Novelty Detection in the Design of Synthesis of Energy Storage Materials: a Case Study of Garnet-Structured Solid Electrolytes

Natalia Kireeva[a,*], Aslan Yu. Tsivadze[a]

[a]*Institute of Physical Chemistry and Electrochemistry RAS, 31 Leninsky prosp, Moscow, 119071, Russian Federation*

Abstract

Recent decades have shown arising growth-on-demand of integrating the machine learning into all areas of chemistry and materials science. In this study, we consider one of the aspects of applying these technologies to gain advantage in the search for new knowledge extracted from experimental data obtained in ever-growing number of studies. The novelty detection approaches are aimed to identify the artefacts in these data that may be of importance in many directions. The analysis of "outliers" in details of the synthesis in the research studies of garnet-structured solid electrolytes was chosen as the object of demonstration of one of the practical applications of this methodology. Particular attention was paid to the choice of precursors. The thermodynamic data such as the heat of formation from the pure oxides as well as the results of drop solution calorimetry for simple oxides were involved as the descriptors of the studied systems. The overall performance of novelty/outlier detection of all types of outliers was characterized for the data described varying the complexity of description using ROC-AUC statistics and was assessed to be 0.71 – 0.72 using the Area-Under-Curve statistics. It was found that all “outlier” compounds related to those as the result of using the rare precursors in synthesis were successfully identified. The complementary regression analysis was performed to elucidate the relationship between the data diversity and the complexity of data description.

Keywords: novelty detection, solid electrolytes, garnets, design of synthesis, machine learning

1. Introduction

Recent decades have designated the integration of artificial intelligence into all the areas of chemistry and materials science[1, 2, 3, 4]. Materials informatics was emphasized as an efficient way towards the rational synthesis of new compounds and new properties. It was efficiently applied in the design of the materials with desired characteristics[5, 6, 7, 8, 9], in the design of synthesis and for the autonomous laboratories[10, 11, 12, 13], for natural language processing to obtain and analyze experimental data in chemistry and materials science [14, 15, 16, 17, 18], for modeling the microstructure [19, 20], for the analysis of the output of physicochemical methods of characterization[21, 22], for inverse design of materials[23, 24] and in many other areas of their application.

Research of the electrochemical energy storage materials is one of the important directions aimed at general wealth and technological growth. The recognized demand for all-solid-state batteries has led to recent progress in solid-state electrolyte research [25, 26, 27, 28, 29] for multifold applications [30, 31, 32]. Solid state electrolytes are able to replace the organic ones in the batteries of new generation and the cathode materials with the enhanced stability, energy and power densities. The performance of all-solid state batteries is largely defined by the choice of the architecture and the electrode/electrolyte combination [33]. The solid electrolyte characteristics can be considered as the bottleneck of battery performance. Up-to-date garnet-structured solid electrolytes first studied by Weppner and Thangadurai groups [34, 35, 36, 37] are the acknowledged alternatives in all-solid-state batteries which are considered in line with LGPS, argyrodites and LiPONs [33].

The role of the synthesis conditions [38] in the product characteristics is the object of active investigation. Novelty

*Corresponding author

Email address: kireeva@phyche.ac.ru (Natalia Kireeva)

detection approaches can assist in finding the non-standard synthesis routes followed by performing the analysis on the benefits in the functional characteristics gained as a result of the discrepancy in the methods of synthesis.

Nowadays, the use of machine learning methods in synthesis optimization is becoming a common practice in chemistry and materials science. The stage of this involvement is a matter of debate due to the difficult questions on embeddedness of a theoretical background based on kinetics and thermodynamics[39, 38, 40]. In contrast to organic synthesis where the fundamental leap forward was achieved in numerous directions (e.g., in 2001, Hartmuth C. Kolb, M G Finn and K. Barry Sharpless introduced "click reactions" [41]), the questions on capability of accounting the complex processes in solids are still far from being applied routinely in practice [42]. It's worth noting that in [42] author described in the details the general mechanism of the formation of a new phase from the two interacting solid state phases (Figure 1(a)). This bird's-eye view can be used effectively as a basis in the search for new paths in the optimization of the reactions in solids.

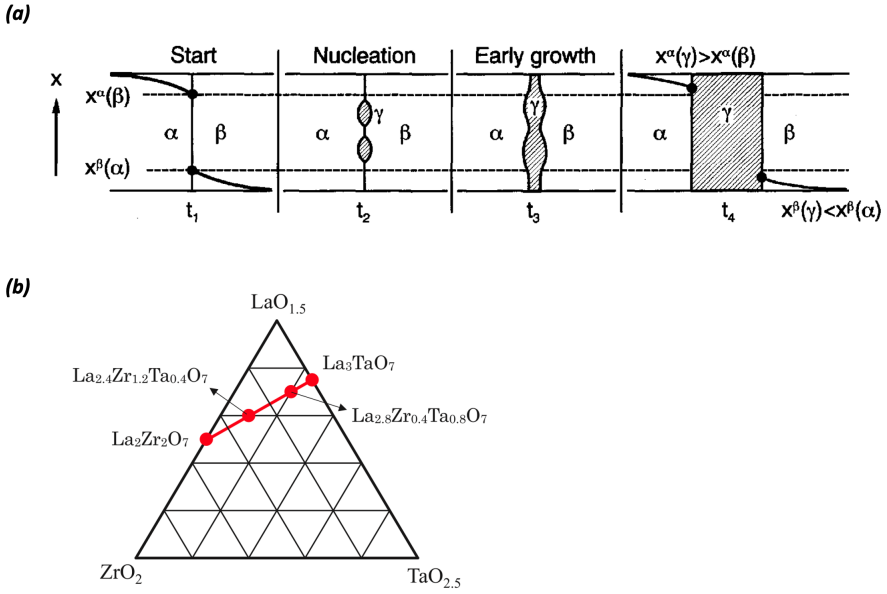

Figure 1: *(a)* The mechanism of a new phase formation from two interacting solid state phases (Figure is reproduced with permission from ref. [42] Copyright 2004, Wiley), *(b)* Quasi-ternary phase diagram of the $LaO_{1.5}$-ZrO_2-$TaO_{2.5}$ system (figure is reproduced with permission from ref. [43] Copyright 2020, Elsevier)

The pertinent examples of integrating machine learning in the full synthesis cycle have been made in numerous studies[12, 10, 13, 44]. Nowadays, it cannot be denied that machine learning methods have already proven themselves to be very effective and in many scenarios critical technologies passing through amounts of experimental data growing in all fields. Such a methodology opens the avenue to multifold applications such as obtaining the metastable compounds away from the equilibrium optimizing the synthesis conditions, to gain the control on the thermodynamics and kinetics of reactions by rationalizing the synthesis. This problem is discussed in [39], where authors have provided with example of the thermodynamic coupling for reaction between $MoCl_5$ and Na_2S, which leads to the formation of MoS_2 and the byproduct NaCl. The formation of the NaCl is exothermic that increases the overall reaction temperature and leading to the melting of the NaCl that in turn results in the enhanced homogeneity and crystallinity of the target product of reaction. One can expect that the progress in using the thermodynamic data in materials informatics will deepen the understanding of the mechanisms of reactions in solids at least due to the enlarged area of visibility for scientists. The role of the nonstandard synthesis route in the final product characteristics due to the new method of synthesis, unconventional treatment or the choice of the precursors is the interesting factor that can be the object of the systematic investigation. In [38] authors have demonstrated the role of the reactivity of the interface of the precursors due to their sequential pairwise participation in the reactions in solids. The developed model provides with the example of how the replacement of the common precursor may redirect the phases evolution to the alternative

reaction intermediates. The replacement of the traditional $BaCO_3$ precursor with BaO_2 resulted in the formation of YBCO in 30 minutes instead of 12 hours. It is well known that the rate of the reactions in some cases has the dominant contribution to the final product formation. In ZrO_2-$YO_{1.5}$ system, the c-δ ($Y_4Zr_3O_{12}$ phase) disorder-order transition is shown to be rate-limited by the cation diffusion and is extremely sluggish[45]. The very long equilibration times are needed to obtain the ordered δ-phase. The authors in [46] assumed that a second yet unseen ordered phase may exist near x = 0.40. The additional investigations showed about 90 energetically permitted intermediates. It is also well known that although the surface energy is formally relatively small, at the nanoscale, when the size of the particle may be quite small, the energetic contribution increases[47]. In [48] it was shown that it can reach the value of 7.9 kJ/mol for TiO_2, which is three times the molar free energy of phase transformation from anatase to rutile. This is a common observation found in many systems, including very classical examples of the systems described by means of the phase diagrams. As was reported for Al_2O_3 (Figure 2(*b*)) the surface area highly affects the enthalpy of the system and may be the reason of the stabilization of the certain phase at given conditions (the slopes represent surface energies). The reactivity of the interface can be assessed through the corresponding enthalpy value. Figure 2(c) shows the the differential heats of water adsorption on hafnia and zirconia surfaces as a function of the method of synthesis and thus the characteristics of the surface [49]. These compounds have the significant importance for diverse technological applications and thus are ones of the most investigated objects. The determination of the value of the excess enthalpy as a function of the surface coverage was recommended as a method for the characterization of the state of the surface for the hydrophilic materials, which surprisingly was not widely used for electrochemical energy storage materials. The involvement of the machine learning in the analysis of the relationship between the synthesis methodology, thermodynamic data and the final product characteristics is the long-standing need.

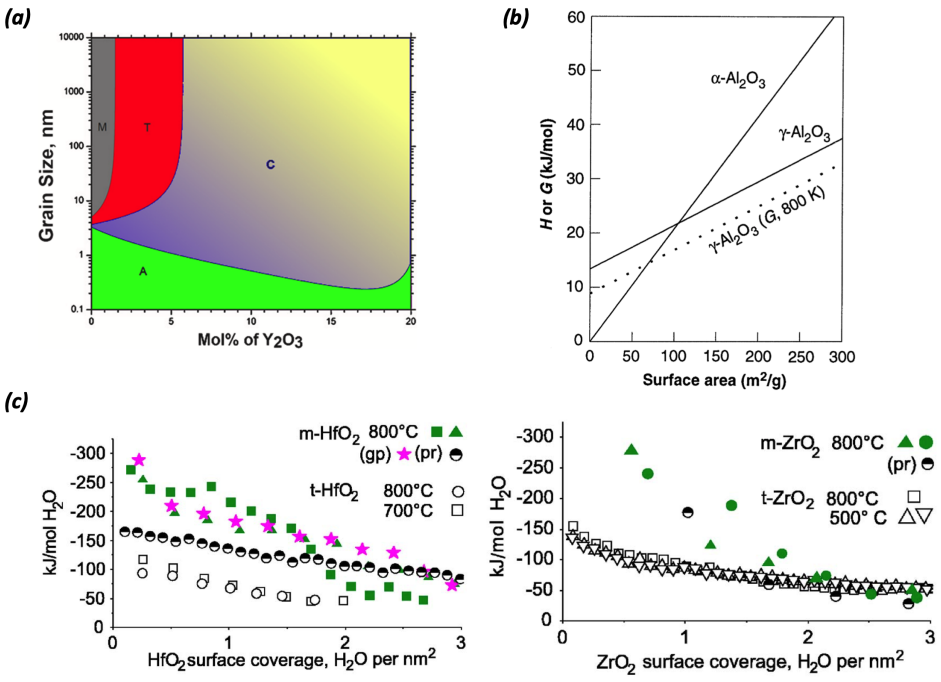

Figure 2: The relation of the microstructure to the target phase stabilization and to the surface energy: *(a)* The phase diagram of the Y_2O_3-ZrO_2 system for dopant and size effects (figure is reproduced with permission from ref. [48] Copyright 2021, Elsevier), *(b)* calculated enthalpy (*H*) of γ-alumina and α-alumina relative to coarse alpha phase, where the slopes represent surface energies (figure is reproduced with permission from ref. [48] Copyright 2021, Elsevier), *(c)* the influence of the defects and the surface state in general as a function of the method of synthesis for HfO_2 and ZrO_2 compounds using the information on differential heats of water adsorption on hafnia and on zirconia surfaces (figure is reproduced with permission from ref. [49] Copyright 2005, AIP Publishing).

