Accelerating Discovery and Design of High-Performance Solid-State Electrolytes: A Machine Learning Approach

Ram Sewak, Vishnu Sudarsanan, and Hemant Kumar*

School of Basic Sciences, Indian Institute of Technology
Bhubaneswar, Argul, Khordha, PIN - 752050, Odisha, India
*hemant@iitbbs.ac.in

Supporting information

S0.1 Dataset and Preprocessing

The collection process yielded information on 170 compositions from NASICONs, with entries containing mobile alkali ions Na and Li. This amount of dataset meets the guideline that the training data should be approximately five times greater than the number of descriptors [1], in order to prevent the 'curse of dimensionality' or overfitting [2]. Significantly, the dataset comprises a combination of low and high ionic conductivity ranging from 10^{-3} to 10^{-12} S/cm. This diversity of data will enhance the reliability of the machine learning models [1]. Due to the limited size of the training dataset and the potential interference of experimental noise, classification was selected as the preferred approach for studying ionic conductivities rather than predicting specific values. This method effectively addresses issues related to noise and overfitting by using decision boundaries to establish multiple classes, allowing the grouping of several orders of magnitudes. Therefore, it is an effective screening method. The dataset containing formulas and corresponding ionic conductivities for NASICON materials is listed in Table S1.

Table S1: List of NASICON compounds with ionic conductivity at 25 ^{0}C.

Formula	Ionic Conductivity	Reference
$NaZr_2(PO_4)_3$	4.50E-06	[5]
$NaSn_{0.5}Zr_{1.5}(PO_4)_3$	4.23E-11	[5]
$NaSnZr(PO_4)_3$	2.47E-10	[5]
$NaSn_{1.5}Zr_{0.5}(PO_4)_3$	7.91E-10	[5]
$NaNbZr(PO_4)_3$	2.49E-08	[5]
$NaMoZr(PO_4)_3$	2.06E-09	[5]
$NaMoTi(PO_4)_3$	3.07E-07	[5]
$Na_3Zr_2(SiO_4)_2(PO_4)$	6.70E-04	[7]

$Na_4Zr_2(SiO_4)3$	8.87E-09	[5]
$Na_{1.2}In_{0.2}Zr_{1.8}(PO_4)_3$	2.08E-07	[8]
$Na_{1.4}In_{0.4}Zr_{1.6}(PO_4)_3$	9.76E-07	[8]
$Na_{1.5}Al_{0.5}Zr_{1.5}(PO_4)_3$	5.70E-06	[9]
$Na_{1.5}Cr_{0.5}Zr_{1.5}(PO_4)_3$	1.00E-05	[9]
$Na_{1.5}Ga_{0.5}Zr_{1.5}(PO_4)_3$	3.40E-06	[9]
$Na_{1.5}In_{0.5}Zr_{1.5}(PO_4)_3$	2.90E-05	[9]
$Na_{1.5}Sc_{0.5}Zr_{1.5}(PO_4)_3$	5.80E-05	[9]
$Na_{1.5}Y_{0.5}Zr_{1.5}(PO_4)_3$	5.60E-05	[9]
$Na_{1.5}Yb_{0.5}Zr_{1.5}(PO_4)_3$	3.00E-05	[9]
$Na_2AlZr(PO_4)_3$	1.20E-06	[9]
$Na_2CrZr(PO_4)_3$	2.50E-05	[9]
$Na_{2.5}Sc_{0.2}Zr_{1.8}(SiO_4)_{1.3}(PO_4)_{1.7}$	3.19E-04	[10]
$Na_3Sc_{1.5}Zr_{0.5}(SiO_4)_{0.5}(PO_4)_{2.5}$	1.17E-04	[10]
$Na_3ScZr(SiO_4)_2(PO_4)$	1.82E-04	[10]
$Na_3Sc_{0.8}Zr_{1.2}(SiO_4)_{1.2}(PO_4)_{1.8}$	1.42E-04	[10]
$Na_{3.5}Sc_{0.5}Zr_{1.5}(SiO_4)_2(PO_4)$	4.88E-04	[10]
$Na_{2.7}Sc_{0.2}Zr_{1.8}(SiO_4)_{1.5}(PO_4)_{1.5}$	8.87E-05	[10]
$Na_{1.3}Al_{0.3}Zr_{1.7}(PO_4)_3$	6.30E-05	[11]
$Na_3MgZr(PO_4)_3$	1.00E-06	[12]
$Na_3MnZr(PO_4)_3$	1.80E-06	[12]
$Na_{2.0}Mg_{0.5}Zr_{1.5}(PO_4)_3$	9.60E-06	[13]
$Na_{2.4}Mg_{0.7}Zr_{1.3}(PO_4)_3$	1.40E-05	[13]
$Na_{1.5}Nb_{0.3}Zr_{1.5}(PO_4)_3$	6.20E-07	[5]
$Na_{1.8}Yb_{0.8}Zr_{1.2}(PO_4)_3$	1.58E-03	[14]
$LiZr_2(PO_4)_3$	1.00E-05	[9]
$Li_{1.2}Zr_{1.8}Ca_{0.2}(PO_4)_3$	1.30E-05	[15]
$Li_{1.4}Zr_{1.6}Ca_{0.4}(PO_4)_3$	8.50E-06	[15]
$Na_{1.2}Zr_{1.8}Fe_{0.2}(PO_4)_3$	2.54E-06	[16]
$Na_2In_{1.0}Zr_{1.0}(PO_4)_3$	3.16E-04	[14]
$Na_2YbZr_{1.0}(PO_4)_3$	3.55E-03	[14]
$Na_2Mg_{0.5}Zr_{1.5}(PO_4)$	3.98E-04	[17]
$Na_3Mg_{1.0}Zr_{1.0}(PO_4)_3$	1.58E-03	[17]
$LiZr_{1.9}Sr_{0.1}(PO_4)_3$	3.44E-05	[18]
$LiZr_{1.8}Sr_{0.2}(PO_4)_3$	1.42E-05	[18]
$Li_2ZrFe(PO_4)_3$	1.14E-08	[28]
$Li_2ZrIn(PO_4)_3$	8.22E-08	[19]
$Li_{1.1}Zr_{1.9}Y_{0.1}(PO_4)_3$	8.90E-03	[20]
$Li_{1.15}Zr_{1.85}Y_{0.15}(PO_4)_3$	7.00E-03	[20]
$Na_3Zr_2(SiO_4)_2(PO_4)$	6.70E-04	[21]
$Na_{3.1}Sc_{0.1}Zr_{1.9}Si_2PO_{12}$	1.32E-03	[21]
$Na_{3.2}Sc_{0.2}Zr_{1.8}Si_2PO_{12}$	1.90E-03	[21]
$Na_{3.3}Sc_{0.3}Zr_{1.7}Si_2PO_{12}$	2.80E-03	[21]
$Na_{3.4}Sc_{0.4}Zr_{1.6}Si_2PO_{12}$	4.00E-03	[21]
$Na_{3.5}Sc_{0.5}Zr_{1.5}Si_2PO_{12}$	3.80E-03	[21]
$Na_{3.6}Sc_{0.6}Zr_{1.4}Si_2PO_{12}$	1.75E-03	[21]
$Na_{3.05}La_{0.05}Zr_{1.9}5Si_2PO_{12}$	9.60E-04	[22]

