport lithium ions, while satisfying all the required metrics^{68,72}; high ionic conductivity at ambient temperature; a wide electrochemical window; good thermal stability and interfacial contact with the two electrodes; high lithium transference; and moderate mechanical strength; has proven to be a non-trivial task. A long, expensive, and experimentally driven process that has been guided more by experience, and intuition than anything else^{73–75}.

The process of discovering and developing new materials has largely relied upon a very iterative and empirical approach which is characterized by a large and uncertain number of varied attempts which are continued until success or at least until the practitioner stops trying. This study proposes a data-driven strategy for solving the problem of identifying potential solid polymer host materials. This approach presents an appealing opportunity to uncover latent patterns, structure and insight that is not immediately apparent to humans from large collections of prior experimental data. But recently, data science and statistical learning techniques have been applied in the field of polymers and materials science in general with some success^{73,76–81}, to aid experimental work. Given the recent success^{54,82–84} of data science in the fields of materials science, using machine learning (ML) presents a compelling avenue for accelerating research and development in the field of SPEs^{85,86}.

Polyinformatics

The sheer scale of many problems in materials science makes any non-trivial materials discovery process a lengthy and difficult one. A task beyond the scope of a simple trial-and-error approach or human intuition. There are an estimated 10^{60} potential drug-like molecules, compared to 10^{80} estimated atoms in the entire universe^{54,83}. ML offers a way to efficiently and automatically explore this vast space of possibilities. In addition, this approach measures our uncertainty, pre-emptively reduces the costs associated with the initial stages of research and development and accelerates materials innovation.

Having revolutionized many fields, machine learning (ML)^{54,82–84}, a subfield of AI that deals with the development of algorithms and models that are capable of automatically learning patterns from data and performing tasks without explicitly programmed instructions⁸⁷, is beginning to make major inroads within the physical science field^{81,86–88}. DeepMind.Inc's triumph exemplifies this trend, and AI's potential to accelerate science and innovation in the field.^{82,89} Meanwhile, the research activity about the design and discovery of battery electrolyte materials using data-driven methods and machine learning has experienced a

surge. The following sections review some of these recent applications at the intersection of the fields of AI and chemistry. But generally, in any of these applications, predictive models are built using data generated from prior experimental work and used to automatically make predictions on previously untested materials.

In the data-driven search for electrolytes, the most significant works have been focused on inorganic solid electrolytes⁹⁰. For example, Sendek et al.^{53,91} trained a logistic regression model on 40 inorganic lithium-ion conductors available from literature with the hope of identifying novel superionic conductors. Despite the small size of the dataset, this group discovered that physics-based descriptors that encode structural information about the materials; average sublattice bond ionicity, the average shortest lithium-lithium distance, the shortest lithium-anion distance, and the average anion-anion coordination number in the anion framework; made it possible to train models that were capable of making good predictions on untested materials. The logistic regression model achieved an F1 score three and a half times greater than that obtained if one were to randomly guess whether given materials were a good lithium-ion conductor. This model helped reveal many new solid electrolytes by screening approximately 12, 000 known lithium-containing compounds from the Materials Project database. This model suggested 21 structures that have the potential to meet ionic conductivity and electrochemical stability requirements for solid electrolytes while using materials that are widely available, and cheap.

On a similar note, Sendek et al.⁵² go on to quantify the likelihood of finding a solid lithium-ion conductor that would be able to meet all the requirements expected of a solid electrolyte. Their study revealed that sulphide-based solid electrolytes show greater promise than oxide-based electrolytes, and that overlooked materials such as bromides and chlorides show enough promise to warrant more research. The model also proved to be more efficient and accurate than traditional screening methods in the analysis of 12 831 materials for promising candidates for solid-state electrolytes in lithium-ion batteries.

In another closely related study, Liu et al.⁹² present a data science approach to SPE development by collecting and analysing electrolyte data published in the literature, as well as supplementing this set of curated data with their independent experimentation and validation to create a comprehensive database that researchers in this study used to build six different models (linear regression, lasso regression, ridge regression, decision tree, random forest and radial basis SVM). It is worth noting that in this study the random forest model, as in many other case studies, performed best as measured by the root-mean-square error in a regression task in which the models were required to predict the ionic conductivity of SPEs. Similarly, Ahmad et al. used more sophisticated convolutional neural networks (CNNs) trained on shear and bulk moduli data to screen over 12, 000 inorganic compounds in search of solid inorganic electrolytes that are capable of suppressing lithium dendrite growth in LMBs. The group found six new stable electrolyte materials, four of which were predicted to be stable in the presence of dendrite growth while simultaneously exhibiting high ionic conductivity.⁹⁰

The work done by Zhang⁹³ et al. stands out among those at the intersection of machine learning and battery materials research because it utilizes the powerful but rarely used unsupervised learning family of algorithms, with only a very limited amount of conductivity data to identify lithium-ion conductors. Their learning scheme discovered 16 new fast Li-conductors whose conductivities were validated using Ab initio molecular dynamics simulations. The most striking and distinctive underlying theme of this study was the ability demonstrated by the unsupervised model to reveal structures and chemistries that were truly novel in the sense that they differed considerably from known inorganic compounds that had been investigated before this study.

The work of Xie⁹⁴ et al. comes across as eye-catching for the simple reason that, an integrated deep learning system is developed to be able to work with any arbitrary solid material as opposed to the usual approach of focusing on just a single specific type of material.

Hatakeyama-Sato⁹⁵ et al. constructed a database of 240 lithium-ion conducting solid polymer electrolytes from scratch and applied gradient boosting deep neural network, and support vector machine regression models on the task of predicting the ionic conductivity of the polymer electrolytes. The gradient boosting model proved to be the best, outperforming, as measured by the coefficient of determination, the other two models. Inspired by these analyses, the researchers went on to synthesize poly (glycidyl ether) derivatives which exhibited considerable ionic conductivity as predicted.

Very recently, Schauser⁹⁶ et al. have described the curation, visualization, and analysis of a database of solid polymer electrolyte conductivity performance, manually extracted from published literature. This study provides new insight and reveals the redundancy of some presumably relevant features that are supposedly important when it comes to predicting the conductivity performance of an SPE. And through visualization of the collective data set, the study also suggested that individual features, apart from activation energy, are poor predictors of conductivity performance across the broad range of polymer families. They also used statistical regression techniques to extract the most important features; polymer molecular weight, glass transition temperature, the existence of electronegative heteroatoms in the monomer, and anion size; that is relevant to determining lithium-ion conductivity. The group also asserts that experimental features can be omitted without compromising the predictive performance of the regression model. This finding bodes well for this study in which I intend to omit experimental features entirely and only make use of the chemical structure of the polymers' repeat unit.

Venkatram⁶⁶ identified 13 relevant properties pertinent to SPE performance, trained Gaussian process regression models based on 5 proxy properties of the aforementioned 13, and identified 16 new SPE candidates by screening a large dataset of 13, 388 experimentally synthesizable polymers which included polyphosphazenes, and fluorine-containing monomers. Venkatram went on to build a deep learning framework that predicts good solvents and non-solvents for the synthesis of new polymers.

While efforts have been made to approximate and predict properties of solid-state electrolytes using regression models^{77,97–99}, past work has not fully focused on identifying potential polymers based on merely monomer chemical composition and structure.

Chapter 3

Motivation

Historically the process of developing new materials is very prolonged⁷³; typically taking an order of 10 to 15 years^{51,54}. This development process has been, to a large extent, an iterative, experimentally-driven, trial-and-error process that relies on upon, and is primarily guided by individual intuition, personal experience, and conceptual insight. Exploring the enormous materials space through this conventional research channel has been largely inefficient, and in most cases unsuccessful. The same can be said for battery materials which have to satisfy multiple conflicting parameters simultaneously.

This chapter aims to describe an attempt to use supervised statistical learning models to identify potential solid polymer electrolyte host materials, for lithium metal batteries, from an overwhelming sea of potential candidates using a relatively small database of historical solid polymer electrolytes.¹⁰⁰

Dataset Development

The data prepared and used in this study was originally compiled by researchers from the Materials Research Laboratory at the University of California, Santa Barbara^{101,102}. This data was manually extracted from 65 publications. Spanning over a wide range of homopolymers and copolymers, this data set contains 655 unique, solvent and additive-free, solid polymer anion-salt, entries with approximately 305 features.