The examples given show that the precursors can be considered as the special objects of machine learning-based screening before the synthesis that will be discussed below in a context of using the "unconventional" precursors

in a synthesis of the garnet-type solid electrolytes. In [50] authors argue that machine learning is well-suited for optimization but not for realizing new exceptional materials. The problem of finding outliers with the exceptional properties (authors meant the unique compounds with a combination of diverse characteristics) mentioned in this paper well correlates with another problem of using the synthesis "outliers" to analyze the effects resulted from the unconventional synthesis routes. In our study, we involve the machine learning methods of novelty/anomaly detection as a way to perform such an analysis. Very recently anomaly detection was the methodology of research in a context of identifying the structural anomalies[51].
In this study, we first discuss the experimental data for garnet-structured solid electrolytes of different compositions. Thereafter, we will continue with the description of the machine learning methodology, which provides an efficient way to extract the knowledge from an ever-growing amount of the experimental data, which has a tendency merely progress over time. The results of chosen novelty detection method in identifying the known outliers will be shown using ROC-AUC statistics. We discuss the methodological problems that have come to the fore in numerous studies in statistical learning in a context of the data complexity problem, the influence of the reduction of data description on the statistical characteristics of the developed models is demonstrated for both the classification and the regression problems.

2. Materials and methods

2.1. Experimental data: garnet-structured solid electrolytes

Garnet-structured oxides of general formula $A_xB_3C_2O_{12}$ are known to crystallize in three polymorphs: one tetragonal and two cubic (Figure 3). The different Li ion distribution among the sites with complete ordering of Li ions accommodating tetrahedral sites of the tetragonal polymorph while Li re-distribution among the tetrahedral (*24d*) and octahedral (*96h* and *48e*) sites depending on Li concentration and the composition of the cubic polymorph defines the difference in the ionic conductivity values of ca. 2 orders of magnitude. The tetragonal polymorph is thermodynamically stable at RT while the superionic cubic phase is stabilized by means of doping strategy or the synthesis route/conditions. Several studies have been published on the local structure and the difference in Li distribution as a function of the doping cations Al^{3+}, Ga^{3+} and Fe^{3+} [52, 53, 54, 55]. Figure 3 presents the information on the experimental data used in this work: *(a)* Three polymorphs of garnet-structured solid electrolytes of Li5, Li6 and Li7 phases with the difference in Li^+ distribution among the sites (for two cubic polymorphs).

Figure 4*(a)* shows the distribution of total conductivity σ_{tot} and activation energy E_a values for the data involved in this study in line with (Figure 4*(b)*) the probability density distribution plots for the information on the heating and processing treatment, Li excess and relative density of samples. The similar analysis was recently performed for the argyrodites [57]. Figure 4(*a*) shows that the most of the compounds are related to the phases with log σ_{tot} values in the range -4.0 to -3 and this trend is distinct from the observed values of the activation energy (E_a), where the values are distributed more monotonically. From the Figure 4(*b*), one can see that the calcination temperature and time are characterized by two representative ranges while for the sintering the experimental data are distributed more uniformly. The specific distribution of temperatures and calcination times may indicate that the heat treatment conditions used in [34] proved to be the starting point for optimizing the synthesis of LLZO in further studies. From the experimental data, the evident trend of forming LLZO at much faster rates using sol-gel synthesis was found [58, 59, 60], where $ZrO(NO_3)_2$ are used as the Zr-containing precursors. The hydrothermal/solvothermal route of synthesis and freeze drying methods were not found in the experimental data for garnet-structured solid electrolytes used in this study, however, one may expect the interesting observations as, first, the c-LLZO formation can be reached at the significantly lower temperatures (that is interesting from the point of view of the kinetic and thermodynamic theoretical aspects) and, second, the Li-ion mass transport characteristics may differ from those for the samples obtained by the conventional solid state synthesis due to the different morphology and the defect state of the product [61]. Another underestimation in the chemistry of garnet-type solid electrolytes is the flash sintering methods which demonstrated the perspectives in the control of the reaction kinetics during the process of phase formation [62]. In this case the amounts of the energy supplied to the system is higher in the shorter period of time (due to Joule heating) that may affect the defect state of the structure of the compounds. Several studies involving this method of synthesis were recently published [63], however, the data used for the modeling in this work do not include information on flash

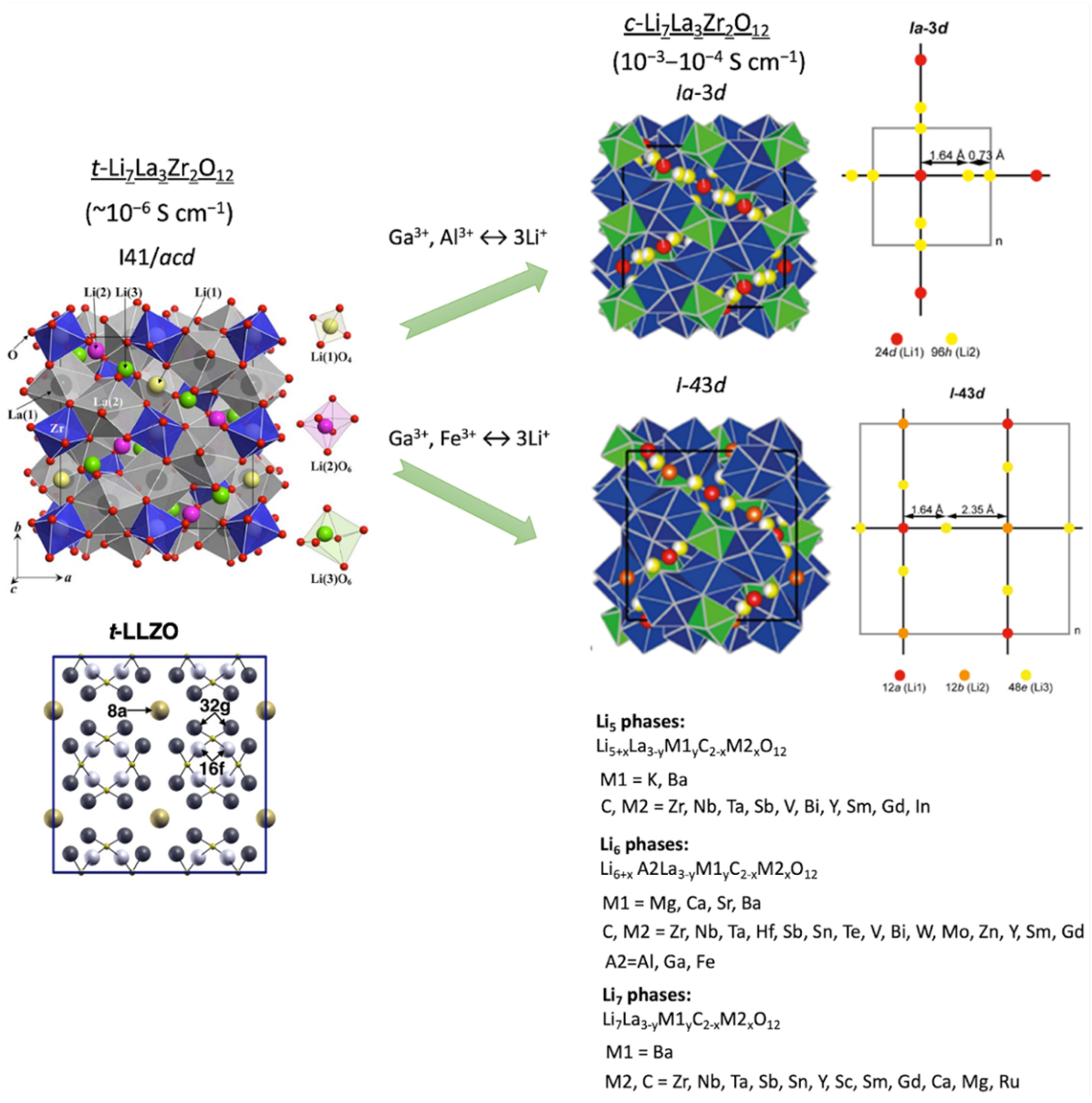

Figure 3: Experimental data used in this work: *(a)* Three polymorphs of garnet-structured solid electrolytes of Li5, Li6 and Li7 phases with the difference in Li^+ distribution among the sites (for two cubic polymorphs). Part of the figure reprinted with permission from [56, 52]. Copyright of American Chemical and Physical Societies.

sintered as well as on hot pressed samples despite the demonstrated effectiveness of these methods as highly distinct microstructure of the samples from that of obtained by the conventional sintering has evident difference which does not require involving any machine learning analysis to be distinguished. In the latter case, the heat in a combination with pressing provides with at least higher density of the samples also introducing the strain that can be relaxed in the certain period of time.

For the analyzed experimental data, the pelletizing pressure has a little relationship with the relative density of the samples according to the analyzed experimental data. The relative density of the samples exceeds 90% for the most of the experimental data. In almost all of the experiments 10 weight % of Li excess were introduced. The mechanochemical treatment encompasses the wide time range of the processing while the high-energy treatment was performed relatively rarely despite its demonstrated efficiency[64] in a number of studies even for poor ion conductors [65] or for the sinterability enhancement [66]. It is worth to note that in [67], author emphasized the role of the nanostructuring for poor ion conductors considering in this study the manifold theoretical insights as a basis for further investigation of the size effects in the systems of different complexity. However, the impact of high energy milling procedures is not evident: first, the point defects are the relevant centers of the reactivity [67] as well as the high-energy ball milling can provide with decreasing the resistance at the interfaces due to the enhanced contact of the particles and with the formation of the non-autonomous phases during the synthesis, second, on the other part, the defect enrichment of the structure can deteriorate the mass transport characteristics in the bulk.

2.2. Investigated alternatives to the conventional Zr and Ta containing precursors

The impact of the choice of the precursors is one of the relatively under-investigated aspects in the research of materials for electrochemical energy storage. In this study, the compounds obtained using non-standard precursors were considered as one of the types of the "outliers". Thus, considering the objects of our study, in [43] authors prepared solid solution system between $La_2Zr_2O_7$ and La_3TaO_7 with a fluorite-type structure $La_{2+x}Zr_{2-2x}Ta_xO_7$ (x = 0.4), using this as the precursor for the synthesis of $Li_{6.5}La_3Zr_{1.5}Ta_{0.5}O_{12}$ allowed to obtain the target composition already at 420°C from the results of TG-DTA measurements. Authors assumed that $La_{2+x}Zr_{2-2x}Ta_xO_7$ (x = 0.4) can effectively act as a precursor oxide because of the same molar ratio of La:Zr:Ta as in $Li_{6.5}La_3Zr_{1.5}Ta_{0.5}O_{12}$. The synthesis of $Li_{6.5}La_3Zr_{1.5}Ta_{0.5}O_{12}$ is performed according to the following equation: $3.25Li_2O + 1.25La_{2.4}Zr_{1.2}Ta_{0.4}O_7 \rightarrow Li_{6.5}La_3Zr_{1.5}Ta_{0.5}O_{12}$ (Figure 1(*b*) presents the corresponding quasi-ternary phase diagram of the $LaO_{1.5}$-ZrO_2-$TaO_{2.5}$ system). The alternative approach was used in [68], where authors have introduced the time-temperature phase diagram for thin films of garnet electrolytes in order to efficiently represent the phase formation process. This information can be related to the problems concerned with the surface energy of the studied system, however, the more precise thermodynamic-based investigations are needed.

For this system, the synthesis of target composition can be considered as a two-step process. First, the formation of solid solution with the same stoichiometry of La, Zr and Ta is performed followed by obtaining the final composition introducing Li_2O in synthesis (the process is presented in the details in Figure 5 (a)). The heat treatment for obtaining the precursor comprises pre-heating at 350°C for 2 h followed by the calcination at 800-1100 °C for 4 h in air. Second, a 50 wt% excess Li_2O was added to increase the reaction area between Li oxide and the precursor oxide. Then, the mixture was calcined at temperature from 400 to 500°C for 12 h in an Ar atmosphere inside Al_2O_3 crucible. The obtained sample exhibited a bulk Li-ion conductivity of $9.4\cdot 10^{-4}$ $S\cdot cm^{-1}$ at 25°C that, according to the authors, slightly outperforms the counterparts of the same composition synthesized by the solid-state route. In [69] authors investigated two alternative routes of synthesis of *c*-LLZO through two different intermediates formation: *(i)* from Li_2CO_3, $La(OH)_3$ and $La_2Zr_2O_7$ (instead of ZrO_2) precursors authors have obtained $La_2Zr_2O_7$ at 650°C followed by formation of the cubic phase single-phase LLZ by heating the initial mixture to 800°C for 1 hour in air and *(ii)* from Li_2CO_3, $La(OH)_3$ and ZrO_2 precursors, Li_2ZrO_7 was formed at 700-850°C with $La_2Zr_2O_7$ as a secondary phase accompanied by the formation of a target *c*-LLZO phase mainly at 700°C-750°C. In [70] it was found that $La_2Zr_2O_7$ is a part of a large lithium containing solid solution region, where the search of structures with high surface reactivity can be of practical interest.