$Na_{3.10}La_{0.1}Zr_{1.9}Si_2PO_{12}$	1.30E-03	[22]
$Na_{3.15}La_{0.1}5Zr_{1.8}5Si_2PO_{12}$	1.50E-03	[22]
$Na_{3.2}La_{0.2}Zr_{1.8}Si_2PO_{12}$	1.80E-03	[22]
$Na_{3.25}La_{0.25}Zr_{1.75}Si_2PO_{12}$	3.20E-03	[22]
$Na_{3.3}La_{0.3}Zr_{1.7}Si_2PO_{12}$	3.40E-03	[22]
$Na_{3.35}La_{0.3}5Zr_{1.6}5Si_2PO_{12}$	2.00E-03	[22]
$Na_{3.4}La_{0.4}Zr_{1.6}Si_2PO_{12}$	1.30E-03	[22]
$Na_3Zr_{1.9}Nb_{0.08}Si_2PO_{12}$	2.06E-04	[6]
$Na_3Zr_{1.7}Nb_{0.2}4Si_2PO_{12}$	2.26E-04	[6]
$Na_3Zr_{1.5}Nb_{0.4}0Si_2PO_{12}$	1.20E-04	[6]
$Na_3Zr_{1.3}Nb_{0.5}6Si_2PO_{12}$	3.31E-05	[6]
$Na_3Zr_{1.2}Nb_{0.8}Si_2PO_{12}$	3.31E-05	[6]
$Na_{1.4}Zr_{1.8}Mg_{0.2}P_3O_{12}$	6.90E-06	[13]
$Na_{1.8}Zr_{1.6}Mg_{0.4}P_3O_{12}$	8.90E-06	[13]
$Na_{2.4}Zr_{1.3}Mg_{0.7}P_3O_{12}$	9.60E-06	[13]
$Na_{2.8}Zr_{1.1}Mg_{0.9}P_3O_{12}$	1.40E-05	[13]
$Na_{2.2}In_{1.2}Zr_{0.8}(PO_4)_3$	2.63E-06	[5]
$Na_{2.5}Cr_{1.5}Zr_{0.5}(PO_4)_3$	1.80E-04	[9]
$Na_{2.5}In_{1.5}Zr_{0.5}(PO_4)_3$	1.00E-04	[9]
$Na_{2.5}Sc_{1.5}Zr_{0.5}(PO_4)_3$	5.60E-04	[9]
$Na_{2.5}Y_{1.5}Zr_{0.5}(PO_4)_3$	4.60E-05	[5]
$Na_{2.5}Yb_{1.5}Zr_{0.5}(PO_4)_3$	1.90E-04	[9]
$Na_{2.6}In_{1.6}Zr_{0.4}(PO_4)_3$	2.81E-06	[8]
$Na_{2.8}In_{1.8}Zr_{0.2}(PO_4)_3$	2.38E-06	[8]
$NaGe_2(PO_4)_3$	1.10E-12	[5]
$NaGeTi(PO_4)_3$	8.50E-12	[5]
$Na_{1.4}Al_{0.4}Ge_{1.6}(PO_4)_3$	7.28E-10	[23]
$NaTi_{0.5}Ge_{1.5}(PO_4)_3$	3.13E-11	[9]
$Na_{1.25}Sn_{0.25}Ge_{1.75}(PO_4)_3$	7.00E-06	[24]
$Na_{1.5}Sn_{0.5}Ge_{1.5}(PO_4)_3$	8.39E-05	[24]
$Na_{1.75}Sn_{0.75}Ge_{1.25}(PO_4)_3$	1.20E-05	[24]
$Li_{1.5}Al_{0.5}Ge_{1.5}(PO_4)_3$	4.00E-04	[25]
$Li1Ge_2(PO_4)_3$	8.00E-08	[26]
$Li_{1.5}Al_{0.5}Ge_{1.5}(PO_4)_3$	1.00E-04	[26]
$NaTi_2(PO_4)_3$	4.43E-10	[5]
$NaGe_{0.5}Ti_{1.5}(PO_4)_3$	5.91E-13	[5]
$NaSn_{0.5}Ti_{1.5}(PO_4)_3$	1.77E-11	[5]
$NaSnTi(PO_4)_3$	6.86E-11	[5]
$NaSn_{1.5}Ti_{0.5}(PO_4)_3$	5.14E-10	[5]
$NaNbTi(PO_4)_3$	1.59E-06	[5]
$Na_{1.4}Al_{0.4}Ti_{1.6}(PO_4)_3$	5.60E-08	[5]
$Na_{1.6}Al_{0.6}Ti_{1.4}(PO_4)_3$	1.10E-07	[5]
$Na_{1.8}Al_{0.8}Ti_{1.2}(PO_4)_3$	1.20E-07	[5]
$Na_{1.9}Al_{0.9}Ti_{1.1}(PO_4)_3$	1.30E-07	[5]
$Na_{1.4}In_{0.4}Ti_{1.6}(PO_4)_3$	1.86E-08	[5]
$Na_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$	1.40E-05	[11]
$LiTi_2(PO_4)_3$	3.60E-08	[27]