Before applying statistical techniques to this data set, it is necessary to transform the data into a form that would best serve the purposes of this research. Using Python's Pandas library in Jupyter Notebooks the SMILES¹⁰³ descriptor column, polymer name, and conductivity measurements at 30 and 60 °C are isolated. The logarithm to the base 10 is then applied to both conductivity measurements. Entries with missing conductivity measurements in both columns are omitted.

This round of filtering reduced the dataset samples from the initial 655 to 271 entries. The dataset is then sorted according to conductivity measurements and only the best-performing entry is retained. Using a PEO benchmark calculated using all the data available on the PEO system in the dataset. The remaining polymers are classified according to this benchmark. A positive label is assigned if the entry has a conductivity greater than the benchmark and a negative label otherwise. Approximately 70% of the polymers were assigned positive labels and the rest were assigned negative labels. And the final dataset contained 52 entries.

In summary, the original UC Santa Barbara dataset with 655 entries and 305 columns is transformed into a much smaller dataset of 52 entries and 2 columns; a SMILES descriptor column and a corresponding label column.

Computer Representation of Molecules

To use statistical learning algorithms, data on molecules that are representative of the problem domain is needed. Furthermore, it is necessary to represent the molecules in a form that is meaningful to the computer – a form that enables the computer to understand and make sense of the data. With the former achieved, here I discuss how we achieve the latter, namely through the SMILES notation. A basic representation used to convert actual molecules into informative and computer-amenable representations.

SMILES

A representation determines how a molecule is translated into a mathematical description that a computer can handle.¹⁰⁴ Much of chemoinformatics depends on the quality of this representation. The representation determines what information is included or excluded, and how the models built using that representation will perform. Therefore, the representation must be an accurate embodiment of the molecule it represents. Several methods exist for representing molecules. Some representations are only human readable, (name, chemical formula, and visual structure), while some are only computer-readable (adjacency matrix), and yet a few are both human and computer readable (non-linear graphs and linear notation).

The Simplified Molecular Input Line Entry Specification (SMILES)^{103,105} is a linear notation designed to be human and computer-readable. The basic syntax of SMILES converts the molecular structure into a sequence of characters in which hydrogen atom presence is implied. The rest of the atoms are represented by their respective atomic symbols, while chemical bonds are represented by, the equals sign (=) for a double bond, the hash symbol (#) for a triple bond, and (:) for an aromatic bond. When neither of these symbols that signify a chemical bond is present, a single chemical bond is implied. Atoms in aromatic rings are represented by the lowercase letters of the element's atomic symbol. And those within cyclic structures are numbered to indicate their point of connection. Branches are represented using a pair of parenthesis, while cyclic forms are represented using numbers or special characters (\$^). The SMILES notation follows the normal valence rules for atoms, therefore standard IUPAC valences are used to fill in missing hydrogen atoms. For instance, carbon has a standard valency of 4, if a carbon atom within SMILES is bonded through three explicit bonds, SMILES syntax implicitly assumes that an additional hydrogen atom is bonded to this carbon. Furthermore, non-common atoms other than Boron (B), Carbon (C), Nitrogen (N), Oxygen (O), Phosphorus (P), Sulphur (S), Fluorine (F), Chlorine (Cl), and Bromine (Br) are incorporated by being written inside square brackets, for instance [Si] for Silicon. Formal charges are also specified within square brackets.

SMILES are commonly used due to their compact form and ease of use. Alternatively, a new text-based structural representation system, BigSMILES, which is an extension of SMILES notation, designed to be more suited for polymers could also be used because it can capture the inherent nature of polymers better than regular SMILES.¹⁰⁶ Despite the obvious benefits of using this form, this new notation is not yet compatible with common scientific computing libraries. All in all, SMILES strings are handy, but they are not of much use to computer algorithms because they are incapable of encoding sufficient structural and chemical

information to allow meaningful relationships to be inferred. To make this possible, it is necessary to convert this linear form into more informative descriptors.

Molecular Fingerprints

Molecular fingerprints¹⁰⁷ are highly effective algorithms capable of encoding the structural, chemical, stereochemical, and physical properties of molecules.¹⁰⁸ Initially developed for substructure searching, molecular fingerprints have found use in clustering, classification, and determining similarity between molecules. Now they have found a place in modelling polymers because they are capable of capturing molecular features relevant to molecular activity, and encoding positive and negative structural information, frequency of occurrence, and substructures.

Using the data set prepared earlier, which consists of two columns – a SMILES column and a corresponding label indicating the probability of the particular polymer exhibiting solid polymer electrolyte characteristics. Each polymer's SMILES string is transformed into a sparse one-dimensional vector, using the Morgan algorithm with a radius of 3 and 512 bits in length. It is worth noting that different lengths of the fingerprint were tried and, there are no significant differences in predictive performance between bit fingerprint vectors of lengths 64, 512, 1024, and 2048. Fingerprint vectors of small size retain enough information to be safely folded into diminishing sizes without compromising model performance.

This data represents the final form of the data set used for statistical learning purposes. Each row entry consists of a molecular fingerprint vector for each polymer SMILES string and the corresponding target label appended at the end.

Supervised Statistical Learning

The use of machine learning algorithms on polymers is predicated on an underlying correlation between the structure of the monomer and a target property. This mathematical relationship can be inferred using statistical learning techniques that attempt to correlate the descriptors in each molecule with a corresponding target label.

Machine learning, a subfield of artificial intelligence consists of a large set of learning algorithms that enable computers to learn the underlying patterns that correlate molecular properties to target labels. For this study, 5 different statistical learning algorithms are trained using the sklearn library in Python. A neural network, a decision tree, a random forest, a support vector machine, and k-nearest neighbours.^{88,109,110} Functions from the sklearn library are used for splitting the data on each iteration and finding the best hyperparameters for each model. The models were trained on a training set of 32 samples and tested on the remaining set of 20 samples through a hundred iterations. The results obtained from these experiments are discussed in detail in chapter 4.

High-throughput Virtual Screening

The PI1M database is an open-source benchmark database with approximately one million polymers in the form of SMILES strings. This database was created for polymer informatics research using a generative machine learning model trained on more than 12, 000 polymers that have been manually compiled from the PoLyInfo database.¹¹¹ PoLyInfo itself is the largest existing database of real polymers. The PI1M database has the advantage of having orders of magnitude more data than the PoLyInfo database, covering a similar chemical space but with even more examples where the PoLyInfo database is sparse. This combination of factors produces a near-perfect pool for screening candidates suited for the goals of this study.

I used the best model, a neural network, to classify polymers in the PI1M database in a high-throughput screening manner. Each SMILES string in the PI1M database is converted into a corresponding molecular fingerprint with a width of 3 and a length of 512 in the same manner as the polymers used to train the statistical learning algorithms. Each fingerprint is then fed into the neural network. A positive label is assigned if the probability output is greater than 50%. The model assigned positive labels to 1,611 polymers as solid-ion conductors.

Chapter 4

This chapter describes the results obtained when 5 machine learning classification models - neural network, decision tree, random forest, SVM and KNN – were tasked with classifying repeating units in the PIIM database into one of two classes based on a PEO benchmark¹¹¹. Each model was created as described in Chapter 3 and assessed using five evaluation metrics following best practice¹¹⁰. And the final measurement is obtained by averaging the performance results from 100 iterations of randomly-sampled training and testing sets. The data used for training and testing, following preprocessing as described in Chapter 3, consisted of a total of 52 samples. The input is a molecular fingerprint of each repeating unit and the corresponding label is a classification designated based on a PEO benchmark⁶⁷. Each iteration involves randomly splitting the data set into a training set of 32 samples and a testing set of 20 samples, which represents a 70% training set and a 30% testing set split. The performance measurement for each evaluation classification metric is calculated based on the results obtained from applying the model to these randomly-sampled training and testing data sets.

From these experiments, the neural network emerged as the best-performing model as shown in Table 4. 1 and more clearly in Fig 4. 1, as measured by the recall, precision, accuracy, f1 score, and the area under the receiver operating curve.

The results described here were obtained from only 52 samples in total. This small dataset size is not unusual in materials science but is undesirable nonetheless. Although, there have been studies that have obtained impressive results with as little as 40 samples^{53,91}. The molecular fingerprints^{107,108,112} used as input are long binary vectors which encode 512 different physical and chemical properties. A large number of features within the fingerprint vectors coupled with the small data set size threatened to consign the models to overfitting as discussed later. Recursive feature elimination and analysis of the 512 features indicated that the majority of the descriptors within the feature vectors are irrelevant. This implies that most of the features consumed by the learning algorithm were not particularly important for the task of predicting lithium-ion conducting ability. And that the learning algorithm would benefit from their elimination. Of the 512 descriptors in the feature vector, only 43% were deemed relevant.