Figure 5 also presents the information on synthesis for the alternative Zr-containing precursors and modified LLZO synthesis routes, $La_2Zr_2O_7$, YSZ and *t*-ZrO_2. In [71] authors have investigated the question on crystallite size effects in a context whether the different crystal structures of ZrO_2 affect the synthesis of LLZO mentioning the surface

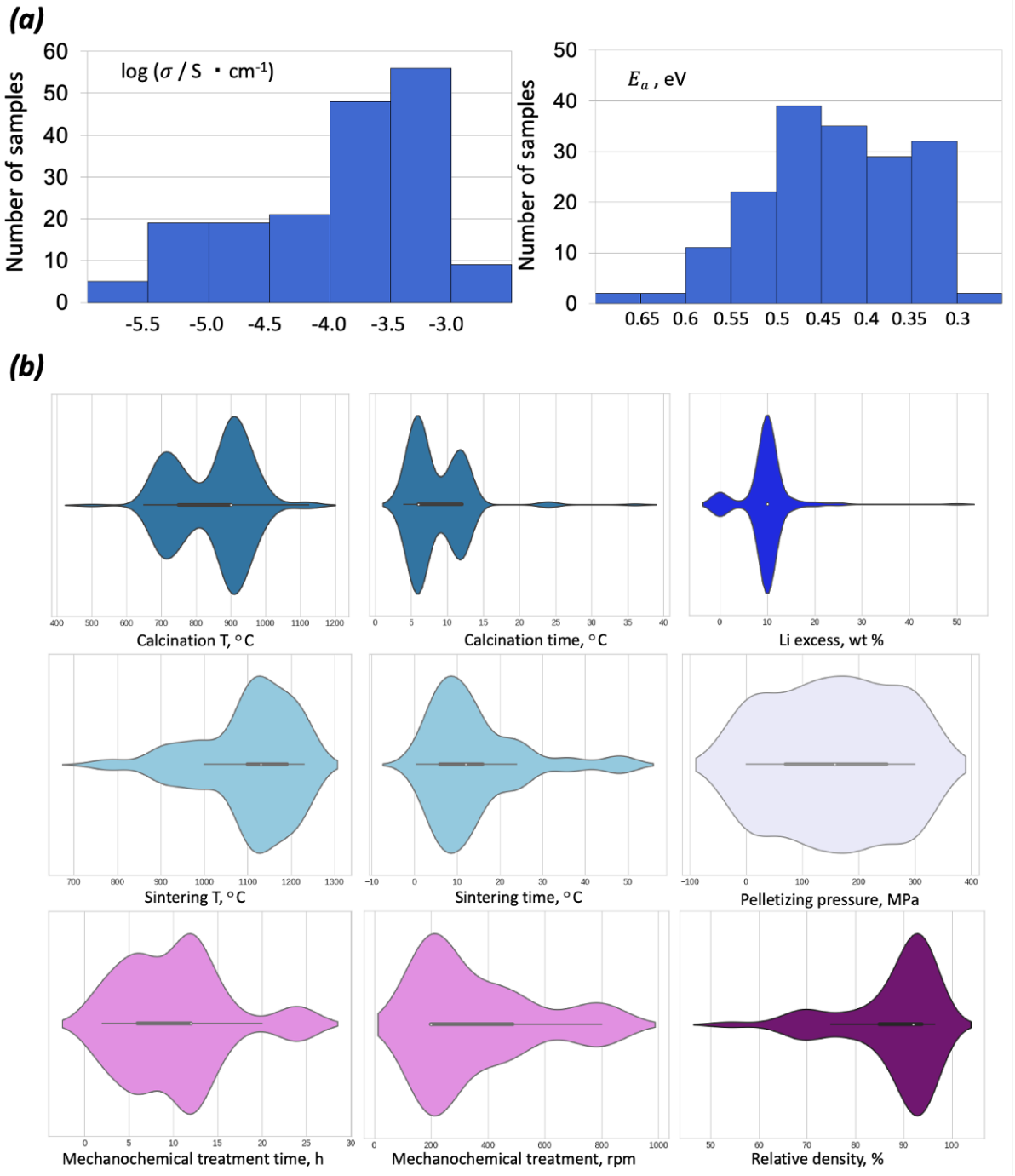

Figure 4: The distribution of total conductivity σ_{tot} and activation energy E_a values *(a)* in line with *(b)*) the probability density distribution plots with the information on the heating and processing treatment regimes, Li excess introduced during the synthesis and relative density of samples.

energy as an important factor highly affecting the synthesis. Three polymorphs of pure ZrO_2 exist at 1 atm: *(i)* monoclinic (*m*), tetragonal (*t*) and cubic (*c*). Monoclinic polymorph (*m*-ZrO_2), which is stable at room temperature (RT) transforms to the tetragonal structure (*t*-ZrO_2) at about 1200°C with an associated enthalpy of transition $\Delta H_{m\rightarrow t}(ZrO_2)$ of 5.94 ±0.4 kJ/mol. Tetragonal polymorph remains stable up to 2377°C, where it transforms to the cubic modification [45]. Monoclinic and tetragonal ZrO_2 precursors were used in the abovementioned study, where *t*-ZrO_2 was produced by caustic fusion of $ZrSiO_4$ followed by heating at 1000°C. The ball milling for 8 hours and 450 rpm in isopropanol was performed with the obtained crystallite size of 190 nm and 122.17 nm for *m*-ZrO_2 and *t*-ZrO_2, respectively (thus, according to Figure 2(*a*, no inclusion of the amorphized phases is expected).The heat treatment was 950°C for 6 hours at the calcination step and 1000°C for 6 hours at sintering step. The presence of $LiNaCO_3$ phase in LLZO synthesized from *t*-ZrO_2 was found. Authors relate this fact with slightly higher values of ionic conductivity ($1.647\cdot10^{-6}$ S·cm^{-1} of LLZO synthesized from *t*-ZrO_2 vs. $1.245\cdot10^{-6}$ S·cm^{-1}) of LLZO synthesized from *m*-ZrO_2. One may assume over-heating during the synthesis resulting in the dominant *t*-LLZO formation. The corresponding scheme of synthesis is shown in Figure 5(*c*). Zircon $ZrSiO_4$ can be obtained as a product of reaction of a quartz crystal (SiO_2) and *m*-ZrO_2 by heat treatment at 1000 K for 3 days. The enthalpy of reaction of the formation of $ZrSiO_4$ from its constituent oxides has been determined $\Delta_r H_{997}(ZrSiO_4)$ = -27.9±1.9 kJ/mol[72]. The studies using *c*-ZrO_2 as a precursor were not found. However, in [73] authors have used polymerized complex method for the synthesis of $Y_4Zr_3O_{12}$ and $Y_{3.93}Yb_{0.07}Zr_3O_{12}$ with yttrium (III) nitrate hexahydrate, zirconium (IV) oxynitrate hydrate and ytterbium (III) nitrate pentahydrate. The X-ray diffraction patterns of powders of YZO undoped and doped with Yb^{3+} match well with those corresponding to the cubic-like fluorite crystal phase of ZrO_2 (PDF-30-1468) when it is stabilized with concentrations higher than 15 mol%. The perspectives of using this compound for the synthesis of, for example, Y-doped LLZO compounds are not clear. The difference with 8 mol% YSZ (being used for the synthesis of garnet-type solid electrolytes as a precursor and discussed below) in the reaction behavior can be expected due to difference in the structure concerned with the long-range order of vacancies in the fluorite-related structure of $Y_4Zr_3O_{12}$. In [43] and [74] authors have unraveled the synthetic methodology for obtaining the target garnet-type compounds by using conventional solid state method as well as the polymerized complex method from $La_2Zr_2O_7$ and La_3TaO_7 Zr and Ta-containing precursors. To our knowledge, these studies and the works of Kimura[69] and [75] were the first ones, where the questions on the influence of using the compounds formed as the intermediates in the synthesis as the precursors were studied. The scheme of the synthesis is given in Figure 5. In [43] at the initial step of the synthesis by the method of polymerized complex $ZrOCl_2\cdot8H_2O$ was used as a source for $La_2Zr_2O_7$ synthesis. The information on the synthesis of garnet-type solid electrolyte materials from the chlorides only was not found in the literature. However, for Li-rich layered oxides the evident benefits for the Li-ion mass transport characteristics were not observed.

In [76] authors have used 3 and 8 % Y-doped ZrO_2 (YSZ) for the synthesis of highly Li-stuffed $Li_{7.06}La_3Zr_{1.94}Y_{0.06}O_{12}$ and $Li_{7.16}La_3Zr_{1.84}Y_{0.16}O_{12}$ by conventional solid state method of synthesis (Figure 5(*b*)). According to the phase diagram represented at Figure 2(*a*), 3 mol% YSZ crystallizes in the tetragonal polymorph while 8 mol% YSZ crystallizes in the cubic modification (the corresponding enthalpies of formation of these compounds from the simple oxides are given below as being used as the descriptors in modeling). Other precursors used in synthesis were La_2O_3 and $LiNO_3$ (10 weight % excess).The powders were ball-milled for 10-24 hours using zirconia balls in isopropanol. Pre-calcination was performed at 700 °C for 12 hours followed by re-grinding for 24 hours and sintering at 950°C for 12 hours. Non-uniform distribution of yttrium is observed for elemental mapping presented for 8 mol% YSZ-doped $Li_{7.16}La_3Zr_{1.84}Y_{0.16}O_{12}$. The Li-ion transport characteristics were found very similar to the studied samples. However, one may expect this as the result of using the same synthesis conditions. In [42] Joachim Maier noticed: "The reactive adsorption step and the reorganization are followed by the nucleation of the oxide phase, that either takes place homogeneously or occur heterogeneously with the aid of crystal defects. In the second case the surface energy of the defect is of prime significance." Thus, the precursors characterizing by the different surface energies and the defect-enrichment states (Figure 2(*c*)) may give very different enthalpies of reactions. Similarly, the problem of the relationship between the association of the defects and the ion conductivity is well-known [45].