$Li_2TiFe(PO_4)_3$	3.34E-07	[28]
$Li_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$	7.00E-04	[29]
$Li_{1.3}Sc_{0.3}Ti_{1.7}(PO_4)_3$	7.00E-04	[29]
$Li_{1.2}Al_{0.2}Ti_{1.8}(PO_4)_3$	3.16E-03	[30]
$Li_{1.5}Al_{0.5}Ti_{1.5}(PO_4)_3$	4.70E-03	[30]
$Li_{1.3}Cr_{0.3}Ti_{1.7}(PO_4)_3$	1.99E-05	[31]
$Li_{1.7}Cr_{1.7}Ti_{1.3}(PO_4)_3$	6.31E-05	[31]
$NaHf_2(PO_4)_3$	8.77E-10	[27]
$Na_{2.4}Hf_2(SiO_4)_{1.4}(PO_4)_{1.6}$	7.30E-04	[32]
$Na_{2.6}Hf_2(SiO_4)_{1.6}(PO_4)_{1.4}$	5.90E-04	[32]
$Na_{2.8}Hf_2(SiO_4)_{1.8}(PO_4)_{1.2}$	6.90E-04	[32]
$Na_3Hf_2(SiO_4)_2(PO_4)$	1.10E-03	[32]
$Na_{3\cdot2}Hf_2(SiO_4)_{2.2}(PO_4)_{0.8}$	2.30E-03	[32]
$Na_{3.4}Hf_2(SiO_4)_{2.4}(PO_4)_{0.6}$	1.40E-03	[32]
$Na_{3.6}Hf_2(SiO_4)_{2.6}(PO_4)_{0.4}$	1.20E-03	[32]
$Na_{3.8}Hf_2(SiO_4)_{2.8}(PO_4)_{0.2}$	3.20E-04	[32]
$Na_{1.4}In_{0.4}Hf_{1.6}(PO_4)_3$	1.86E-07	[32]
$LiHf_2(PO_4)_3$	6.31E-07	[33]
$Li1.1Cr_{0.1}Hf_{1.9}(PO_4)_3$	3.80E-04	[35]
$Li_{1.2}Fe_{0.2}HF_{1.8}\ (PO_4)_3$	1.58E-04	[33]
$Li_{1.2}Cr_{0.2}HF_{1.8}\ (PO_4)_3$	1.00E-04	[33]
$Li_{1.2}Sc_{0.2}HF_{1.8}\ (PO_4)_3$	2.00E-05	[33]
$Li_{1.2}In_{0.2}HF_{1.8}\ (PO_4)_3$	1.26E-04	[33]
$Li_{1.2}Lu_{0.2}HF_{1.8}\ (PO_4)_3$	1.00E-04	[33]
$Li_{1.2}Y_{0.2}HF_{1.8}\ (PO_4)_3$	3.16E-05	[33]
$Li_{1.4}Fe_{0.4}HF_{1.6}\ (PO_4)_3$	2.00E-05	[33]
$Li_{1.4}Sc_{0.4}HF_{1.6}\ (PO_4)_3$	1.00E-04	[33]
$Li_{1.4}In_{0.4}HF_{1.6}\ (PO_4)_3$	1.41E-04	[33]
$Li_{1.4}Lu_{0.4}HF_{1.6}\ (PO_4)_3$	2.00E-05	[33]
$Li_{1.4}Y_{0.4}HF_{1.6}\ (PO_4)_3$	2.00E-05	[33]
$Na_{1.6}Al_{0.2}Hf_{1.8}Si_{0.4}P_{2.6}O_{12}$	4.33E-05	[34]
$Na_{2\cdot2}Al_{0.4}Hf_{1.6}Si_{0.8}P_{2\cdot2}O_{12}$	1.06E-04	[34]
$Na_{2.8}Al_{0.6}Hf_{1.4}Si_{1.2}P_{1.8}O_{12}$	5.85E-04	[34]
$Na_{3.4}Al_{0.8}Hf_{1.2}Si_{1.6}P_{1.4}O_{12}$	2.11E-04	[34]
$Na4.0Al_{1.0}Hf_{1.0}Si_{2.0}P_{1.0}O_{12}$	$_{1.9}$2E-04	[34]
$Na4.6Al_{1.2}Hf_{0.8}Si_{2.4}P_{0.6}O_{12}$	$_{1.3}$0E-04	[34]
$NaSn_2(PO_4)_3$	4.65E-09	[5]
$Na_3Cr_2(PO_4)_3$	1.70E-07	[5]
$Na_3Fe_2(PO_4)_3$	1.20E-07	[5]
$Na_3Sc_2(PO_4)_3$	2.27E-05	[5]
$Na_{1.4}Al_{0.4}Sn_{1.6}(PO_4)_3$	1.41E-08	[5]
$Na_{1.4}In_{0.4}Sn_{1.6}(PO_4)_3$	2.72E-08	[5]
$Na_3V_2(PO_4)_3$	3.00E-08	[36]
$Na_3V_{1.9}Fe_{0.1}(PO_4)_3$	2.00E-06	[36]
$Na_3V_{1.9}Al_{0.1}(PO_4)_3$	1.00E-07	[36]
$Na_{3.1}V_{1.9}Ni_{0.1}(PO_4)_3$	1.20E-06	[36]
$Na_3V_{1.9}Cr_{0.1}(PO_4)_3$	4.00E-08	[36]