Fig 4.1. shows the overall performance of each learning algorithm as measured by the accuracy, recall, precision, f1 score and area under the receiver operating curve. All the horizontal bars for each model appear above the 50% standard mark. Based on this evidence all 5 models demonstrate genuine predictive ability. The horizontal bar for the area under the receiver operating curve indicates just how better the model is performing as compared to random guessing.

The neural network and support vector machine models performed markedly better than the random forest, decision tree and k-nearest-neighbours models. This result is typical because we would expect more sophisticated learning algorithms, the neural network and SVM, to perform better than simpler learning algorithms^{92,113,114}. The neural network and SVM were closely matched and were and shoulders above the other three learning algorithms.

The best-performing model, however, was the small feedforward neural network. The model had just one hidden layer. The model obtained the highest scores in each evaluation metric. The model obtained an accuracy of 75%, an f1 score of 75%, a precision of 80%, and a recall of 75%. The neural network used a ReLU activation function, a hidden layer with 512 nodes, an adaptive learning rate and 5, 000 iterations. However, the neural network was one of the slowest models, managing to only better the random forest. The model narrowly outperformed a radial basis function SVM. The SVM obtained an accuracy score of 74.6%, a precision of 78.2%, a recall of 74.6% and an f1 score of 74.4%. It is also worth noting that the support vector machine did perform better than the neural network on a speed basis.

Unsurprisingly, the k-nearest-neighbours model performed worst on all performance metrics. Despite coming in last on this data set, the model performs better than the standard 50% benchmark on each performance measure. Therefore this model is still expected to perform well enough to achieve predictive abilities that normally would not be possible due to resource, time, and human error constraints. The decision tree and random forest models were seen to perform significantly better than the KNN model. The random forest performs slightly better than the decision tree on recall, precision, f1 score, the area under the receiver operating curve and accuracy, but loses out on speed.

Table. 4.1. Summary of model performance

Model	AUC	Accuracy	Precision	Recall	F1 Score
Random forest	0.7038	0.6885	0.7444	0.6885	0.6852
k-nearest-neighbours	0.5750	0.6154	0.6026	0.6154	0.6053
Neural network	0.7605	0.7499	0.7959	0.7499	0.7468
Support vector machine	0.7549	0.7461	0.7882	0.7461	0.7443
Decision tree	0.6703	0.6669	0.6046	0.6669	0.6622

Figure. 4.1. A plot of model performance.

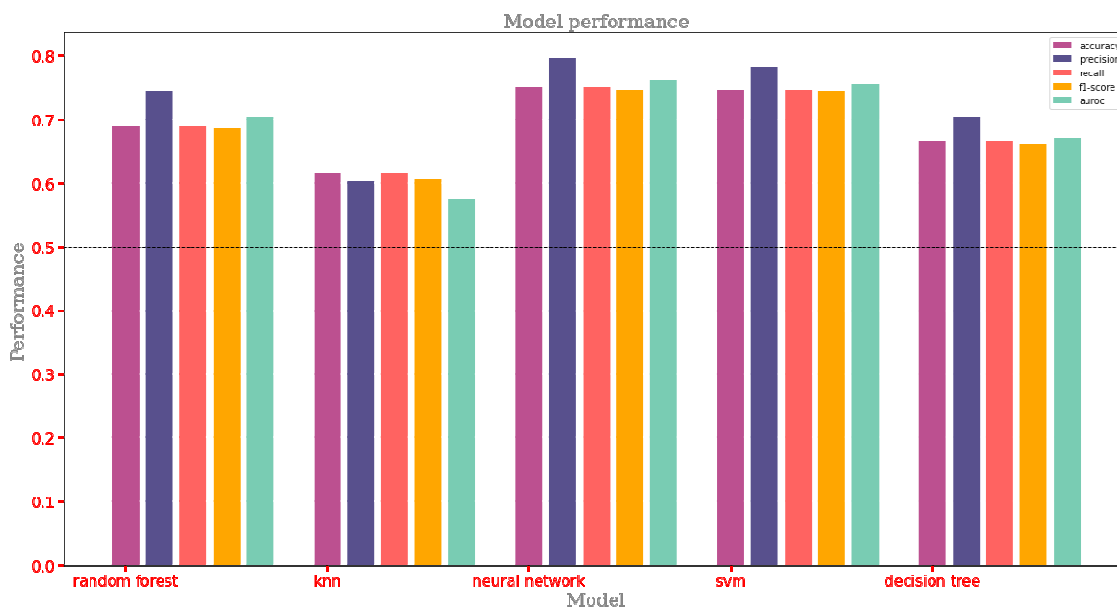


Figure 4.1: A plot of 5 performance metrics for 5 different machine learning models developed in Chapter 3 - random forest, k-nearest-neighbours, neural network, support vector machine and decision tree.

The results described here were obtained after applying the neural network model, on approximately one million polymers from the PI1M polymer database, using the polyethylene oxide (PEO) system as a benchmark for classification. The PEO system as described in Chapter 2⁶⁷, is an archetypal solid polymer electrolyte. In this classification context, a positive label implies that the repeat unit's SMILES string is expected to possess a high probability of leading to a corresponding polymer that exhibits characteristics that are desirable for solid electrolyte applications. On the contrary, a negative label would imply the opposite is true – meaning there is less likelihood that the polymer would lead to a suitable solid polymer electrolyte.

The neural network model assigned 1611 positive labels from the 995,800 polymers available in the PI1M database. This set of 1611 polymers consists of a diverse set of polymer families, all possessing polar atoms, groups and substructures in one form or the other. The most common polar groups contained either oxygen or nitrogen. Most of these polymer families, namely polysiloxanes, polyphosphazenes, fluorine-containing polymers, and polyethers are familiar and have been extensively studied and discussed in Chapter 2^{48,58,72,115–117}. The list of polymers also consists of some polymers that have received less attention according to the literature. The list of obscure polymers includes varied polymer

chemistries, including but not limited to polymers containing the carbonyl group, [PH] functional group, lead, calcium, sulphur and bromine-containing polymers, and many heterocyclic polymers. Polycarbonates¹¹⁸ are among the most promising host materials for solid polymer electrolytes. Despite the large number and diversity of the list of polymers, it is worth noting the absence of the polycarbonate family.

The oxygen atom is present in 98% of the 1611 potential polymers. The most prevalent and arguably most important functional group in this list is the larger oxyethylene group. This functional group was present in 1295 polymers. Representing more than 80% of the list of positively labelled polymers.

The polyether group is the second most common functional group in this data set. Its presence accounts for 1142 polymers, which represents approximately 71% of the list of polymers. The polyphosphazene functional group was present in 645 polymers, which accounts for 40% of the list. The list contains 341 polymers from the polyester family. Polysiloxanes and Fluorine-containing polymer families had 223 and 227 polymers respectively. The list also contained a handful of phosphorus, and sulphur-based polymers, 34 and 9 respectively. Aromatic-based polymers are also present in the list. Surprisingly accounting for 945 polymers in the list.

Table 4.2 contains a small sample of 8 polymers identified using the neural network and images of their corresponding monomer structures in Fig 4. 2. Table 4. 3 shows 8 repeat units obtained by Venkatram¹¹⁹ et al. for comparison.