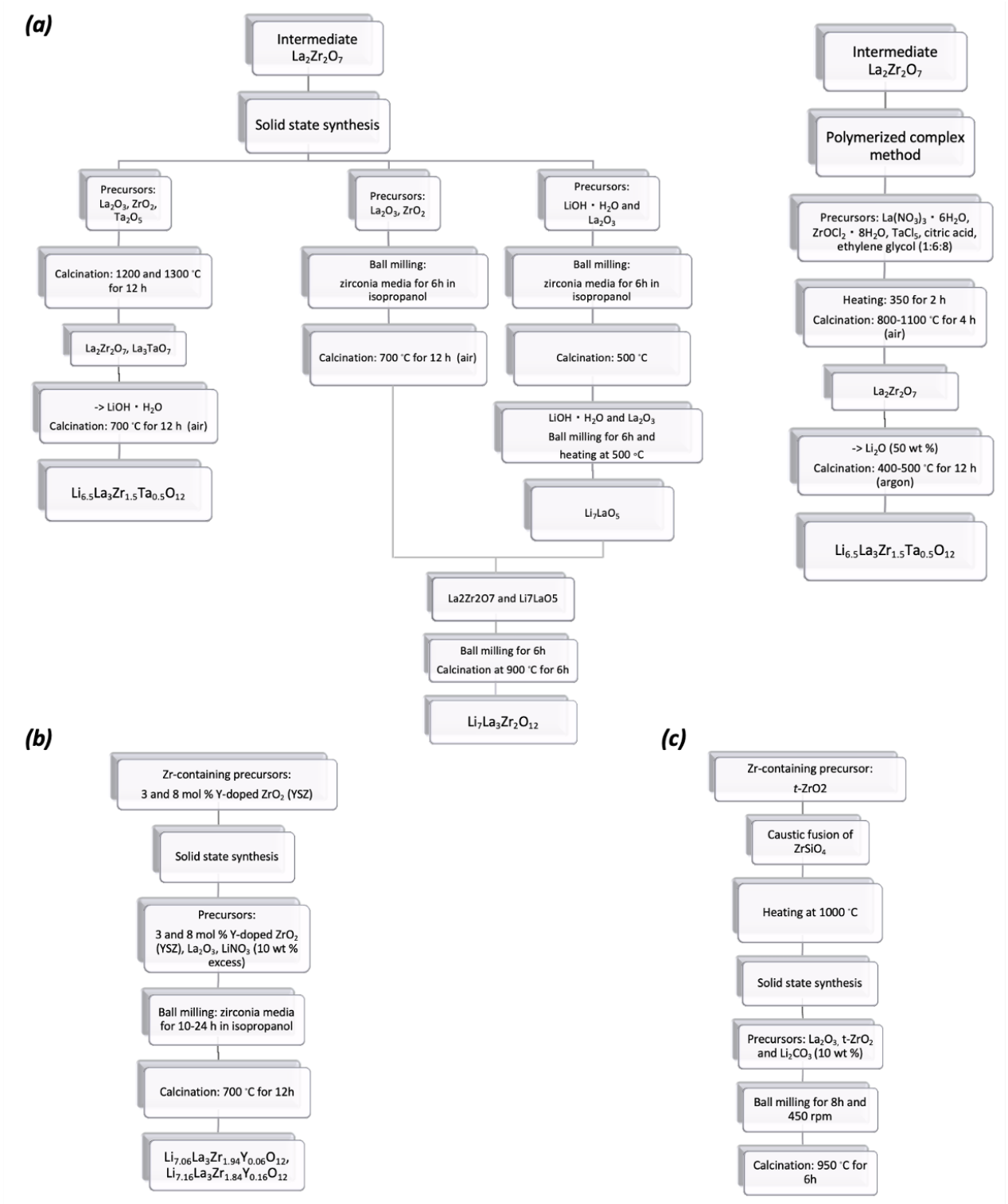

Figure 5: Information on synthesis using the alternative Zr- and Ta-containing precursors ($La_2Zr_2O_7$, YSZ and *t*-ZrO_2, La_3TaO_7) and the modified LLZO synthesis routes.

2.3. Machine learning in a design of synthesis

The most promising practice for planning synthesis at the moment is the automation of literature analysis followed by the use of natural language processing technologies (NLP). In such a way the important findings have been made in numerous studies. However, each time these findings are almost completely the result of the analysis depth of the obtained results and the assumptions have been made. In [77] using the data set of over 30 000 text-mined solid-state synthesis reactions and guided by the assumptions on the *(i)* good synthesizability of the compounds, *(ii)* performing the synthesis experiments in a one-shot fashion and *(iii)* predicting the "optimal" synthesis conditions authors have underlined the importance of the heating time for the compositions containing (1) Mn, Ru, Rb, In, Ti, Cd, Th, Hf, Pb and Te cations and (2) of the heating temperature for the compositions including Li, Mo, Ba, Sr, V, Bi and Te. In [78] authors have used the text mining and natural language processing to extract the information on the synthesis protocols for gold nanoparticles associating the synthesis routes/conditions with a shape of the particles. In [15] authors have used the information on sintering and calcination temperature and time, method of synthesis and solvent, which was extracted from the literature and using the NLP software learned model to discriminate synthesis details for $SrTiO_3$ and $BaTiO_3$ compounds as well as to elucidate the formation of certain polymorph of MnO_2. In [79] authors have introduced the similarity metric in the assessment of the similarity in the synthesis routes for inorganic materials. The proposed approach has allowed to capture the correlation between the target and the precursors as well as the dependency between the different precursors in the same experiment. Authors have randomly masked a part of the precursors while remained data have been used to imputing the set of precursors. This methodology has a high potential to be used for solving the more general problem of the imputation of the experimental data in chemistry and materials science. Among the dilated set of examples authors refer to the study[80], where $NaHPO_4$ have been used to synthesize $Na_3TiV(PO_4)_3$ while the common precursors are Na_2CO_3 and $NH_4H_2PO_4$ Additionally, authors unexpectedly found that approximately half of the target materials were synthesized using at least one uncommon precursor. The similarity-based approach has allowed to transfer the use of the former precursor for the synthesis of $Na_3V_2(PO_4)_3$. In [81] the authors jointly used the automatic data extraction methodology with state-of-the-art algorithm Autonomous Reaction Route Optimization with Solid State Synthesis (ARROWS), an algorithm designed to optimize the solid-state synthesis including the problem of the precursors selection. Based on the initial ranking of precursors combinations according to the DFT results on the reaction energies the authors conducted a systematic search over a wide range of temperatures and synthesis times. The outcome of reactions was analyzed by XRD thus identifying all the products. The information on the intermediates obtained during the synthesis was also obtained. The mentioned study is a continuation of many others among which one can refer [82]. The developed methodology can be considered one of the most attractive to date due to the obvious prospects for its easy extension to other objects with the only limitation associated with synthesis.

2.4. Methods of novelty/anomalies detection

The concept of the density of the probability distribution P is laid at the core of the richness of methods for anomalies/novelty detection[83]. In practice of the materials science and chemistry this means that any experimental data may be analyzed in the context of how the obtained observations fit into a bigger picture of already obtained knowledge that can be enforced by initially statistical learning-based principle formulated by Vladimir Vapnik "*never to solve a problem which is more general than the one we actually need to solve*" [84] laying in the basis of numerous studies [83]. Anomalies detection techniques can be efficiently applied for the allied purpose of circumscribing the experimental data sufficient for the detailed analysis of the certain studied phenomena thus removing the observations involving no new knowledge. In the underlying principle, the latter resonates with the training data attribution methodology [85].

Discussing the problems of the so called "outliers", the real data that differ from what may be expected and can be identified as obeying no statistical law, one may also address to the experience of statistical learning, where the importance of the identification and minimization of such observations are positioned as one of the cornerstones[86]. The assumption on the majority of the confident "normal" data is usually made. Thus, using the terminology of machine learning, the task is formulated as the unsupervised learning, where the data described by means of some selected features and the learning process itself does not require the knowledge on the associated with these data properties. The decision on the "normality" or "outlierness" of the data point is passed according to the density of the probability distribution of the analyzed data. The binary classification of the data to the nominal data and the "outliers" is usually performed

by involving the methods that can be classified into four main groups: *(i)* boundary-based, *(ii)* reconstruction-based, *(iii)* embedding similarity-based, *(iv)* classic density estimation methods and *(v)* distance-based. Figure 6 presents *(a)* the schematic representation of these main categories of methods widely used for the outlier/anomalies detection: boundary-based, reconstruction-based, density estimation-based and distance-based methods and *(b)* Deep Autoencoding Gaussian Mixture Model.

Among the boundary-based methods several approaches including the "traditional" ones are most often used: *(i)* one-class classifiers, One-Class Support Vector Machines (OC-SVM) [83], Support Vector Domain Description (SVDD) [87], *(ii)* deep one-class classification [88] (DSVDD) and *(iii)* one-class neural networks [89]. It is widely acknowledged that the performance of OC-SVM can be sub-optimal on complex, multivariate datasets[87, 84] and this problem is discussed for a number of examples as a trade-off between reducing the density uncertainty for the multidimensional description of the data and the "spatial" uncertainty choosing simplified data description [90].

The reconstruction-based approaches use the concept of the reconstruction (quantization) error or the energy values for estimating data density distribution. Among the reconstruction-based methods the most popular and known methods are different types of autoencoders (deep autoencoders [91, 92], neural generative models (variational autoencoders [93], contrastive autoencoders [94], adversarial autoencoders [95], autoencoders with nonlinear dimensionality reduction [96], different types of generative adversarial networks GANs [97]), hybrid approaches (Deep Autoencoding Gaussian Mixture Model [98], deep anomaly detection using geometric transformations[99], kernel density estimators (Kernel Density Estimation [100] and Robust Kernel Density Estimation [101]), Robust-PCA [102].

Embedding similarity-based and distance-based methods for outlier/anomalies detection includes PaDiM [103], Isolation Forest (iForest) [104] and Local Outlier Factor (LOF) [105]. The anomaly identification in the latter case is performed using the distance between the embedding vectors of a test example and the reference vectors representing the normality of the training dataset. The similar principle is put behind the well-known dimensionality reduction technique of Self-Organizing Maps [106], where the quantization error is used as a measure characterizing the similarity of the considered sample to the nominal data. In the former case, the pre-trained networks are used.

Among the classic density estimation methods for outlier detection one should distinguish manifold learning-based (Regularized Principal Manifolds [107], Bayesian estimation of assignments of objects to mixed classes [86] and Generative Topographic Maps [108]. The "outlierness"of the data is evaluated using the predictions of the probability of mixture membership for each sample. Adversarial Autoencoders introduced by Pidhorskyi et al. [95] and Zong et al. [98] can be related to some extent to this category as hybrid approaches.

Recently a comprehensive review describing in the details the variety of the anomalies approaches was published[109]. In this study, we use the energy values evaluated by Deep Autoencoding Gaussian Mixture Model[98] as a criteria for identifying the novelty/outliers in the data.

Deep Autoencoding Gaussian Mixture Model In its original form, the autoencoders that are nowadays at the root of one of the most popular family of statistical learning methods were introduced in the late 80's [110]. This type of neural networks was used as a basis for the deep autoencoders gained traction during the last decades simultaneously with onrush of the methodology of deep learning [111]. The method involved in this study, Deep Autoencoding Gaussian Mixture Model (DAGGM), is one of the representatives of deep autoencoders enhanced by Gaussian Mixture Model thus integrating it with a probabilistic approach.

DAGGM consists of two components: the compression network and the estimation network. The compression network functionalizes in an enforced manner by outputting the data of two types: *(i)* the reduced low-dimensional representation learned by a deep autoencoder and *(ii)* the data derived using the reconstruction error evaluated by the decoder component. The estimation networks uses both type of information to evaluate the likelihood/energy by using Gaussian Mixture Model (GMM). Given the low-dimensional representation, the estimation network estimates the probability density of the data by using GMM. During the training with some initialized mixture component distribution ϕ, the mixture means μ, and the mixture covariance Σ, the network estimates the parameters without using procedures as Expectation-Maximization (EM) by means of multilayer neural network to predict the mixture membership for each sample.

$$p = MLN(z;\Theta_m) \tag{1}$$

$$\gamma = softmax(p) \tag{2}$$

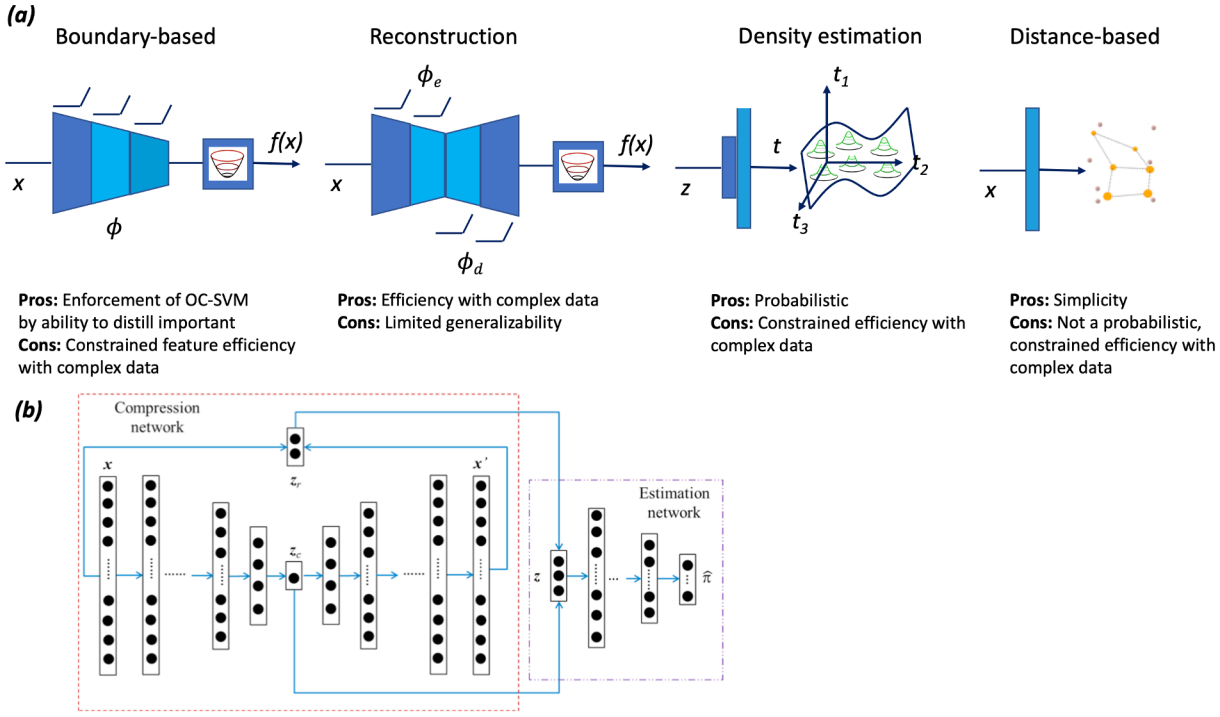

Figure 6: Schematic representation of main categories of methods widely used for the outlier/anomalies detection: boundary-based, reconstruction-based, density estimation-based and distance-based methods; *(b)* Deep Autoencoding Gaussian Mixture Model.

where γ is a K-dimensional vector for the soft mixture-component membership prediction, K is a number of Gaussians and p is the output of a multilayer network parameterized by Θ_m. This proposed architecture allows one to preserve and use the capabilities of the basic algorithm as well as to gain the probabilistic assessment of the data.