$Na_{3.4}Sc_2(SiO_4)_{0.4}(PO_4)_{2.6}$	4.00E-03	
$Na_{3.4}Mg_{0.4}Cr_{1.6}(PO_4)_3$	1.58E-02	
$Na_{3.65}Mg_{0.65}Cr_{1.35}(PO_4)_3$	6.31E-03	
$LiCrTa(PO_4)_3$	6.20E-06	
$Li_3Cr_2(PO_4)_3$	2.50E-08	
$Na_{3.1}Zr_{1.95}Mg_{0.05}Si_2PO_{12}$	3.5E-03	[37]
$Na_{3.1}Zr_{1.95}Ca_{0.05}Si_2PO_{12}$	2.1E-03	[37]
$Na_{3.1}Zr_{1.95}Sr_{0.05}Si_2PO_{12}$	1.8E-03	[37]
$Na_{3.1}Zr_{1.95}Ba_{0.05}Si_2PO_{12}$	1.20E-03	[37]
$Na_{3.4}Zr_{1.8}Co_{0.2}Si_2PO_{12}$	1.55E-03	[38]
$Na_{3.4}Zr_{1.8}Ni_{0.2}Si_2PO_{12}$	6.18E-04	[38]
$Na_{3.4}Zr_{1.8}Zn_{0.2}Si_2PO_{12}$	7.8E-03	[38]
$Na_{3.4}Zr_{1.8}Y_{0.2}Si_2PO_{12}$	3.52E-04	[38]
$Na_{3.3}Zr_{1.7}Pr_{0.3}Si_2PO_{12}$	1.27E-03	[39]
$Na_{3.3}Zr_{1.7}Eu_{0.3}Si_2PO_{12}$	1.08E-03	[39]
$Na_{3.3}Zr_{1.7}Lu_{0.3}Si_2PO_{12}$	8.30E-04	[39]
$Na_3Zr_{1.9}Nb_{0.1}Si_2PO_{12}$	2.10E-04	[40]
$Na_3Zr_{1.9}Ti_{0.1}Si_2PO_{12}$	3.77E-04	[40]
$Na_{3.2}Zr_{1.8}Fe_{0.2}Si_2PO_{12}$	7.53E-04	[38]
$Na_{3.2}Zr_{1.8}Al_{0.2}Si_2PO_{12}$	4.39E-04	[38]
$Na_{3.2}Zr_{1.9}Ga_{0.1}Si_2PO_{12}$	1.06E-03	[41]

S0.2 Features selection

A total of 27 features and their determination methods for every compound in the data set tabulated in Table S4. This list includes descriptors for structural, physical, chemical, and elemental properties, encompassing a broad spectrum of potential properties.

However, utilizing multiple features has limitations, as it reduces the processing speed of the algorithm and increases its complexity due to the high dimensionality of the data. Additionally, many of these features exhibit correlation with each other, which needs to be identified in order to enhance model performance and reduce overfitting. We have used Pearson's correlation coefficient to initially check the correlation between features (Figure S1). It is a statistical measure that can be used to determine the strength and direction of the linear relationship between two variables. In feature selection, Pearson's correlation coefficient can be used to identify highly correlated features that may cause collinearity problems. Once the correlation between features has been calculated, highly correlated features have been removed to reduce redundancy and improve model performance. A threshold value > 0.90 can be set to determine which features to remove based on their mutual correlation. After that, principal component analysis (PCA) was used to find the highest independent weightage of features by transforming the original dataset into a new set of variables called principal components. These principal components are linear combinations of the original features that capture most of the variance in the data as shown in Figure S2. By using PCA, the most important features have been identified, that contribute to the variance in the data and reduce the dimensionality of the dataset while preserving most of the information present

Table S2: List of features used in the model evaluation process with methodology for determining features.