1	*N=P(*) (OCCCC) OCCOCCOCC
2	*N=P(*) (OCCC) OCCCCCCCCOCCOCCOC
3	*N=P(*) (OCCCC) OCCOCCOCC
4	*N=P(*) (OCCOCCOCCOC) OCCOCCOCC1COCCO1
5	*OCC(F)(F)C(F)(F)C(F)(F)C(F)(F)CC(F)(F)C(F)(F)CCCCCCC[Si](C)(C)O*
6	*CCC(*)O[Si](C)(C)CCCC
7	*CC(CO*)(COCCOCCOCCOC)OC(=O)CCCCCCCCCCCC(=O)OCCC
8	*C(C(=*)CCCCCCCCC(=O)OCCOCCOC)C(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F

Table 4. 3. SMILES strings for 8 repeat units by Venkatram¹¹⁹ and Allcock¹²⁰

1	<chem>*N=P(*) (OCC)OCC</chem>
2	<chem>*N=P(*) (OCCC)OCCC</chem>
3	<chem>*N=P(*) (OCCCC)OCCCC</chem>
4	<chem>*N=P(*) (OCCCC)OCCCCC</chem>
5	<chem>*N=P(*) (OCC(F)(F)C(F)F)OCC(F)(F)C(F)F</chem>
6	<chem>*C[Si](C)(C)[Si](*) (C)C</chem>
7	<chem>*CCCCCOC1(CO*)CC1</chem>
8	<chem>*CCOCCOCCOC(=O)CCCCSCCCCC(=O)O*</chem>

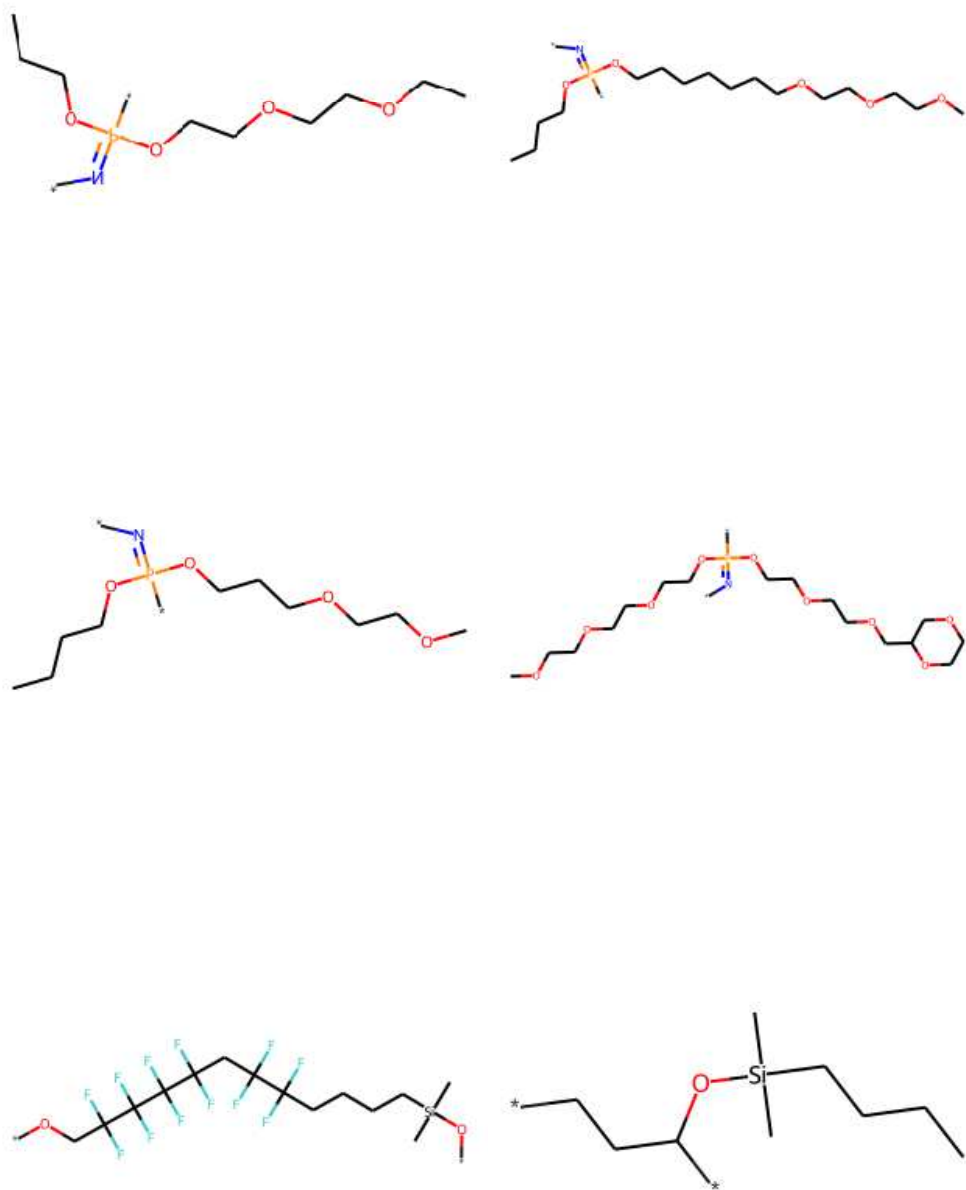


Figure 4.2 A sample of 6 repeat unit structures. The first four represent polyphosphazene repeat units. The bottom left is a fluorine-containing monomer and the bottom right is a polysiloxane. Comparisons with literature show that these units resemble, but whose substituents and branches differ slightly from, some repeat units that have been previously reported^{119,120}

Chapter 5

Overview

The primary goal of this study was to develop statistical learning models capable of identifying novel polymers for solid-state electrolyte applications. Based on the evidence in Chapter 4, which is discussed here, the experiments in Chapter 3 were successful and the goals of building reliable statistical models and identifying new polymers were achieved.

This chapter discusses the results obtained from applying the best model, a small feed-forward neural network, to the PIIM polymer database with the task of identifying solid polymer electrolytes based on the repeat unit. The high-throughput process of classifying each of the 995, 800 polymers in the database revealed 1611 potential homopolymer repeat units as described in Chapter 4. This list of polymers is comprised of many variants of a quite diverse and large number of previously known families of polymers, namely polyphosphazenes, fluorine-containing polymer family, polyethers, and polysiloxanes.⁷² Additionally, the list also contains several less common monomer units. The repeat units suggested by the model are valid and coherent monomer structures that are consistent with the literature review in Chapter 2^{48,58,63,64,92,115}. A sample of 8 structures from this list shows a striking resemblance with common solid polymer electrolytes that have been reported in the literature.^{120,121}

Although all five models performed extremely well, while the neural network outperformed every other model. The model generalizes well to unseen samples despite the risk of overfitting due to the small size of the data set and the large number of descriptors in the input feature vector. Given these unconducive learning conditions, the learning algorithm performed consistently well on different samples. Despite this difficult learning environment, the resulting model was ultimately able to obtain unexpectedly high evaluation scores in accuracy and recall with 75%, an f1 score of 74%, and a precision of 80%. The model's ability to generalize beyond the training set is reassuring. And based on this evidence I have reason to believe that the neural network was able to capture important underlying patterns about solid polymer electrolytes and then able to use that knowledge to make predictions that are reliable. With this confidence in mind, the results obtained from experiments conducted in chapter 3 and described in chapter 4 are discussed here.

Analysis of the observations described in Chapter 4 suggests that the neural network recognizes and captures important underlying chemical patterns associated with ionic conductivity and earmarked in the literature as important to solid ion conductors, not least key substructures and functional groups embedded within the repeat units^{120,122}. Monomers within the list of potential candidates possess repeating units with the electronegative atoms nitrogen (N), oxygen (O), and fluorine (F) associated with forming lithium-ion complexing sites which are necessary for salt dissociation and ion solvation⁷⁰. The model recognizes the importance of oxygen moieties as an indispensable agent in lithium-ion solvation and

transport as indicated by the fact that 98% of the positively labelled monomers in the list of potential polymers have an oxygen atom in one or more functional groups.

Research shows that oxygen moieties are critical to ion transport and segmental motion with solid polymer electrolytes. Therefore it is unsurprising to see so many oxygen and oxygen-containing functional groups present within multiple polymer families, including but not limited to polyphosphazenes, polyethers, polysiloxanes and even organolead-containing polymers. The prevalence of the oxygen atom also raises the question of whether the neural network might be overfitting to oxygen moieties or perhaps generally overfitting to electronegative atoms for that matter.

Broadly, most of the polymer repeat units assigned positive labels by the neural network model have common main back-bone chain polymers from the polyphosphazene, polyether and polysiloxane polymer families. Frequently, the main branch occurs with diverse side chains grafted onto the backbone. The functional groups, substructures and polymer families revealed by the model closely follow the general trends associated with ionic conductivity in solid polymers and the consensus on solid polymer electrolytes. And are now discussed here in more detail.

Oxyethylene

The oxyethylene side group dominates the list of functional groups present in the monomer units. This observation supports the theory that oxyethylene groups are an indispensable part of solid polymer electrolytes¹²⁰. Appearing as long side chains grafted to inorganic polymer backbones. Long ethyleneoxy side chains are associated with increased ionic conductivity and, and fire retardant properties. Chains that are too long, however, have been reported to be detrimental to performance. It has been observed that polymers with more than eight ethyleneoxy units in the side chain tend to crystallize and their ionic conductivity diminishes from that point onwards.