2.5. Methods of dimensionality reduction

***t*-Distributed Stochastic Triplet Embedding** In this study, we use the method of Stochastic Triplet Embedding for the dimensionality reduction[112]. This method exploits the concept underlying in a human system of judgments based on the principles of relative similarity of objects (A is more similar to B than C) which is realized in the framework of stochastic neighbor approach by maximizing the sum of the log probabilities of fairness of a given statement over all considered triplets.

$$\max_{X} \sum_{\forall(i,j,l)\subseteq(\tau)} logP_{ijl} \tag{3}$$

Authors introduce a Student-t kernel with α degrees of freedom. Hence, the probabilities are defined based on the local similarities as follows:

$$p_{ijl} = \frac{(1+\frac{||x_i-x_j||^2}{\alpha})^{-\frac{\alpha+1}{2}}}{(1+\frac{||x_i-x_j||^2}{\alpha})^{-\frac{\alpha+1}{2}} + (1+\frac{||x_i-x_l||^2}{\alpha})^{-\frac{\alpha+1}{2}}} \tag{4}$$

During the process of training, the relative coordinates of the samples are perturbed to improve the decision of the model according to equation (4). This procedure similar to the metric learning principle allows both to close the coordinates of the similar observations in the map minimizing the distance between them while to spread apart the dissimilar according to the data description objects. In this study, we use this method in line with the energy values obtained from the DAGGM to visualize the experimental data with the indication of the compounds, for which the probability being the novelty/outlier is higher.

2.6. *Machine learning methods used for regression problem*

Support Vector Machines method (SVM)[84] was involved for model development as it is implemented in LIBSVM package [113] with settings of using ε-SVR and Radial Basis Function (RBF) kernel.
The second approach is the Probabilistic Backpropagation Bayesian Neural Networks (PBP) [114]. Given data $\mathscr{D} = \{x_n, y_n\}_{n=1}^{N}$, where $x_n \in \mathbb{R}^{\mathscr{D}}$ and corresponding scalar variables $y_n \in \mathbb{R}$, $y_n = f(x_n; \mathscr{W}) + \varepsilon_n$, where $f(\cdot; \mathscr{W})$ is the output value with weights given by $\mathscr{W}$ and noise variables ε_n.
The likelihood of dependent variable given weights $\mathscr{W}$ and the noise precision γ is defined as following:

$$p(y|\mathscr{W}, X, \gamma) = \prod_{n=1}^{N} \mathscr{N}(y_n|f(x_n; \mathscr{W}), \gamma^{-1}) \tag{5}$$

PBP does not use the point estimates of the weights during the training instead the set of the Gaussians is generated:

$$p(\mathscr{W}|\lambda) = \prod_{l=1}^{L} \prod_{i=1}^{V_l} \prod_{j=1}^{V_{l-1}+1} \mathscr{N}(w_{ij,l}|0, \lambda^{-1}) \tag{6}$$

where $w_{ij,l}$ is the weights and λ is a precision parameter. The details on how the prior for λ is defined are given in the original publication.
The posterior distribution for the parameters $\mathscr{W}$, γ and λ can then be obtained according to Bayes' rule as follows:

$$p(\mathscr{W}, \gamma, \lambda|\mathscr{D}) = \frac{p(y|\mathscr{W}, X, \gamma)p(\mathscr{W}|\lambda)p(\lambda)p(\gamma)}{p(y|X)} \tag{7}$$

The output predictions are performed using predictive posterior distribution:

$$p(y_{target}|x_{target}, \mathscr{D}) = \int p(y_{target}|x_{target}, \mathscr{W}, \gamma)p(\mathscr{W}, \gamma, \delta|\mathscr{D})d\gamma d\lambda d\mathscr{W} \tag{8}$$

where $p(y_{target}|x_{target}, \mathscr{W}, \gamma) = \mathscr{N}(y_{target}|f(x_{target}), \gamma)$. At the end of the forward stage, PBP computes the logarithm of the marginal probability of the dependent variable. At the stage of backward propagation, the network propagates the gradient of this quantity with respect to the means and the variances of the approximate Gaussian posterior, which in turn are used to update the corresponding values of the means and the variances of the posterior approximation of the Gaussians. The weights are updated according to the Bayes' rule:

$$s(w) = Z^{-1} f(w) \mathscr{N}(w|m, v) \tag{9}$$

where Z is the normalization constant. The updated values for the means and the variances are obtained using the gradient of the logarithm of the normalization constant Z:

$$m^{new} = m + v\frac{\partial logZ}{\partial m} \tag{10}$$

$$v^{new} = v - v^2 \left[\left(\frac{\partial logZ}{\partial m} \right)^2 - 2 \left(\frac{\partial logZ}{\partial v} \right) \right] \tag{11}$$

The third method is Deep Gaussian Processes (DGP) [115]. DGP are a deep belief network based on Gaussian process mapping. The Gaussian processes are organized in the hierarchy in a way similar to the neural networks, where the inputs of one layer are the outputs of the previous layer. A single layer model is equivalent to the individual GP or the GP latent variable model. The benefits of this methodology is two-fold: *(i)* deep models are known to have the advantages over the shallow methods as they are able to extract more complex relationships from the processed data and *(ii)* the proposed method allows one to perform some sort of feature weighting that significantly improves the stability of the models for the data of different complexity. Principally, authors have realized the variational inference to marginalize the latent variables in the hierarchy variationally. The probabilistic nature of the algorithm additionally contributes to the model stability that otherwise can be sensitive to the outlier objects.

The proposed architecture can be reprsented as a graphical model with three kinds of nodes: *(i)* the leaf nodes $Y \in \mathbb{R}^{N \times D}$, where D is the number of Gaussian process priors, *(ii)* the intermediate latent spaces in amount of the hidden layers in the architecture $X_h \in \mathbb{R}^{N \times Q_h}$, where $h = 1,..., H$ (H is the number of hidden layers) and *(iii)* the parent latent node Z that can be unobserved in the architectures of the special purposes $Z = X_H \in \mathbb{R}^{N \times Q_h}$.
For the architectures containing only two hidden units, the generative process takes the form:
$y_{nd} = f_d^y(x_n) + \varepsilon_{nd}^y$, where $d = 1,..., D$; $x_n \in \mathbb{R}^Q$
$x_{nq} = f_q^x(z_n) + \varepsilon_{nq}^x$, where $q = 1,..., Q$; $z_n \in \mathbb{R}^{Q_z}$.
where f^y and f^x are the input and the output, respectively. As each layer contains a significant number of model parameters to be taken into account, authors have proposed two features that enforce this method at the same time allowing to solve the problem of regularization. At the first step, they have introduced the automatic relevance determination (ARD) covariance functions for the GPs:
$K(x_i, x_j) = \sigma_{ard}^2 e^{-1/2 \sum_{q=1}^{Q} w_q (x_{i,q} - x_{j,q})^2}$

Thus, the weights are effectively introduced into the model for each latent dimension. This allows one to automatically determine the most important information ignoring the possible deficiencies in the data description. To avoid the difficulties for Bayesian treatment of the introduced nonlinearities, which are, however, of principal importance for the weighting, the special pseudo-inputs are introduced that are known as the inducing points. Their number K is defined in the configuration of the model. This allows to realize some sort of the surrogate function defining the performance of the models developed during the training procedure. For the details concerned to the Bayesian training, please, refer to the original publication.

2.7. Descriptors

The parameters used in the study as the descriptors can be related to several categories: ***(i)*** a composition of the compounds, ***(ii)*** atomic characteristics (Shannon ionic radii, atomic scattering factor, dielectric polarizability and atomic weight values) [116], ***(iii)*** details of synthesis including Li excess, temperature of decomposition for Li precursors, calcination and sintering time and temperature.
The atomic characteristics are well-known and successfully applied in materials informatics as a simple materials fingerprints [117, 3].
In recent years the descriptors describing the process of synthesis were efficiently used in materials informatics [15, 118, 10, 119]. Authors of this study have used this type of descriptors for modeling the functional characteristics of the materials for the electrochemical energy storage [120, 121, 122]. The important role of these descriptors is in the possibility to take into account the number of factors related to the synthesis that otherwise remain not described. The thoroughness of the introduced description for the compounds of different compositions largely defines the precision of the models, however, placing the question on the compromise between the complexity and the uncertainty in the spatial data description and the inherent capabilities of the machine learning methods. The need in using the special data imputation techniques can arise as a result of increased complexity in data description. In this study, several descriptors from this category were used. The heat treatment were described by means of four values referred to calcination and sintering temperature and time, respectively. These values were normalized to zero-to-one scale using the minimal and maximal temperature and time values taken from the collected experimental data. The temperature of decomposition of Li precursors was also normalized. The information on Li excess was introduced without the normalization.
Particular attention in the analysis of novelty detection in design of synthesis in this study was given to the choice of C cation precursors. The thermodynamic data such as the heat of formation from the pure oxides as well as the results of drop solution calorimetry for simple oxides were involved as the descriptors of the studied systems. The enthalpy of formation of *c*-YSZ $\Delta H_{f,ox}$ was defined with respect to the oxides *m*-ZrO_2 and C-type $YO_{1.5}$ as equal to the weighted sum of the enthalpies of transition plus the enthalpy of mixing in the cubic solid solution [45]:
$\Delta H_{f,ox}$ (c-YSZ) = [(1–*x*) $\Delta H_{m \to c}(ZrO_2)$ + $x\Delta H_{C \to c}$]+ ΔH_{mix}
For 8YSZ and 3YSZ this value was defined as $\Delta H_{f,ox}$ (8YSZ) = +0.78 kJ/mol and $\Delta H_{f,ox}$ (3YSZ) = +4.2 kJ/mol, respectively ($\Delta H_{m \to c}(ZrO_2)$ = 6.1 kJ/mol). For other oxides, the data provided in a literature was taken. We also indicate here the enthalpy value for $La_2Zr_2O_7$ $\Delta H_{f,ox}$ ($La_2Zr_2O_7$) = -136.1 kJ/mol.

In a case of several cations accommodated at the C site of garnet structure this value was taken for the major one in the composition. We consider this decision as appeared the best found.
The complete data description involves the following descriptors: compositional (nine elements in the string), atomic characteristics (Shannon ionic radii, atomic scattering factor, dielectric polarizability (according to Shannon), atomic weights), the synthesis details (Li excess, T of decomposition of Li precursors, enthalpies of formation from the pure oxides as well as the results of drop solution calorimetry for simple oxides) for C cation precursors, calcination time and temperatures (two-stage heat treatment was assumed), sintering temperatures (two stage heat treatment was assumed) (overall 48 descriptors). The reduced data description involves the following descriptors: Li content, La content, Atomic scattering factor for cation C, Shannon ionic radius of Li substituent, polarizability of cation C, Shannon ionic radius of the substituent of the cation C, content of cation C, Calcination time (stage 1), content of Li substituent, content of substituent of cation C, sintering temperature (stage 1), calcination temperature (stage 1), calcination temperature (stage 2), T of decomposition of Li precursor, enthalpies of formation from the pure oxides as well as the results of drop solution calorimetry for simple oxides) for C cation precursors (overall 21 descriptors). The content of the pool was defined as a result of the analysis performed in our recent study [121].