Feature	Description	Determination approach
S	Ionic conductivity at room temperature (S/cm2)	Previously reported data
a	Lattice parameter (Å)	Previously reported data
b	Lattice parameter (Å)	Previously reported data
c	Lattice parameter (Å)	Previously reported data
a/b	Ratio of a to b parameters	a/b
a/c	Ratio of a to c parameters	a/c
h	Planer diagonal length (Å)	$\sqrt{a^2+b^2}$
d	Cell diameter (Å)	$\sqrt{c^2+h^2}$
V	Calculated volume according to crystal system	For trigonal = $a^2c\cdot\sin(60^0)$ and for monoclinic = $abc\cdot\sin(\beta)$
R	Ionic radius of Na/Li	Reported
MR	Atomic radius of M atom	
MPR	Atomic radius of M'	
ER	Calculated effective ionic radius	R·N (Stoichiometric occupancy)
MER	Calculated effective atomic radius	MR·N
MPER	Calculated effective atomic radius	MPR·N
EV	Calculated effective volume of Na/Li	$\frac{4}{3}\pi\cdot$R·N
MEV	Calculated effective volume of M atom	$\frac{4}{3}\pi\cdot$MR·N
MPEV	Calculated effective volume of M' atom	$\frac{4}{3}\pi\cdot$MPR·N
EN	Electronegativity of Na/Li	Tabulated
MEN	Electronegativity of M atom	Tabulated
MPEN	Electronegativity of M' atom	Tabulated
EEN	Effective electronegativity Na/Li	EN·N
MEEN	Effective electronegativity of M atom	MEN·N
MPEEN	Effective electronegativity of M' atom	MPEN·N
M	Stoichiometric occupancy of Li/Na at M2 site	Calculated Na/Li formulaic occupancy compared to pristine structure
PO_4	Stoichiometric number of PO_4 group present in formula	Incremented based on formulaic count of PO_4 in compound
SO_4	Stoichiometric number of SO_4 group present in formula	Incremented based on formulaic count of SO_4 in compound

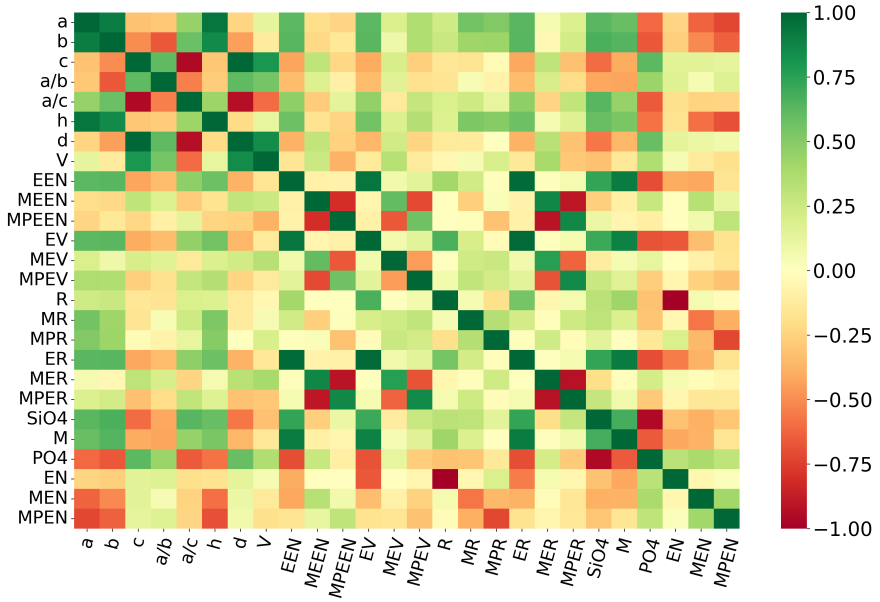

Figure S1: Pearson's correlation matrix corresponding to all features considered in analysis

Table S3: Summary of the splitted dataset for training and testing.

	Training	Testing
Good(1) (%)	21.8	19.7
Bad(0) (%)	78.2	80.3
Good/Bad	0.28	0.24

in the data. Here, we have chosen the characteristics with the highest weighting of each principal component contributing to 95% of the data's variance. This has provided us with 9 features with a maximum value of Pearson correlation coefficient among them less than approximately 65% shown in Figure S3.

S0.3 Model selection

In this work, we compare three machine learning (ML) algorithms such as logistic regression, linear support vector machine, and XG Boost, to see which one provides the most accurate predictions for the good ionic conductor class. The details of the Logistic regression model can be found in the main manuscript.

Support vector machine (SVM) classifier uses a separating hyperplane with a soft margin that allows for some misclassification [1]. The SVM model is based on the principle of finding a hyperplane to use as a boundary that divides the output classes as cleanly as possible. XGBoost algorithm is a decision-tree-based ensemble ML algorithm that uses a parallelized implementation of the sequential tree-building approach [4]. The XGBoost algorithm learns by gradually incorporating an ensemble of decision trees, each of which has the capacity to reduce the ensemble's overall error rate. As the new trees are trained to im-

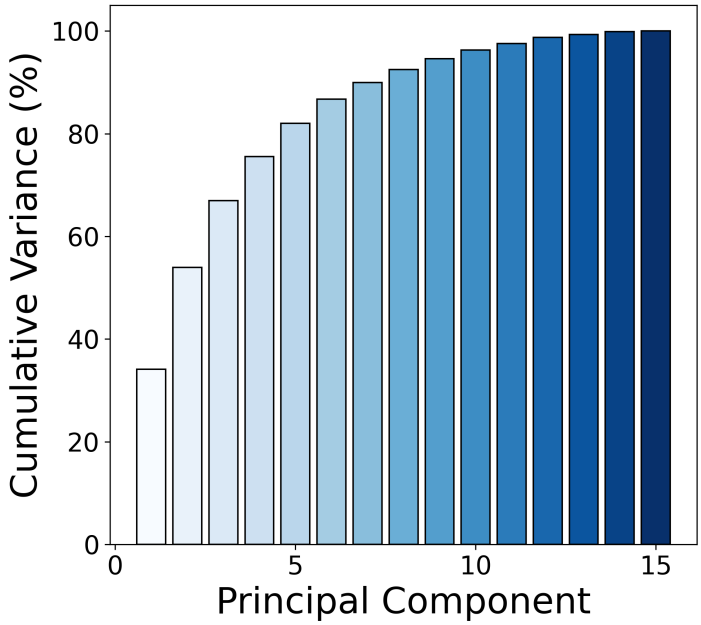

Figure S2: cumulative variance of principal components for 15 features obtained by discarding features having more than 90% correlation.