Hopper et al.¹²³ demonstrated that ionic conductivity increases with the increasing length of oxyethylene side chains, up to a maximum of 6 oxygens, but tailing off at higher ethylene oxide content. Results from the neural network classifier are consistent with this observation. Analysis of repeat units with oxyethylene side groups shows a slightly lower average oxygen content of 4 in the ethylene oxide side groups within the polyphosphazene group and 3 in polysiloxanes.

Polyphosphazene

The neural network assigned a large number of positive labels to polymers from the phosphazene family as described in Chapter 4. The analysis of this group of polyphosphazenes reveals repeating units similar to oxyethylene-grafted polyphosphazenes (MEEPs). A common group of polymers is prominent within the SPE literature^{121,124}. These comb-like polymer chains are known to lead to low-glass transition temperature polymers which lead to amorphous regions that play an important role in lithium-ion conduction. Low glass transition temperatures enable local segmental motion dynamics of chains which aid ion hopping and ultimately faster ionic conductivity in polymers.

The lithium conducting abilities of polyphosphazenes are well-known and have been studied previously. Particularly by Allcock and Shriver, who pioneered the work and extensively studied them.^{120,125} First investigated in the 1980s and early days of solid polymer electrolyte research, polyphosphazenes are usually liquid-like at room temperature. In this system, an inorganic backbone made of alternating nitrogen and phosphorus atoms with side chains attached to the phosphorus atom stretches to yield a grafted polymer structure that promotes lithium ion movement.

This system typically outperforms the model PEO system because the phosphazene functional group possesses more segmental mobility due to lower torsional barriers between the nitrogen and phosphorus units along the backbone.¹²¹ At room temperature, the practical ionic conductivity of dry solid MEEP can reach as high as three orders of magnitude greater than PEO⁶⁹.

MEEP, poly [bis (methoxy-ethoxy-ethoxy) phosphazene], is the paragon of the polyphosphazene system. Many polyphosphazenes based on the MEEP system have been reported in the past mainly focusing on liquid and gel electrolytes. Lithium iron phosphate metal cells incorporating MEEP with high ionic conductivity at 30 °C, good capacity retention, and an electrochemical window of 4.4 V versus lithium metal have been demonstrated.

However, the lithium transference number for MEEP is usually very low. MEEP dissociates lithium salts by utilizing the highly electronegative nitrogen atoms to coordinate lithium ions in addition to oxygen atoms. This high degree of dissociation ability means that more charge carriers are involved in conductivity. This class of polyphosphazenes has also been observed to effectively suppress lithium dendrite growth in lithium metal cells.^{120,124}

The electrochemical stability, low glass transition temperature and resultant segmental motion explain why polyphosphazenes exhibit sufficient ionic conductivity at 30 °C and warrant further investigation because polyphosphazenes are deemed safer than organic solvent electrolytes because they have been observed to suppress lithium dendrite growth in lithium metal cells.

Polysiloxane

Polysiloxanes are, like polyphosphazenes, highly flexible inorganic backbones, with a wide variety of substituents and low glass transition temperatures. Overall they have shown lower room temperature performance and mechanical properties than polyphosphazenes. But given this combination of desirable properties and appropriate substituent groups, polysiloxanes should therefore make for suitable host materials for lithium ions. The model assigned positive labels to approximately 223 different, disubstituted comb-like polysiloxane polymers of diverse length scales and substituent variants.

Despite the development of polysiloxane electrolytes, their conductive mechanisms are still unclear and their properties are still lacking. This underwhelming result might explain why polysiloxane-based SPE progress has stalled in recent years; Hopper et al¹²⁶ provide clear

guidance on how research could go about developing high-performance polysiloxane-based polymer electrolytes.

To increase ionic conductivity and optimize their solvation abilities, Smid et al^{58,126} suggested incorporating oligo-ethylene oxide substituents into the polysiloxane, by grafting the substituents into the polysiloxane backbone. This structure has shown great promise as an SPE because it improves ionic conductivity and lowers the glass transition temperature while promoting more segmental motion of the chains. This configuration has led to some of the highest ambient temperature ionic conductivity yet recorded for any polymer electrolyte.

This trend is replicated in many of the 223 polysiloxane repeat units. Some of the structures assigned positive labels by the classifiers are quite similar to those developed by Smid and co-workers. The new polysiloxane variants in the list of potential candidates coupled with the knowledge from the work of Smid and Hopper could lead to breakthroughs in polysiloxane SPEs.

Aromatic

The list of positive labels also has 945 aromatic compounds with benzene-like rings as part of their monomer structure but not necessarily obeying Huckel's rule. This is quite a large number considering that the literature only has a handful of aromatic compounds that have been investigated in the past.^{122,127} Most heterocyclic compounds seem to serve as additives that enhance electrochemical stability within liquid electrolytes instead of host matrixes for lithium ions^{128,129}.

Aromatic compounds are perhaps overlooked partly due to their failure to passivate lithium. But given these newly found leads and the fact that aromatic compounds have shown promising ionic conductivity, high lithium transference numbers and excellent mechanical strength, in the past, these new monomer units warrant more investigation.^{130,131}

Apart from polyethers, polyphosphazenes, benzene-like compounds and polysiloxanes, the remaining set of monomers is made up of polymers that are a mix of a few alkynes, fluorine and sulphur-containing monomers, and other polymer families such as carbonyl coordinating polymers. This set of monomers generally consists of less known and investigated polymers that could lead to breakthroughs or might simply represent flaws in the classification models.

Chapter 6

This study aimed to build statistical learning models capable of identifying a suitable host matrix for a lithium battery solid-state electrolyte. The results, obtained from experiments conducted in chapter 3, described in chapter 4, and discussed in chapter 5, show that all the primary and secondary goals of this study were accomplished. And the results exceeded expectations. High-performing and reliable predictive models were successfully developed, and the models, pending further experimental validation, are capable of identifying polymers that are expected to perform better than the archetypical polyethylene oxide system that has been the standard for solid polymer electrolytes since the 1970s.

The novelty of this work is that a capable classification model was developed and used to screen a very large and untested benchmark polymer database of nearly a million hypothetical and real polymers and hundreds of novel polymers were identified. A thorough cross-reference, in chapter 5, against the solid polymer electrolyte literature shows that the list of candidate polymers suggested by the best-performing model consists of polymer repeat units that are indeed coherent, and consistent with solid polymer electrolyte research and chapter two's literature review. All five statistical learning models from this study are reliable as indicated by high evaluation scores on multiple performance metrics – recall, precision, accuracy, f1 score, and area under the receiver operating curve. Based on this evidence one would have reason to believe that the polymers suggested by the models are certainly promising and worthy of deeper investigation through molecular dynamics simulations, density functional theory or experimental work.

Since most, at least 98%, of the polymers within the PIIM polymer database are hypothetical and remain untested because they have never been made before, it is important to consider their synthetic accessibility and the significant undertaking required for them to materialize. For instance, the polyphosphazene variants are really good candidates but whose synthetic route is relatively complex.

Although the empirical approach to research - characterized by carrying out an umpteen of attempts until success, or at least until the practitioner eventually stops trying - has worked well enough for small problems. This approach has proven to be an expensive, lengthy and painstaking process that is fraught with uncertainties. This approach is undeniably unsustainable, and inadequate given the scale of modern materials science and engineering problems, and the speed with which technological innovation is desperately needed. Solid polymer electrolytes have been a victim of this approach because finding an ideal polymer that simultaneously satisfies all battery electrolyte requirements has failed for more than 40 years of research. This study addressed these problems by providing a better approach for navigating the vast chemical space of polymers and by identifying promising polymers that hold tremendous potential while using very little time and resources. In addition, this study also highlighted the challenges and strategies of working with small but diverse datasets with high-dimensional representations.

The models guide researchers by quantifying uncertainty and making it possible for them to make informed decisions going into experimental work that would ultimately involve

synthesizing, characterizing and testing positively labelled polymers that have been earmarked.

This work began by discussing the motivation, history and challenges of developing lithium metal batteries. It went on to describe and execute a rational and efficient approach for addressing these challenges and enabling improvements through solid polymer electrolytes. By leveraging an existing solid polymer electrolyte database five different predictive models were developed and evaluated to provide a means to identify polymers that are most likely to be good host materials for an all-solid-state lithium metal battery electrolyte. Using the best model, a neural network, the number of potential candidate polymers was reduced from approximately one million to less than two thousand novel repeat units that are largely untested. This list of potential polymers consisted of popular and unpopular polymer families that have been reported in the literature, most notably polyphosphazenes, fluorine-containing polymers, polyethers and polysiloxanes. And the models demonstrated an ability to capture relevant solid polymer electrolyte patterns with the prevalence of prominent electronegative atoms and functional groups that were not present within the training set but are associated with high ionic conductivity in polymers.