2.8. Computational procedures

2.8.1. Validation and statistical parameters of models

The overall performance of novelty/outlier detection of all types of outliers was characterized using neural networks with deep autoencoder architecture to describe data of varying complexity using ROC-AUC statistics, where the Receiver Operating Characteristic (ROC) is determined as a result of binary classification, where the confusion matrix of true positives, true negatives, false positive and false negatives is used to determine the true positive and false positive rates, TPR and FPR, respectively. The ROC curve (TPR as a function of FPR values) was build, where the decision on the success in prediction was taken as a result of the comparison of the class("outlier"/"normal") labels for each individual sample pre-defined based on the analysis of the experimental data and taking into account the set of descriptors (using the complete or reduced data description) with the labels obtained as a result of the estimation of the energy value by DAGGM, with the threshold values defined as 0.195 and 0.435 for the complete and reduced data description, respectively.
Predictive performance of regression models was evaluated using the ten-fold external cross-validation (10-CV) procedure where the entire dataset was divided into ten non-overlapping pairs of the training and test sets of compounds. The models were obtained on the training set followed by their validation on the corresponding test set. The parameters optimization was performed using the tuning set (subset of the training set). The determination coefficient R^2 and root mean square error (*RMSE*) were used to evaluate the ability of the models to quantitatively predict the target property value. At the end of the procedure, all of the compounds in the initial data set are evaluated. The initial pool of descriptors both for the complete and the reduced data description was reshuffled N times (10 time for DGP and 100 times for SVM and PBP) and the calculations were repeated. The resulting performance metrics coefficients were averaged. The parameters for performance evaluation are evaluated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_{(pred,i)} - Y_{(exp,i)})^2}{\sum_{i=1}^{N}(Y_{(exp,i)} - \bar{Y}_{exp})^2} \tag{12}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{1}{N}(Y_{(pred,i)} - Y_{(exp,i)})^2} \tag{13}$$

Here, $Y_{(exp,i)}$ and $Y_{(pred,i)}$ are, respectively, experimental and predicted values of the modeling property, N is the number of data points, $\bar{Y}_{exp}$ is the average value of the experimental property.
In this study we use the p-value as the criteria of the model performance. We suggest it as the additional criteria to characterize the possible effects related with the different complexity of the data description.
The performed analysis on the continuity of the changes in the target characteristics as a function of the similarity defined by the descriptors was performed for both levels of complexity in data description using the special index of Structure-Activity Landscape Analysis (SALi) adopted from the practice used in the literature in the fields of cheminformatics and artificial intelligence in medicinal chemistry [123].

2.8.2. Configuration of the models

The following parameters were set for **Deep Autoencoding Gaussian Mixture Model (DAGMM)**: *(i)* for the complete data description defined by 45 descriptors in the model the number of epochs for training was limited by 500, the patience is 5, learning rate milestone = 50, batch size = 32, latent dim = 1, number of Gaussians in mixture model = 4, lambda energy = 0.1, lambda covariance = 0.005, *(ii)* for the reduced data description defined by 15 descriptors in the model the number of epochs for training was limited by 500, the patience is 5, learning rate milestone = 50, batch size = 12, latent dim = 1, number of Gaussians in mixture model = 12, lambda energy = 0.1, lambda covariance = 0.005.

Support Vector Machines performance was optimized in the grid search by varying three parameters as follows within the given range: $C = 2^{-5}, 2^{-3}...2^{15}$, $\varepsilon = 0.0001, 0.001 \ldots 10$ and $\gamma = 2^{-15}, 2^{-13}...2^{3}$.

Gaussian Processes were used with the following parameters: RBF kernel, back constraint = False, number of inducing points = 20, maximum number of iterations = 1200 and 1800 (for fixed and unfixed noise variance, respectively).

3. Results and analysis

3.1. Identification of different types of "outliers"/"novelty" in the synthesis in relation with the influence of the complexity of the data description

The impact of the complexity of the data description is one of the key problems being discussed in the context of novelty/outlier detection methods. The known discussions addressed both the general methodological issues as well as the shortcomings of the certain types of the methods. Additionally, considering the complexity of data description one should address the practical aspects of using experimental data: extending the data description by introducing more factors, which affect the target functional characteristics of interest, increases the necessity to involve the special techniques for data imputation or to reduce the size of the experimental data since some parameters can be not described in the literature. The question may arise how to assess the performance of the data description sufficient for the formulated problem and if one should always expect the need to find the certain compromise between the uncertainty in the data description and the capacities of the methods used for solving the problem. In this study, we consider these questions for one practical example of the problem of identification of different types of "outliers"/"novelty" and relying on the results obtained for one particular data set describing the data with a specific pool of the descriptors. Therefore, the results are assumed to be not of general character in nature.

In Figure 7 (left) the correlation matrix for two studied descriptors' pools is shown: *(i)* data description used in our previous study of modeling the functional ion transport characteristics of the garnet-type solid electrolytes while augmented with one type of descriptors has been introduced in this study [121], *(ii)* the reduced data description using the descriptors selected from the complete set according to the analysis of the contribution based on the Shapley value analysis performed in our previous study (Figure 8), however, preserving the information on the synthesis conditions and augmented with one type of descriptors introduced in this study. This correlation matrix allows to visually assess the types of descriptors where the sharp discontinuity in its values are observed. This also allows to assess the collinearity of these introduced parameters. From the Figure 7 (left) one may infer on the presence of the outliers in the data with respect to the synthesis conditions. It also allows to pre-assess the diversity in the presented compositions of the compounds and in the synthesis details/conditions. This correlation matrix is shown in line with the representation of the landscape of the changes in Li-ion conductivity value as a function of the similarity among the individual compounds based on Euclidean distance and defined by the chosen set of the descriptors. For this purpose, the corresponding SALi index relating the difference in the property value to the distance between the compounds is visualized. One can see that the reduced representation of the data is characterized by the increased number of the so called "cliffs" in the landscape. Our results obtained for the regression and, in some extent, for binary classification problems allows us to conclude on the demonstrated usefulness of such an analysis being performed to preserve the Occam's razor principle. It is assumed that this illustrative representation by using the heatmap plots may be helpful to support the decision on the perspectives of the methods to reach the desired performance with the involved data. However, this is more reasoned to assume the analysis of the inherent structure of the data description as more universal and feasible approach.

Below the data complexity is discussed from the point of view *(i)* of the relationship with the performance of the regression models as well as *(i)* of the dimensionality reduction problem supported by the the energy/distance evaluation provided as a landscape of the reduced data representation.

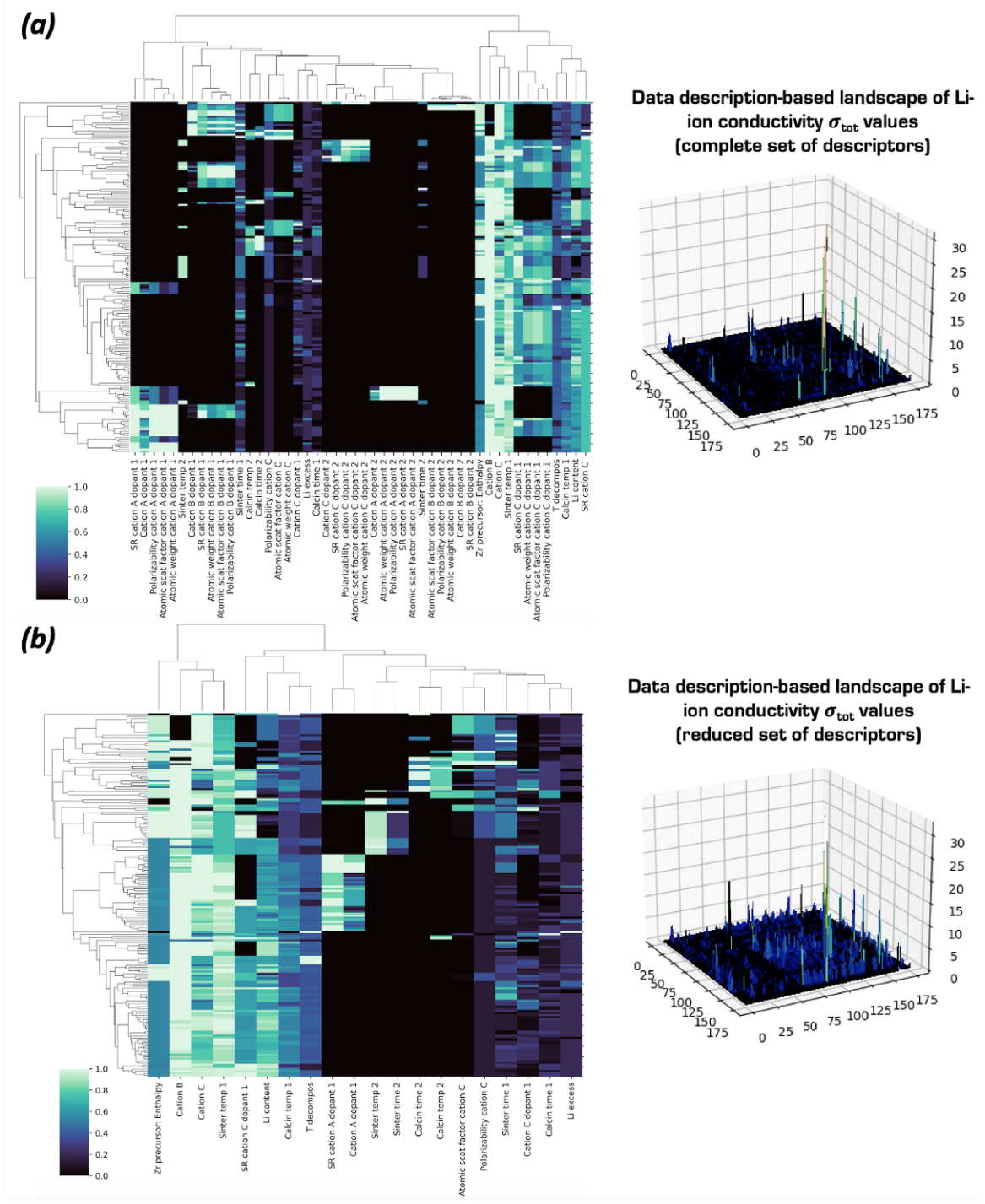

Figure 7: Data description as a clustermap with information on the variation in the descriptors' values for: *(i)* complete pool of parameters *(ii)* reduced pool of descriptors preserving the synthesis-related parameters.