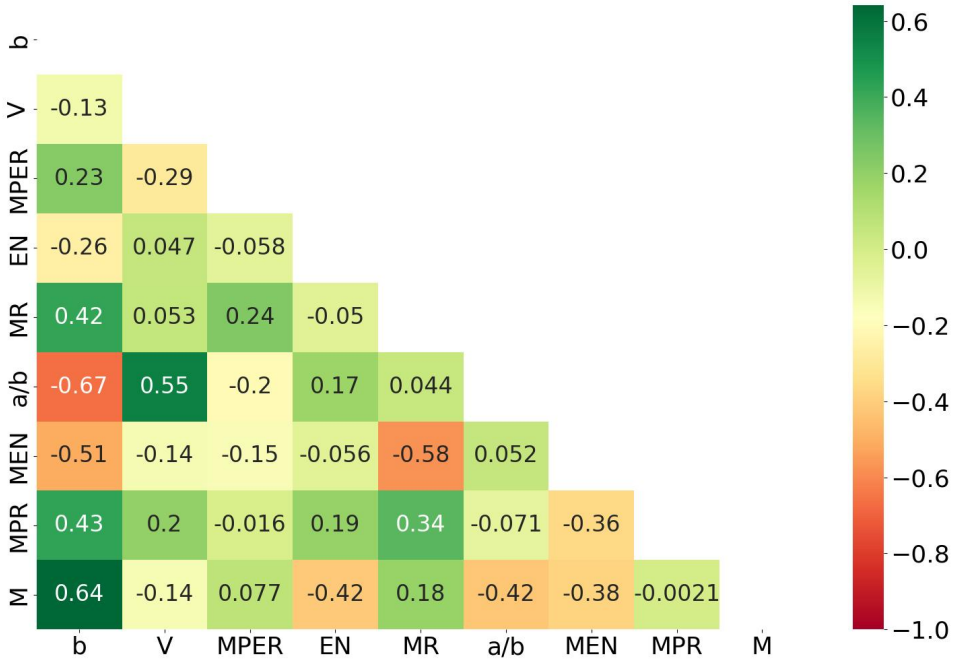

Figure S3: Pearson's correlation matrix for features filtering out by PAC analysis

Table S4: List of features used in the model evaluation process with methodology for determining features.

ML Model	Training accuracy (TAC %)	Cross validation score (CVS %)	Testing accuracy (TSAC %)	Standard deviation
Logistic regression	87.4	87.5	82.3	0.06
Support vector machine	85.7	87.4	82.3	0.03
XGBoost	82.2	76.2	80.4	0.05

prove upon the areas where the old tree failed, the model's overall performance improves. The mixed-type data can be processed effectively by tree-based algorithms. Ensemble models, such as those based on trees, can reduce overfitting by combining many weak learners into a single robust one.

The training accuracy (TAC) and 5-fold cross-validation score (CVS) and testing accuracy (TSAC) percentage for these three different ML models are compared in Table S4. The CSV was employed to measure the model's predictive capacity at the specified feature set size. It provided a concise assessment of the extent of underfitting or overfitting observed with given combinations of features. Models that perform well in training but poorly in cross-validation are likely overfitted because they cannot generalize or extrapolate beyond their training dataset. Whereas models that perform poorly in both training and cross-validation point to underfitting, or a lack of data or appropriate features for prediction. The hyper-parameters of each model are listed below:

Logistic regression

- Solver = saga, used due to both the number of samples and the number of features being large.
- penalty = l2, default one is used to reduce model generalization error, and regulate overfitting with a regularization strength of 6.000569183179
- tolerence = 0.000000000001

Support Vector Machine

- kernel= linear, a linear function used to separate the data, which helps to get positive and negative weightage of features.
- C=2.0, aims to correctly classify as many training examples as possible by minimizing the margin violations, even if the margin is smaller.
- gamma=0.001, The lowest values chosen for gamma, which include farther points, are also considered when selecting the decision boundary.

XGBoost Classifier

- colsample_bytree= 0.5, half of the fraction of features (columns) to be randomly sampled for each tree in the model.