All in all, this work has contributed to knowledge and practice by demonstrating generalizing statistical learning models and expanding existing solid polymer electrolyte possibilities that hitherto have remained unknown.

Chapter 7

Despite the success of this study, a few improvements can be made to the data, the statistical learning models, and the experimental procedure used, in order to achieve more robust models and better results.

Firstly, the application of machine learning on materials science data sets is limited by the availability of high-quality data sets of sufficient size. Machine learning models benefit greatly from large quantities of data. The data set used in this study is quite small. The small size may even have prevented the statistical learning algorithms from reaching their full potential. The generation, collection and addition of more experimental data would certainly improve the performance of the models.

The best-performing model, the neural network, labelled 1611 polymers are potentially better than the prototypal polyethylene oxide system. Although the list of possible polymers for electrolyte applications is certainly encouraging, it can also be quite overwhelming. The number of candidates needs to be reduced even further by applying filtering criteria that incorporate the electrochemical and thermal stability of the polymer as well as the mechanical strength. Ideally, this large number of candidates will have to be reduced to a list of between 3 and 5 promising candidates that would be synthesised, characterised and tested. This list can be made even more useful by sorting the polymers from most promising to least promising using their predicted probabilities. This prioritization would allow resources to be dedicated to the material that is most likely to succeed and potentially used more economically.

The quality of input data, the molecular fingerprints can be improved by removing irrelevant features identified through recursive feature elimination as described in Chapter 4. Removing unimportant would be expected to improve the overall performance of the model regarding speed and predictive power. The quality of results can also be improved by incorporating the electrochemical stability window, mechanical strength, and thermal stability into the classification criteria in addition to the ionic conductivity at two different temperatures which was the sole classification criteria in this study.

The choice of learning algorithms used in this study is very basic. This decision was intentional since this study is meant to represent an initial exploratory inquiry to demonstrate the utility and benefits of using even the simplest methods available. Moving forward, more powerful learning algorithms could be employed. In particular, convolutional neural networks and graph neural networks have proven to be highly effective for screening compounds. The use of these state-of-the-art architectures could potentially lead to enormous improvements, new insights and breakthroughs. Recent research shows that they can outperform the best-performing feed-forward neural network used in this study on tasks involving predicting unknown physical and chemical properties of molecules because they incorporate a broader range of structural and chemical information. Graph neural networks are particularly appealing because they do not require molecular fingerprinting but instead work directly on the structural representation of molecules. This enables them to have unique access to

important information that is relevant to extracting latent patterns within the data and making predictions.

Finally, the models developed in this study have potential applications, beyond solid-state polymer electrolytes for lithium batteries, in proton exchange membrane (PEM) hydrogen electrolyzers. PEM electrolyzers require a solid-state membrane to separate the anode and cathode just like lithium-based batteries. Only, instead of transporting lithium ions, the solid-state membrane in this case is required to allow the movement of hydrogen ions between the two electrodes. Given the parallels between the nature of these applications and their requirements, it is easy to imagine using these models for proton exchange membranes.

References

1. Gates, B. How to avoid a climate disaster: the solutions we have and the breakthroughs we need. 370.
2. David, C. & Mackay, J. C. Sustainable Energy — without the hot air This Cover-sheet must not appear in the printed book . (2009).
3. Watkiss, P., Downing, T., Handley, C. & Butterfield, R. The Impacts and Costs of Climate Change. *Modelling support for Future Actions – Benefits and Cost of Climate Change Policies and Measures* 77 (2005).
4. IPCC. IPCC 2021 Technical report. *Ipcc* 150 (2021).
5. How Exactly Does Carbon Dioxide Cause Global Warming? - You Asked.
<https://news.climate.columbia.edu/2021/02/25/carbon-dioxide-cause-global-warming/>.
6. Arrhenius, S. & Holden, E. S. ON THE INFLUENCE OF CARBONIC ACID IN THE AIR UPON THE TEMPERATURE OF THE EARTH Author (s): S . Arrhenius and Edward S . Holden Source: Publications of the Astronomical Society of the Pacific , Vol . 9 , No . 54 (February 1 , Published by: Astronomica. *Publications of the Astronomical Society of the Pacific* 9, 14–24 (1897).
7. Schneider, S. The worst-case scenario. *Nature* 2009 458:7242 **458**, 1104–1105 (2009).
8. Global climate projections. Chapter 10 (Miscellaneous) | ETDEWEB.
<https://www.osti.gov/etdeweb/biblio/20962171>.
9. Where greenhouse gases come from - U.S. Energy Information Administration (EIA).
<https://www.eia.gov/energyexplained/energy-and-the-environment/where-greenhouse-gases-come-from.php>.
10. Group to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Solomon, W. I. *et al.* How do Human Activities Contribute to Climate Change and How do They Compare with Natural Influences?
11. Lun, Y. H. V. & Tung, S. L. D. *Sustainable energy. Green Energy and Technology* (2020). doi:10.1007/978-3-030-31387-6_2.
12. Renewables Were The World's Cheapest Energy Source in 2020 | World Economic Forum.
<https://www.weforum.org/agenda/2021/07/renewables-cheapest-energy-source/>.
13. Zhang, X., Yang, Y. & Zhou, Z. Towards practical lithium-metal anodes. *Chem Soc Rev* **49**, 3040–3071 (2020).
14. Hesse, H. C., Schimpe, M., Kucevic, D. & Jossen, A. *Lithium-ion battery storage for the grid - A review of stationary battery storage system design tailored for applications in modern power grids. Energies* vol. 10 (2017).

15. Behabtu, H. A. *et al.* A review of energy storage technologies' application potentials in renewable energy sources grid integration. *Sustainability (Switzerland)* **12**, 1–20 (2020).
16. Goodenough, J. B. & Park, K. The Li-Ion Rechargeable Battery: A Perspective. (2013).
17. Korthauer, R. *Lithium-ion batteries: Basics and applications. Lithium-Ion Batteries: Basics and Applications* (2018). doi:10.1007/978-3-662-53071-9.
18. Wu, Y. *LITHIUM-ION Fundamentals and Applications*. (2018).
19. Hatakeyama-sato, K. Synthesis of Lithium-ion Conducting Polymers Designed by Machine Learning-based Prediction and Screening. **16**, 2018–2020 (2019).
20. Deng, D. Li-ion batteries: Basics, progress, and challenges. *Energy Sci Eng* **3**, 385–418 (2015).
21. Olof Ramström. Scientific Background on the Nobel Prize in Chemistry 2019 LITHIUM-ION BATTERIES. *The royal swedish academy of sciences* **50005**, 0–13 (2019).
22. Berman, K. *et al.* The Lithium Ion Battery and the EV Market. (2018).
23. Scrosati, B. History of lithium batteries. *Journal of Solid State Electrochemistry* **15**, 1623–1630 (2011).
24. Reddy, M. V., Mauger, A., Julien, C. M., Paoletta, A. & Zaghib, K. Brief History of Early Lithium-Battery Development. *Materials 2020, Vol. 13, Page 1884* **13**, 1884 (2020).
25. Grey, C. P. & Hall, D. S. Prospects for lithium-ion batteries and beyond—a 2030 vision. *Nat Commun* **11**, 2–5 (2020).
26. Nitta, N., Wu, F., Lee, J. T. & Yushin, G. Li-ion battery materials: Present and future. *Materials Today* **18**, 252–264 (2015).
27. Federal Aviation Authority. EVENTS WITH SMOKE, FIRE, EXTREME HEAT OR EXPLOSION INVOLVING LITHIUM BATTERIES Date Device Carrier Aircraft Type. (2021).
28. Zalosh, R., Gandhi, P. & Barowy, A. Lithium-ion energy storage battery explosion incidents. *J Loss Prev Process Ind* **72**, 104560 (2021).
29. Lowell, D. & Huntington, A. Electric Vehicle Market Status. *M.J. Bradley & Associates* (2019).
30. Turcheniuk, K., Bondarev, D., Amatucci, G. G. & Yushin, G. Battery materials for low-cost electric transportation. *Materials Today* **42**, 57–72 (2021).
31. Azevedo, M. *et al.* Lithium and cobalt-a tale of two commodities. *McKinsey&Company Metals and Mining* p1-25 (2018).
32. Mostert, C., Ostrander, B., Bringezu, S. & Kneiske, T. M. Comparing electrical energy storage technologies regarding their material and carbon footprint. *Energies (Basel)* **11**, (2018).
33. Zablocki, A. Energy storage: Fact sheet (2019). *Eesi* **2040**, 1–8 (2019).