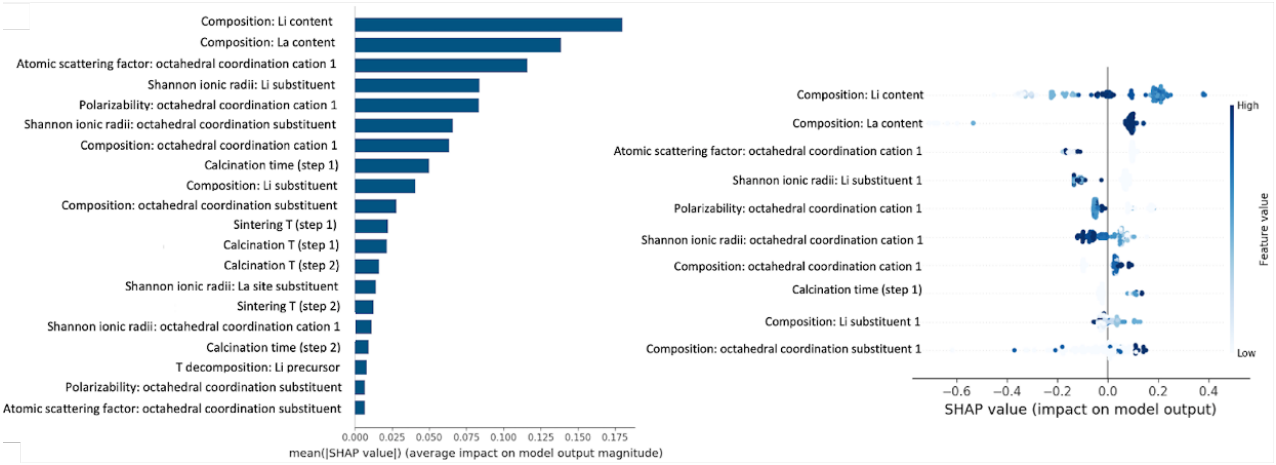

Figure 8: Shapley value based model explainability. The average impact on the model prediction is accompanied by the distribution of Shapley values for the considered data: (a) for σ_{tot} conductivity values. Figure reprinted with permission from [121]. Copyright of Elsevier

3.2. Quantitative results of outliers identification using the Area Under Curve (AUC) values

The overall performance of novelty/outlier detection of all types of outliers was characterized using neural networks with deep autoencoder architecture combined with the Gaussian mixture model to describe the data of varying complexity. The analysis was performed for both types of the data description, complete and reduced, to compare the efficiency of the data description in line with the efficiency of the novelty detection approaches. The decision on the success of the chosen DAGGM approach in the detection of the novelty/anomalies in the experimental data is performed based on the obtained values of Receiver Operating Characteristic with the evaluated values of the Area Under Curve (AUC). Using ROC-AUC statistics provides with 0.71 – 0.72 of the Area-Under-Curve values. However, one should take into account that this insignificant difference can be explained by re-defined list of the compounds that are considered as the "novelty"/"outliers" as it was not correct to use the same characterization for the reduced data description. The unchanged classification to outliers/nominal data provides with significantly worse performance that may be considered as insufficient (0.58 in AUC values). Using the provided data description DAGGM demonstrated the performance that does not affected by the dimensionality of the data. From the analysis of the results one may conclude on the sensitivity of DAGGM to the local changes in the data description. Even the change in one position in the descriptor fingerprint (indeed, we are referring to the significant change in its value) impacts the value of the energy function of DAGGM. It is worth to note that all the “outlier” compounds related to those as the result of using the rare precursors in synthesis were successfully identified. The unexpected adverse observation is that the two similar compounds can be characterized by the extremely different values of energy and, therefore, related to different classes (the nominal data or the outliers) of data as a result of the chosen batch parameter of neural networks. One of the source of the errors are considering the compounds obtained by distinct from the solid state synthesis routes as the outliers. This includes the compounds synthesized by sol-gel, co-precipitation and Pechini methods. The compounds obtained by these methods differ at least in their morphology and for sol-gel synthesis one may infer on the general trend of forming LLZO at much faster rates. Thus in the most of cases the samples obtained by sol-gel techniques are identified correctly. Also the difference can be concerned to the different defect states as of the surface as well as of the bulk structure.

The aim of this study was to demonstrate the perspectives of using the novelty detection techniques for identifying the unconventional synthesis routes. Using the available experimental data on ion transport characteristics in solid electrolytes of garnet-type we in addition have deliberately introduced the information on the compounds obtained by using unconventional for the most of data the cation C precursors, the available data also contained the examples that can be classified as those differ from the most of the data. For the former case, the chosen method of the novelty detection DAGGM successfully identified all the examples. The performance of the models can be improved by the enrichment of the target data especially for co-doped samples of different compositions.

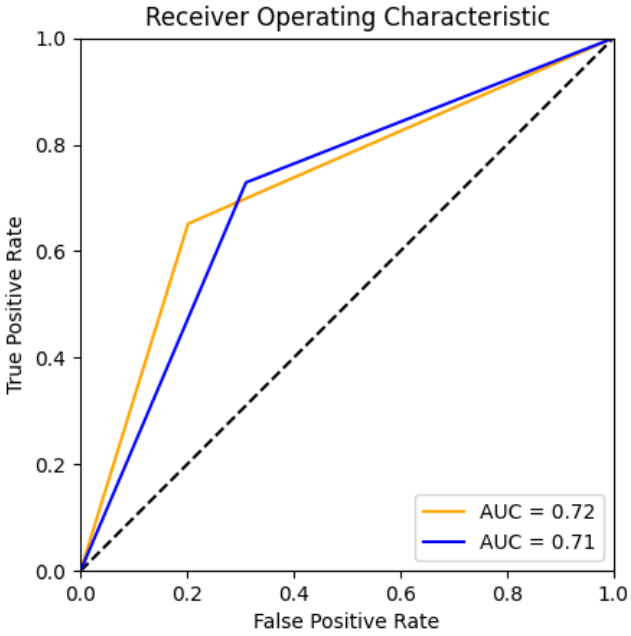

Figure 9: Receiver Operating Curve in line with corresponding Area Under Curve value characterizing the performance of the obtained novelty detection models describing the composition and the synthesis routes of the garnet-structured solid state electrolytes.

3.3. Relationship between the outliers detection and the performance of composition-synthesis conditions-structure-property relationship modeling using the coefficient of determination R^2, root mean squared error RMSE and p-value statistics as the selected criteria

Three machine learning methods for regression problem have been involved as those providing with minimum value of the average prediction error in our previous studies [121, 122]: Support Vector Machines (SVM), Deep Gaussian Processes (DGP) and Probabilistic Backpropagation Neural Networks (PBP). We assume the overall "normality" of the data and thus the interval of the analysis is limited by 75 percents of the data to be considered as the nominal. The another assumption made in this study is in the a-priori better quality of the complete data description as the lower number of the "cliffs" were shown in Figure 7 well coincides with the predictive performance of models. The outlierness of the data was introduced by means of the ranging the data ascending according to the values of the energy function evaluated by DAGGM. Hereafter, the obtained results for the data of varying complexity are analyzed in two aspects: *(i)* the overall performance as a function of the ML method as well as the complexity and *(ii)* the character of behavior of the system. The statistical parameters (R^2 and *RMSE*) of regression models for Li-ion total conductivity σ_{tot} value are represented in Figure 10 with the corresponding standard errors as a function of the part of the data (Q) involved in modeling (the corresponding values are given in Table 1). In all the cases, the statistical parameters of the models for the complete data description and the entire data involved in modeling is higher with the best performance of the models obtained using PBP (R^2 = 0.75±0.02 and *RMSE* = 0.379 ± 0.015). For the data of reduced complexity the best performance was reached using DGP method and is characterized by R^2 = 0.69±0.003 and *RMSE* = 0.417 ± 0.003. DGP has shown a smoother trend in the changes of the statistical characteristics of models depending on the data used in the simulation than its counterparts. However, the opposite trend is observed for the complete and the reduced data description: the predictive error decreases as a function of the data amount for the complete data description while the insignificant increase is observed for the reduced data. It is worth to note that the std. errors for the averaged results are of order of magnitude lower for the DGP. Both SVM and PBP methods have revealed the lower boundary that may be recommended for the separation the areas of lower density of the probability distribution. For the complete data description, this boundary corresponds to 85 percents of data considered as nominal/"normal" and thus used in modeling. For the reduced data description, PBP shows the enhanced predictive performance after

10 percents of data are removed, afterwards the predictive error changes insignificantly. This difference demonstrates the importance of the analysis of the landscape of the target characteristics as a function of the similarity between the compounds defined by the data description. SVM approach has demonstrated the similar trend for the data of both complete and reduced description and this trend corresponds well with that of PBP used for the data described by the complete set of the descriptors. The performance and the trend of the obtained results allows one to conclude on the tendency of the methods to preserve all the data as the normal to provede with reliable predictions. Due to introducing the feature weighting in the algorithm DGP is highly stable to the data variations and thus can be considered as a good choice for the data comprising the different types of the "outliers". Figure 11 shows the p-value evaluated using the experimental values and the results of the predictions of the regression models for total Li-ion conductivity σ_{tot} value as a function of the part of the data involved in modeling. From this figure one can consider 85 percents of the data as a recommended threshold for studied experimental data and the machine learning methods used for modeling.

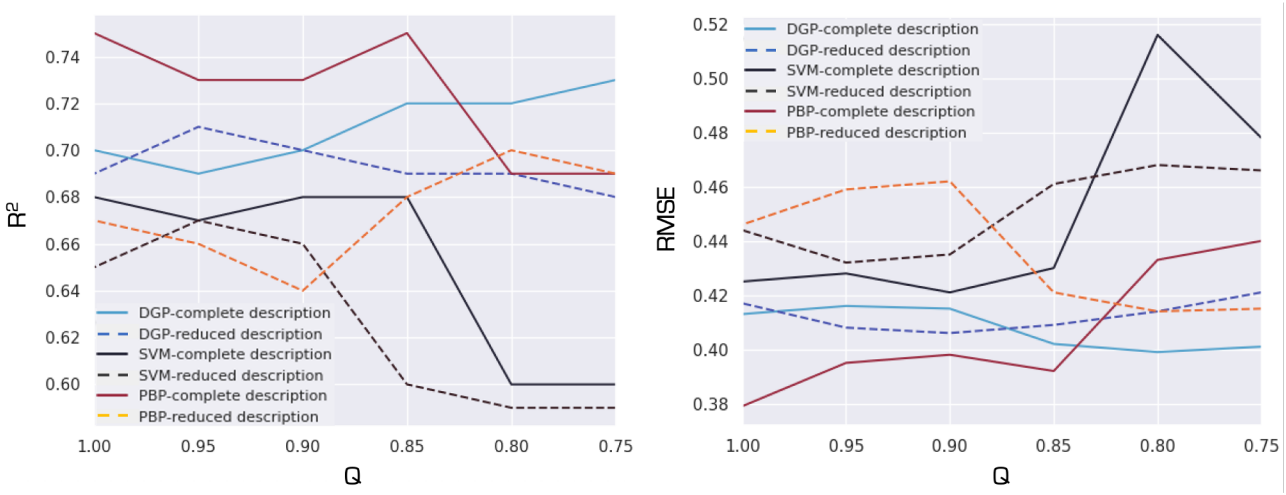

Figure 10: Statistical parameters (R^2 and *RMSE*) of regression models for Li-ion total conductivity σ_{tot} value as a function of the part of the data involved in modeling. The values were evaluated using 10-fold cross-validation for complete and reduced data description and three machine learning methods: Deep Gaussian Processes (DGP), Support Vector Machines (SVM) and Probabilistic Bayesian Backpropagation Neural Networks (PBP) (here Q is the part of the data used in modeling).

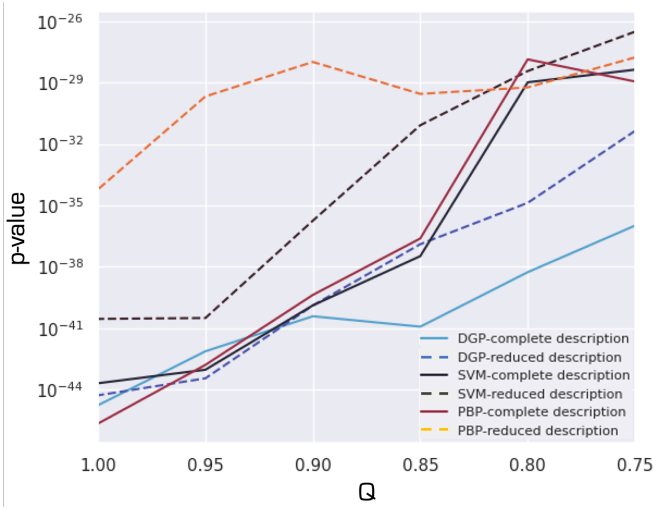

Figure 11: p-value evaluated using the experimental values and the results of the predictions of the regression models for total Li-ion conductivity σ_{tot} value as a function of the part of the data involved in modeling. The values were evaluated using 10-fold cross-validation for complete and reduced data description and three machine learning methods: Deep Gaussian Processes (DGP), Support Vector Machines (SVM) and Probabilistic Bayesian Backpropagation Neural Networks (PBP) (here Q is the part of the data used in modeling).

Table 1: Statistical parameters for modeling the Li-ion conductivity σ_{tot}: coefficient of determination R^2 and root mean square error (*RMSE*) with averaged (the number of models are shown in the brackets) std. errors for ten-fold cross-validation for complete and reduced data description (here Q is the part of the data used in modeling).