- gamma= 0.001, penalty =l2, minimum loss reduction required to make a further partition on a leaf node of the tree.
- learning_rate =0.001, step size at each iteration while moving towards the minimum of the loss function.
- n_estimators= 400, number of trees to be used in the model

Table S5: Formation energy calculations

Formation energy of $LiGe_2(PO_4)_3$ = -2.00738 eV/atom					
	Formation energy of $Li_{1+x}A_xGe_{2-x}(PO_4)_3$				
A^{+3}	$LA_{0.167}GP$	$LA_{0.33}GP$	$LA_{0.5}GP$	$LA_{0.667}GP$	$LA_{0.833}GP$
Cr	-2.005	-2.003	-1.995	-2.014	-2.015
Fe	-1.993	-1.980	-1.972	-1.966	-1.961
Ga	-2.019	-2.031	-2.032	-2.064	-2.076
Y	-2.066	-2.125	-2.174	-2.253	-2.316
Nb	-2.033	-2.061	-2.105	-2.148	-2.198
A^{+2}	Formation energy of $Li_{1+2x}A_xGe_{2-x}(PO_4)_3$				
Ca	-2.052	-2.107	-2.141	-2.195	-2.248
Mg	-2.046	-2.091	-2.132	-2.171	-2.210
Sr	-2.045	-2.105	-2.134	-2.188	-2.243

References

[1] Bishop, C. M. Pattern Recognition and Machine Learning, Springer, New York, NY, 2006.

[2] Theodoridis, S.; Koutroumbas, K. Pattern Recognition, 4th edn, Academic Press, Burlington, MA, 2008.

[3] Cox, D. Analysis of Binary Data. Chapman and Hall, London, 1969.

[4] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proc. of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016, 785–794.https://doi.org/10.1145/2939672.2939785

[5] Guin,M., Tietz, F., Survey of the transport properties of sodium superionic conductor materials for use in sodium batteries, J. Power Sources 273 (2015),s 1056-1064

[6] Wang, W., Li,D., Zhao, J., Solid phase synthesis and characterization of $Na_3Zr_{2-y}Nb_{0.8y}Si_2PO_{12}$ system, Solid State Ionics 51 (1992), 97-100.

[7] Goodenough, J.B., Hong, H.Y-P., Kafalas, J.A., Fast Na^+−ion transport in skeleton structures, Materials Research Bulletin 11 (1976) 203-220.

[8] Losilla E. R., Aranda M. A., Bruque S., Sanz J., París M. A., Campo J. and West A. R. 2000 Sodium mobility in the NASICON series $Na_{1+x}Zr_{2-x}In_x(PO_4)_3$ Chem. Mater. 12, 2134-42 (2000).

[9] Winand, J.M., Rulmont, A., Tarte, P., Ionic conductivity of the $Na_{1+x}M^{III}_x$ $Zr_{2-x}(PO_4)_3$ systems (M = Al, Ga, Cr, Fe, Sc, In, Y, Yb). J Mater Sci 25, 4008–4013 (1990). https://doi.org/10.1007/BF00582473

[10] M.A. Subramanian, P.R. Rudolf, A. Clearfield, The preparation, structure, and conductivity of scandium-substituted NASICONs, J. Solid State Chemistry 60, (1985) 172-181.

[11] Kazakevi̇̂ius, E., et. al, Characterization of $Na_{1.3}Al_{0.3}Zr_{1.7}(PO_4)_3$ solid electrolyte ceramics by impedance spectroscopy, Solid State Ionics 271 (2015), 128-133.

[12] Feltz, A., Barth, S., Preparation and conductivity behaviour of $Na_3M^{II}Zr(PO_4)_3$, (M^{II}:Mn,Mg,Zn), Solid State Ionics, 9 (1983), 817-821.

[13] Yong, Y., Fengqi, Z., Wenqin, P., Hydrothermal crystallization and structural investigation of $Na_{1+2x}Zr_{2-x}Mg_x(PO_4)_3$ systems ($0 < x < 1$), Materials Chemistry and Physics 35, (1993) 28-30.

[14] Delmas, C., Ionic conductivity in Nasicon-type phases $Na_{1+x}Zr_{2-x}L_x(PO_4)_3$ (L = Cr, In, Yb), Solid State Ionics 4, (1981) 209-214.

[15] Xie,H., Lib, Y., Goodenough, John B., NASICON-type $Li_{1+2x}Zr_{2-x}Ca_x(PO_4)_3$ with high ionic conductivity at room temperature, RSC Adv. 1, (2011) 1728-1731.

[16] Shrivastava, O.P., Kumar, N., Chourasia, R. Synthesis, crystallographic characterization and ionic conductivity of iron substituted sodium zirconium phosphate $Na_{1.2}Zr_{1.8}Fe_{0.2}(PO_4)_3$. J Mater Sci 42, 2551–2556 (2007). https://doi.org/10.1007/s10853-006-1230-1

[17] Cherkaoui, F., Viala, J.C., Delmas, C., Hagenmuller, P., Crystal chemistry and ionic conductivity of a new Nasicon-related solid solution $Na_{1+x}Zr_{2-x2}Mg_{x2}(PO_4)_3$ Solid State Ionics 21 (1986) 333-337.

[18] Kumar S., Balaya P., Improved ionic conductivity in NASICON-type Sr^{2+} doped $LiZr_2(PO_4)_3$ Solid State Ionics 296 (2016) 1–6.

[19] Safronov, D., Stenina, I., Maksimychev, A., Shestakov, S., Yaroslavtsev, A., Phase transitions and ion transport in NASICON materials of composition $Li_{1+x}Zr_{2-x}In_x(PO_4)_3$ (x=0-1) Russ. J. Inorg. Chem. 54 (2009) 1697.

[20] Li, Y., Liu, M., Liu, K., Wang, C-A, High Li+ conduction in NASICON-type $Li_{1+x}Y_xZr_{2-x}(PO_4)_3$ at room temperature, J. Power Sources 240 (2013) 50–3.