34. Cheng, X. B., Zhang, R., Zhao, C. Z. & Zhang, Q. Toward Safe Lithium Metal Anode in Rechargeable Batteries: A Review. *Chem Rev* **117**, 10403–10473 (2017).
35. 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* **164**, A1731–A1744 (2017).
36. Zhang, H. *et al.* Chem Soc Rev electrolytes : advances and perspectives †. *Chem Soc Rev* (2017) doi:10.1039/c6cs00491a.
37. Wheatle, B. K., Fuentes, E. F., Lynd, N. A. & Ganesan, V. Design of Polymer Blend Electrolytes through a Machine Learning Approach. *Macromolecules* **53**, 9449–9459 (2020).
38. Li, J., Ma, C., Chi, M., Liang, C. & Dudney, N. J. Solid electrolyte: The key for high-voltage lithium batteries. *Adv Energy Mater* **5**, 1–6 (2015).
39. Zhang, S. S. A review on electrolyte additives for lithium-ion batteries. **162**, 1379–1394 (2006).
40. Wang, R., Cui, W., Chu, F. & Wu, F. Lithium metal anodes: Present and future. *Journal of Energy Chemistry* **48**, 145–159 (2020).
41. Chen, S., Dai, F. & Cai, M. Opportunities and Challenges of High-Energy Lithium Metal Batteries for Electric Vehicle Applications. *ACS Energy Lett* **5**, 3140–3151 (2020).
42. Xia, S., Wu, X., Zhang, Z., Cui, Y. & Liu, W. Practical Challenges and Future Perspectives of All-Solid-State Lithium-Metal Batteries. *Chem* **5**, 753–785 (2019).
43. Electric vehicle life cycle analysis and raw material availability. 1–11 (2017).
44. Dale, A., Fant, C., Strzepek, K., Lickley, M. & Solomon, S. Climate model uncertainty in impact assessments for agriculture: A multi-ensemble case study on maize in sub-Saharan Africa. *Earths Future* **5**, 337–353 (2017).
45. Li, A. *et al.* A review on lithium-ion battery separators towards enhanced safety performances and modelling approaches. *Molecules* **26**, (2021).
46. Brennan, John W. & Barder, T. E. Battery Electric Vehicles vs . Internal Combustion Engine Vehicles : A United States-Based Comprehensive Assessment. *Arthur D. Little* **48** (2016).
47. Barkenbus, J. N. Prospects for electric vehicles. *Sustainability (Switzerland)* **12**, 1–13 (2020).
48. Agrawal, R. C. & Pandey, G. P. Solid polymer electrolytes: Materials designing and all-solid-state battery applications: An overview. *J Phys D Appl Phys* **41**, (2008).
49. Wang, Z. *et al.* Toward safer solid-state lithium metal batteries: a review. *Nanoscale Adv* **2**, 1828–1836 (2020).

50. Zhao, W., Yi, J., He, P. & Zhou, H. Solid-State Electrolytes for Lithium-Ion Batteries: Fundamentals, Challenges and Perspectives. *Electrochemical Energy Reviews* **2**, 574–605 (2019).
51. Maine, E. & Seegopaul, P. Accelerating advanced- materials commercialization. *Nature Publishing Group* **15**, 487–491 (2016).
52. Sendek, A. D., Cheon, G., Pasta, M. & Reed, E. J. Quantifying the Search for Solid Li-Ion Electrolyte Materials by Anion: A Data-Driven Perspective. *Journal of Physical Chemistry C* **124**, 8067–8079 (2020).
53. Cubuk, E. D., Sendek, A. D. & Reed, E. J. Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data. *Journal of Chemical Physics* **150**, (2019).
54. Tabor, D. P. *et al.* Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater* **3**, 5–20 (2018).
55. Zhang, H., Zhou, M. Y., Lin, C. E. & Zhu, B. K. Progress in polymeric separators for lithium ion batteries. *RSC Adv* **5**, 89848–89860 (2015).
56. Solid-State Landscape. 2021 (2021).
57. Nestler, T. *et al.* Separators and electrolytes for rechargeable batteries: Fundamentals and perspectives. *Physical Sciences Reviews* **4**, 1–29 (2019).
58. Yue, L. *et al.* All solid-state polymer electrolytes for high-performance lithium ion batteries. *Energy Storage Mater* **5**, 139–164 (2016).
59. Nor, S., Yusuf, F. & Arof, A. K. Polymer Electrolytes for Lithium Ion Batteries and Challenges: Part II. 201–230 (2020).
60. Manthiram, A. A reflection on lithium-ion battery cathode chemistry. *Nat Commun* **11**, 1–9 (2020).
61. Miller, T. F., Wang, Z. G., Coates, G. W. & Balsara, N. P. Designing polymer electrolytes for safe and high capacity rechargeable lithium batteries. *Acc Chem Res* **50**, 590–593 (2017).
62. Varzi, A., Raccichini, R., Passerini, S. & Scrosati, B. Challenges and prospects of the role of solid electrolytes in the revitalization of lithium metal batteries. *J Mater Chem A Mater* **4**, 17251–17259 (2016).
63. Wang, H. *et al.* Reviewing the current status and development of polymer electrolytes for solid-state lithium batteries. *Energy Storage Mater* **33**, 188–215 (2020).
64. Zhao, S., Wu, Q., Ma, W. & Yang, L. Polyethylene Oxide-Based Composites as Solid-State Polymer Electrolytes for Lithium Metal Batteries: A Mini Review. *Front Chem* **8**, 1–7 (2020).
65. Wright, P. V. Electrical Conductivity in Ionic Complexes of Poly (ethylene oxide)". 319–327 (1975).

66. Venkatram, S. No Title. (2021).
67. Jiang, Y. *et al.* Development of the PEO based solid polymer electrolytes for all-solid state lithium ion batteries. *Polymers (Basel)* **10**, 1–13 (2018).
68. Zhang, Q., Liu, K., Ding, F. & Liu, X. Recent advances in solid polymer electrolytes for lithium batteries. *Nano Res* **10**, 4139–4174 (2017).
69. Dias, F. B., Plomp, L. & Veldhuis, J. B. J. Trends in polymer electrolytes for secondary lithium batteries. (2000).
70. Arya, A. & Sharma, A. L. *Polymer electrolytes for lithium ion batteries*: a critical study. *Ionics* (Ionics, 2017). doi:10.1007/s11581-016-1908-6.
71. Zhou, D., Shanmukaraj, D., Tkacheva, A., Armand, M. & Wang, G. Polymer Electrolytes for Lithium-Based Batteries: Advances and Prospects. *CHEMPR* **5**, 2326–2352 (2019).
72. Agrawal, R. C. & Pandey, G. P. Solid polymer electrolytes: Materials designing and all-solid-state battery applications: An overview. (2008) doi:10.1088/0022-3727/41/22/223001.
73. Mistry, A., Franco, A. A., Cooper, S. J., Roberts, S. A. & Viswanathan, V. How Machine Learning Will Revolutionize Electrochemical Sciences. *ACS Energy Lett* **6**, 1422–1431 (2021).
74. Chen, L. *et al.* Polymer informatics: Current status and critical next steps. *Materials Science and Engineering R: Reports* **144**, (2021).
75. Mannodi-Kanakkithodi, A. *et al.* Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond. *Materials Today* **21**, 785–796 (2018).
76. Chen, A., Zhang, X. & Zhou, Z. Machine learning: Accelerating materials development for energy storage and conversion. *InfoMat* **2**, 553–576 (2020).
77. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *Journal of Physical Chemistry C* **122**, 17575–17585 (2018).
78. Kauwe, S. K., Rhone, T. D. & Sparks, T. D. Data-driven studies of Li-ion-battery materials. 1–9 (2019).
79. Tao, L., Chen, G. & Li, Y. Article Machine learning discovery of high-temperature polymers Machine learning discovery of high-temperature polymers. *Patterns* **2**, 100225 (2021).
80. Cai, J., Chu, X., Xu, K., Li, H. & Wei, J. Machine learning-driven new material discovery. *Nanoscale Adv* **2**, 3115–3130 (2020).
81. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science* **6**, (2019).
82. Callaway, E. ‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures. *Nature* **588**, 203–204 (2020).

83. Chen, G. *et al.* Machine-learning-assisted de novo design of organic molecules and polymers: Opportunities and challenges. *Polymers (Basel)* **12**, (2020).
84. Kitano, H. Artificial Intelligence to Win the Nobel Prize and Beyond: Scientific Discovery. 39–49 (2016).
85. Liu, Y., Zhou, Q. & Cui, G. Machine Learning Boosting the Development of Advanced Lithium Batteries. **2100442**, 1–17 (2021).
86. Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: Realization of the ‘fourth paradigm’ of science in materials science. *APL Mater* **4**, (2016).
87. Chen, C. *et al.* A Critical Review of Machine Learning of Energy Materials. *Adv Energy Mater* **10**, 1–36 (2020).
88. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
89. Jumper, J. *et al.* Highly accurate protein structure prediction. (2021) doi:10.1038/s41586-021-03819-2.
90. 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* **4**, 996–1006 (2018).
91. Sendek, A. D. *et al.* Machine learning-assisted discovery of many new solid li-ion conducting materials. *ArXiv* (2018) doi:10.1021/acs.chemmater.8b03272.s001.
92. Liu, M., Clement, C., Liu, K., Wang, X. & Sparks, T. D. A data science approach for advanced solid polymer electrolyte design. *Comput Mater Sci* **187**, 110108 (2021).
93. Zhang, Y. *et al.* Unsupervised discovery of solid-state lithium ion conductors. *Nat Commun* **10**, 1–7 (2019).
94. Xie, T. Deep Learning Methods for the Design and Understanding of Solid Materials. (2020).
95. Hatakeyama-sato, K., Tezuka, T. & Nishikitani, Y. Synthesis of Lithium-ion Conducting Polymers Designed by Machine Learning-based Prediction and Screening Synthesis of Lithium-ion Conducting Polymers Driven by Machine Learning-based Prediction and Screening.
96. Schauser, N. S., Kliegle, G. A., Cooke, P., Segalman, R. A. & Seshadri, R. Database Creation, Visualization, and Statistical Learning for Polymer Li + - Electrolyte Design. (2021) doi:10.1021/acs.chemmater.0c04767.
97. Elton, D. C., Boukouvalas, Z., Butrico, M. S., Fuge, M. D. & Chung, P. W. Applying machine learning techniques to predict the properties of energetic materials. 1–12 (2018) doi:10.1038/s41598-018-27344-x.

98. Tao, L., Chen, G. & Li, Y. Machine Learning Discovery of High-Temperature Polymers The Bigger Picture Graphical Abstract Data Science Maturity Keywords. 1–20.
99. Joshi, R. P. *et al.* Machine Learning the Voltage of Electrode Materials in Metal-Ion Batteries. *ACS Appl Mater Interfaces* **11**, 18494–18503 (2019).
100. Gómez-Bombarelli, R. *et al.* Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent Sci* **4**, 268–276 (2018).
101. Schauser, N. S., Kliegle, G. A., Cooke, P., Segalman, R. A. & Seshadri, R. Database Creation, Visualization, and Statistical Learning for Polymer Li + - Electrolyte Design. (2021) doi:10.1021/acs.chemmater.0c04767.
102. Schauser, N. S., Kliegle, G., Cooke, P. & Segalman, R. A. Supporting Information for: Database Creation , Visualization , and Statistical Learning for Polymer Li + Electrolyte Design.
103. Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J Chem Inf Comput Sci* **28**, 31–36 (1988).
104. Wigh, D. S. A review of molecular representation in the age of machine learning. 1–19 (2022) doi:10.1002/wcms.1603.
105. England, N. W. Automated Analysis and Validation of Open Polymer Data. (2011).
106. Lin, T. *et al.* BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules. (2019) doi:10.1021/acscentsci.9b00476.
107. Capecchi, A., Probst, D. & Reymond, J. L. One molecular fingerprint to rule them all: drugs , biomolecules , and the metabolome. *J Cheminform* 1–15 (2020) doi:10.1186/s13321-020-00445-4.
108. Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. 742–754 (2010).
109. Domingos, P. A few useful things to know about machine learning. *Commun ACM* **55**, 78–87 (2012).
110. Wang, A. Y. T. *et al.* Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chemistry of Materials* **32**, 4954–4965 (2020).
111. Ma, R. & Luo, T. Machine Learning and Deep Learning PI1M: A Benchmark Database for Polymer Informatics PI1M: A Benchmark Database for Polymer Informatics. 0–21 (2020) doi:10.1021/acs.jcim.0c00726.
112. Pattanaik, L. & Coley, C. W. II Molecular Representation: Going Long on Fingerprints. *CHEMPR* **6**, 1204–1207 (2020).
113. Roy, N. K., Potter, W. D. & Landau, D. P. Polymer property prediction and optimization using neural networks. *IEEE Trans Neural Netw* **17**, 1001–1014 (2006).

114. Sumpter, B. G. & Noid, D. W. Neural networks as tools for predicting materials properties. *Annual Technical Conference - ANTEC, Conference Proceedings* **2**, 2556–2560 (1995).
115. Choudhury, S. *et al.* Solid-state polymer electrolytes for high-performance lithium metal batteries. *Nat Commun* **10**, 1–8 (2019).
116. Chen, J. *et al.* All Solid Polymer Electrolytes for Lithium Batteries. *Progress in Chemistry* **32**, 481–496 (2020).
117. Zhang, X., Daigle, J. & Zaghib, K. Comprehensive Review of Polymer Architecture for All-Solid-State Lithium Rechargeable Batteries. (2020).
118. Sun, B., Mindemark, J., Edström, K. & Brandell, D. Polycarbonate-based solid polymer electrolytes for Li-ion batteries. *Solid State Ion* **262**, 738–742 (2014).
119. Venkatram, S., Joseph, R. & Brettmann, B. *MACHINE LEARNING BASED MODELS FOR THE DESIGN OF SOLID POLYMER ELECTROLYTES A Dissertation Presented to The Academic Faculty MACHINE LEARNING BASED MODELS FOR THE DESIGN OF SOLID POLYMER ELECTROLYTES.* (2021).
120. Recent Progress with Ethyleneoxy Phosphazenes as Lithium Battery Electrolytes Shih-To Fei and Harry R. Allcock* Department of Chemistry, The Pennsylvania State University University Park, PA 16802, U.S.A. **1127**, (2009).
121. Ingber, T. T. K. *et al.* Increasing the Lithium Ion Mobility in Poly (Phosphazene) -Based Solid Polymer Electrolytes through Tailored Cation Doping Increasing the Lithium Ion Mobility in Poly (Phosphazene) -Based Solid Polymer Electrolytes through Tailored Cation Doping. (2021) doi:10.1149/1945-7111/ac148d.
122. Bergfelt, A., Lacey, M. J., Bowden, T., Hedman, J. & Christofer, S. RSC Advances. 16716–16725 (2018) doi:10.1039/c8ra00377g.
123. Hooper, R., Lyons, L. J., Moline, D. A. & West, R. Novel siloxane polymers as polymer electrolytes for high energy density lithium batteries. 121–128 (2002).
124. Mindemark, J., Lacey, M. J., Bowden, T. & Brandell, D. Beyond PEO—Alternative host materials for Li+-conducting solid polymer electrolytes. *Progress in Polymer Science* vol. 81 114–143 Preprint at <https://doi.org/10.1016/j.progpolymsci.2017.12.004> (2018).
125. Jankowsky, S., Hiller, M. M., Stolina, R. & Wiemh, H. Performance of polyphosphazene based gel polymer electrolytes in combination with lithium metal anodes. **273**, 574–579 (2015).
126. Hooper, R. *et al.* Highly Conductive Siloxane Polymers. 931–936 (2001).
127. Manuscript, A. metal batteries. (2016) doi:10.1039/x0xx00000x.
128. Adachi, M., Tanaka, K. & Sekai, K. Aromatic Compounds as Redox Shuttle Additives for 4 V Class Secondary Lithium Batteries. *J Electrochem Soc* **146**, 1256–1261 (1999).
129. Zhang, L., Zhang, Z. & Amine, K. Redox Shuttle Additives for Lithium-Ion Battery.

130. Shima, K. *et al.* Reaction mechanisms of aromatic compounds as an overcharge protection agent for 4 V class lithium-ion cells. **161**, 1264–1274 (2006).
131. Yoo, D. *et al.* Tuning the Electron Density of Aromatic Solvent for Stable Solid-Electrolyte-Interphase Layer in Carbonate-Based Lithium Metal Batteries. **1802365**, 1–10 (2018).