[21] Ma, Q., et. al., Scandium-Substituted $Na_3Zr_2(SiO_4)_2(PO_4)$ Prepared by a Solution-Assisted Solid-State Reaction Method as Sodium-Ion Conductors, Chem. Mater., 28 (2016) 4821–4828.

[22] Zhang, Z.,et. al., A Self-Forming Composite Electrolyte for Solid-State Sodium Battery with Ultralong Cycle Life, Adv. Energy Mater., 7 (2017) 1601196.

[23] Zhang, Q., Wen, Z., Liu,Y., Song, S., Wu, X., Na^+ ion conductors of glass–ceramics in the system $Na_{1+x}Al_xGe_{2-x}(PO_4)_3(0.3 \leq \leq 1.0)$, J. Alloys and Compounds 479 (2009) 494-499.

[24] Gandi, S., et.al., High Na-ion conducting $Na_{1+x}[Sn_xGe_{2-x}(PO_4)_3]$ glass-ceramic electrolytes: structural and electrochemical impedance studies J. Am. Ceram. Soc. 101 (2018) 167–77.

[25] Safanama, D., Adams, S., High efficiency aqueous and hybrid lithium-air batteries enabled by $Li_{1.5}Al_{0.5}Ge_{1.5}(PO_4)_3$ ceramic anode-protecting membranes, J. Power Sources 340 (2017) 294-301.

[26] Weiss, M., Weber, D. A., Senyshyn, A., Janek, J. R., Zeier, W. G., Correlating transport and structural properties in $Li_{1+x}Al_xGe_{2-x}(PO_4)_3$ (LAGP) prepared from aqueous solution, ACS Appl. Mater. Interfaces 10 (2018)10935-44.

[27] Aono, H., Suginoto, E., Ionic conductivity and sinterability of NASICON-type ceramics: the systems $NaM_2(PO_4)_3 + yNa_2O$ (M=Ge, Ti, Hf, and Zr), J. Am. Ceram. Soc. 79 (1996) 2786-8.

[28] Ruffo, R., Mari, C., Catti, M., Structural and electrical characterization of the NASICON-type $Li_2FeZr(PO_4)_3$ and $Li_2FeTi(PO_4)_3$ compounds, Ionics 7 (2001) 105-108

[29] Aono. H., et. al., Ionic conductivity and sinterability of lithium titanium phosphate system, Solid State Ionics 40 (1990) 38–42.

[30] Rettenwander, D., et. al., A microcontact impedance study on NASICON-type $Li_{1+x}Al_xTi_{2-x}(PO_4)_3$ ($0 \leq x \leq 0.5$) single crystals, J. Mater. Chem. A, 4 (2016) 1506-13

[31] Svitan'ko, A.I., Novikova, S.A., Safronov, D.V. et al. Cation mobility in $Li_{1+x}Ti_{2-x}Cr_x(PO_4)_3$ NASICON-type phosphates. Inorg Mater 47, 1391–1395 (2011). https://doi.org/10.1134/S0020168511120181

[32] Vogel, E.M., Cava, R.J., Rietman, E., Na^+ ion conductivity and crystallographic cell characterization in the Hf-nasicon system $Na_{1+x}Hf_2Si_xP_{3-x}O_{12}$, Solid State Ionics 14 (1984) 1-6.

[33] Aono, H., Sugimoto, E., Sadaoka, Y., Imanaka, N., Adachi, G-Y, Electrical properties and crystal structure of solid electrolyte based on lithium hafnium phosphate $LiHf_2(PO_4)_3$ Solid State Ionics 62 (1993) 309–16.

[34] Wang, W., Liu, X., Study on sodium fast ion conductors in the $Na_{1+3x}Al_xHf_{2-x}Si_2xP_{3-2x}O_{12}$ system, Solid State Ionics 89 (1996) 165-168.

[35] Losilla, E. R., et al., NASICON to scandium wolframate transition in $Li_{1+x}M_xHf_{2-x}(PO_4)_3$ (M=Cr, Fe): structure and ionic conductivity, Solid State Ionics 112 (1998) 53-62.

[36] Novikova, S.A., Larkovich, R.V., Chekannikov, A.A. et al. Electrical Conductivity and Electrochemical Characteristics of $Na_3V_2(PO_4)_3$-Based NASICON-Type Materials. Inorg Mater 54 (2018) 794–804. https://doi.org/10.1134/S0020168518080149

[37] Song, S., Duong, H. M., Korsunsky, A. M., Hu, N., Lu, L. (2016). A Na^+ superionic conductor for room-temperature sodium batteries. Scientific reports, 6(1), 32330.

[38] Jolley, A. G., Cohn, G., Hitz, G. T., Wachsman, E. D. (2015). Improving the ionic conductivity of NASICON through aliovalent cation substitution of Na 3 Zr 2 Si 2 PO 12. Ionics, 21, 3031-3038.

[39] Wang, X., Chen, J., Mao, Z., Wang, D. (2022). Effective resistance to dendrite growth of NASICON solid electrolyte with lower electronic conductivity. Chemical Engineering Journal, 427, 130899.

[40] Chen, D., Luo, F., Zhou, W., Zhu, D. (2018). Influence of Nb5+, Ti4+, Y3+ and Zn2+ doped Na3Zr2Si2PO12 solid electrolyte on its conductivity. Journal of Alloys and Compounds, 757, 348-355.

[41] Huang, C., Yang, G., Yu, W., Xue, C., Zhai, Y., Tang, W., Song, S. (2021). Gallium-substituted Nasicon Na3Zr2Si2PO12 solid electrolytes. Journal of Alloys and Compounds, 855, 157501.