Q	N	d	**DGP** Avg. R^2 with std. error and p-value	**DGP** Avg. *RMSE* and std. error	**SVM** Avg. R^2 with std. error and p-value	**SVM** Avg. *RMSE* and std. error	**PBP** Avg. R^2 with std. error and p-value	**PBP** Avg. *RMSE* and std. error
1.0	172	48	0.70±0.003, 1.72e-45 (10)	0.413±0.003 (10)	0.68±0.03, 2.05e-44 (100)	0.425±0.020 (100)	0.75±0.02, 2.19e-46 (100)	0.379±0.015(100)
0.95	163	48	0.69±0.003, 2.85e-29 (10)	0.416±0.004 (10)	0.67±0.03, 9.39e-44 (100)	0.428±0.023 (100)	0.73±0.02, 1.62e-43 (100)	0.395±0.015(100)
0.9	154	48	0.70±0.003, 3.85e-41 (10)	0.415±0.003 (10)	0.68±0.02, 1.31e-40 (100)	0.421±0.015 (100)	0.73±0.02, 4.24e-40 (100)	0.398±0.017(100)
0.85	145	48	0.72±0.002, 1.19e-41 (10)	0.402±0.002 (10)	0.68±0.07, 3.34e-38 (100)	0.430±0.039 (100)	0.75±0.02, 2.43e-37 (100)	0.392±0.017(100)
0.8	136	48	0.72±0.002, 5.49e-39 (10)	0.399±0.004 (10)	0.60±0.15, 1.04e-29 (100)	0.516±0.071 (100)	0.69±0.03, 1.38e-31 (100)	0.433±0.020(100)
0.75	127	48	0.73±0.004, 1.04e-36 (10)	0.401±0.005 (10)	0.60±0.09, 4.32e-29 (100)	0.478±0.050 (100)	0.69±0.02, 1.13e-29 (100)	0.440±0.019(100)
1.0	172	21	0.69±0.003, 5.28e-45 (10)	0.417±0.003 (10)	0.65±0.03, 2.86e-41 (100)	0.444±0.017 (100)	0.67±0.03, 6.08e-35 (100)	0.446±0.021(100)
0.95	163	21	0.71±0.004, 3.57e-44 (10)	0.408±0.005 (10)	0.67±0.03, 3.16e-41 (100)	0.432±0.019 (100)	0.66±0.03, 2.12e-30 (100)	0.459±0.027(100)
0.9	154	21	0.70±0.003, 1.29e-40 (10)	0.406±0.003 (10)	0.66±0.03, 1.80e-36 (100)	0.435±0.016 (100)	0.64±0.04, 1.03e-28 (100)	0.462±0.030(100)
0.85	145	21	0.69±0.006, 1.28e-37 (10)	0.409±0.006 (10)	0.60±0.05, 8.29e-32 (100)	0.461±0.027 (100)	0.68±0.03, 2.82e-30 (100)	0.421±0.024(100)
0.8	136	21	0.69±0.003, 1.34e-35 (10)	0.414±0.002 (10)	0.59±0.04, 3.69e-29 (100)	0.468±0.022 (100)	0.70±0.03, 5.87e-30 (100)	0.414±0.020(100)
0.75	127	21	0.68±0.003, 4.44e-32 (10)	0.421±0.003 (10)	0.59±0.05, 3.11e-27 (100)	0.466±0.026 (100)	0.69±0.03, 1.72e-28 (100)	0.415±0.022(100)

3.4. Distinguishing outliers by means of dimensionality reduction techniques using associated information of novelty detection

In this study, the idea one of the first times mentioned in [110] is demonstrated in a way very close to the original formulation by means of using the dimensionality reduction technique, Stochastic Triplet Embedding. The dimensionality reduction effectively condenses the multiparametric information distilled from the available experimental data to, in our case, the two-dimensional representation (more generally, to the dimensionality reduced comparing to the original one) to provide with the so called "mapping" to show the investigated compounds in a correspondence to their relative distribution according to the similarity defined by the descriptors (the set of the input parameters) introducing the landscape of the energy value as a criteria of the outlierness/normality of the data. This type of representation may be associated with the related additional information by using the color indication for the points corresponding to the individual compounds. This color indication of the points may relate to, for example, the functional characteristics under the investigation. In Figure 12 the distributed data are attributed with Li-ion conductivity σ_{tot} values (in log of σ_{tot} / S·cm^{-1}) and activation energy values (E_a, eV). The background in Figure 12 may serve as the indication of the samples were synthesized using the non-conventional routes or, of distinct from the most represented compositions. The lighter background colour corresponds to the higher probability for the compound/compounds to be the outlier. One can see that the compounds with the lower conductivity values are grouped in the map in one constrained region, the larger area of the map is intended to be used for the screening of new compounds. Thus, the areas with already assessed values of energy function using novelty detection techniques are preferential. The enrichment of the analyzed data here is beneficial. The comparative analysis of these two maps can provide with the compounds, which have these two target characteristics in disagreement.

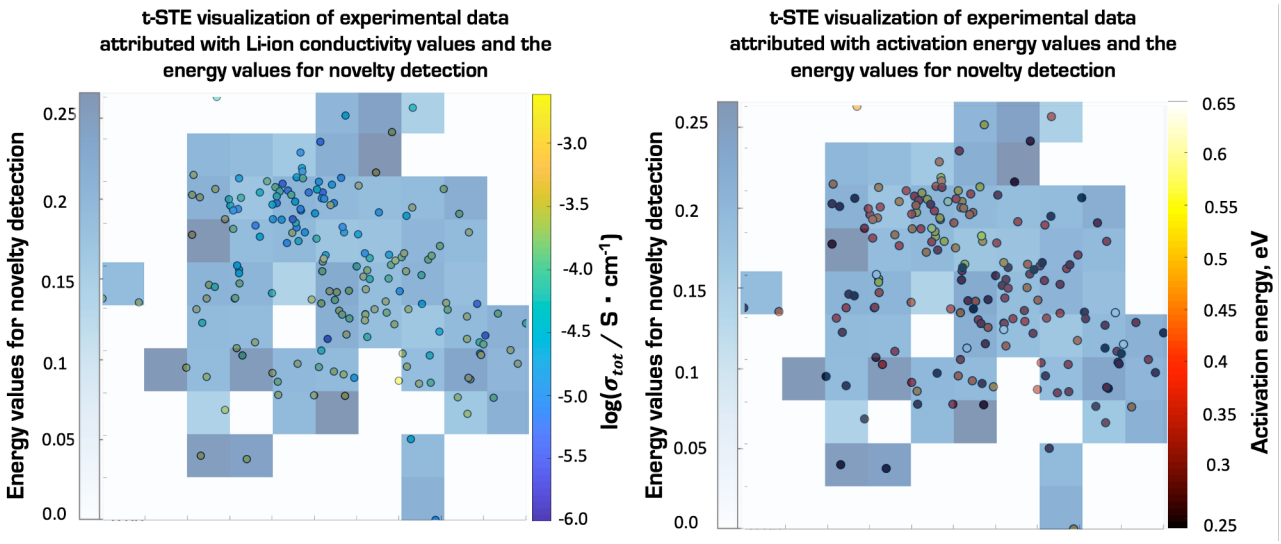

Figure 12: *t*-STE visualization of the experimental data attributed with Li-ion transport characteristics and the energy values for the novelty detection: each point on the map is the individual compound (the dimensionality reduction was performed from the dimensionality of 48).

4. Conclusions

In this study, we propose using the novelty detection approaches aimed at identifying the novelty in the experimental data for garnet-structured solid electrolytes for the analysis of the non-standard synthesis details as the object demonstrating one of the practical applications of high demand. Particular attention in the analysis of novelty detection in the design of synthesis was given to the choice of C cation precursors. The thermodynamic data such as the heat of formation from the pure oxides as well as the results of drop solution calorimetry for simple oxides were involved as the descriptors of the studied systems. The overall performance of novelty/outlier detection of all types

of outliers was characterized for the data description of varying complexity using ROC-AUC statistics and is assessed as 0.71 – 0.72 of the Area-Under-Curve values. It was found that all the "outlier" compounds related to those as the result of using the rare precursors in synthesis were successfully identified. The complementary regression analysis was performed for studying the relationship between the data diversity and the complexity. The conclusions on the requirements for the initial characterization of the data to find the optimized degree of the complexity of the data description were made based on the results obtained.

5. Acknowledgements

Authors acknowledge Prof. Alexandre Varnek for fruitful discussions of our work. Matplotlib[124] and Seaborn[125] packages were used for the modeling and the presentation of the results. Free access to the "Database of properties of chemical elements"[116] is highly appreciated.

6. Data availability

The data were submitted as Supporting Information.

References

[1] K. Rajan (Ed.), Data-driven Discovery for Accelerated Experimentation and Application, Elsevier, 2013.
[2] K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, U. Celano, Machine learning in nanoscience: Big data at small scales, Nano Letters 20 (1) (2020) 2–10, pMID: 31804080. arXiv:https://doi.org/10.1021/acs.nanolett.9b04090, doi:10.1021/acs.nanolett.9b04090.
URL https://doi.org/10.1021/acs.nanolett.9b04090
[3] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S. P. Ong, A critical review of machine learning of energy materials, Advanced Energy Materials 10 (8) (2020) 1903242. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201903242, doi:https://doi.org/10.1002/aenm.201903242.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201903242
[4] T. Gao, W. Lu, Machine learning toward advanced energy storage devices and systems, iScience 24 (1). doi:10.1016/j.isci.2020.101936.
URL https://doi.org/10.1016/j.isci.2020.101936
[5] D. Lee, D. You, D. Lee, X. Li, S. Kim, Machine-learning-guided prediction models of critical temperature of cuprates, J. Phys. Chem. Lett. 12 (26) (2021) 6211–6217. doi:10.1021/acs.jpclett.1c01442.
URL https://doi.org/10.1021/acs.jpclett.1c01442
[6] A. G. Patel, L. Johnson, R. Arroyave, J. L. Lutkenhaus, Design of multifunctional supercapacitor electrodes using an informatics approach, Mol. Syst. Des. Eng. 4 (2019) 654–663. doi:10.1039/C8ME00060C.
URL http://dx.doi.org/10.1039/C8ME00060C
[7] Z.-Y. Zhang, X. Liu, L. Shen, L. Chen, W.-H. Fang, Machine learning with multilevel descriptors for screening of inorganic nonlinear optical crystals, The Journal of Physical Chemistry C 125 (45) (2021) 25175–25188. arXiv:https://doi.org/10.1021/acs.jpcc.1c06049, doi:10.1021/acs.jpcc.1c06049.
URL https://doi.org/10.1021/acs.jpcc.1c06049
[8] N. Kireeva, V. P. Solov'ev, Machine learning analysis of microwave dielectric properties for seven structure types: The role of the processing and composition, Journal of Physics and Chemistry of Solids 156 (2021) 110178. doi:https://doi.org/10.1016/j.jpcs.2021.110178.
URL https://www.sciencedirect.com/science/article/pii/S0022369721002444
[9] H. Takeda, H. Fukuda, K. Nakano, S. Hashimura, N. Tanibata, M. Nakayama, Y. Ono, T. Natori, Process optimisation for nasicon-type solid electrolyte synthesis using a combination of experiments and bayesian optimisation, Mater. Adv. 3 (2022) 8141–8148. doi:10.1039/D2MA00731B.
URL http://dx.doi.org/10.1039/D2MA00731B
[10] S. R. Young, A. Maksov, M. Ziatdinov, Y. Cao, M. Burch, J. Balachandran, L. Li, S. Somnath, R. M. Patton, S. V. Kalinin, R. K. Vasudevan, Data mining for better material synthesis: The case of pulsed laser deposition of complex oxides, Journal of Applied Physics 123 (11) (2018) 115303. arXiv:https://doi.org/10.1063/1.5009942, doi:10.1063/1.5009942.
URL https://doi.org/10.1063/1.5009942
[11] L. Velasco, J. S. Castillo, M. V. Kante, J. J. Olaya, P. Friederich, H. Hahn, Phase–property diagrams for multicomponent oxide systems toward materials libraries, Advanced Materials 33 (43) (2021) 2102301. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202102301, doi:https://doi.org/10.1002/adma.202102301.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202102301

[12] D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C. J. Brabec, B. Maruyama, K. A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation, Nature Reviews Materials 3 (5) (2018) 5–20. doi:10.1038/s41578-018-0005-z.
URL https://doi.org/10.1038/s41578-018-0005-z
[13] E. Borvick, A. Y. Anderson, H.-N. Barad, M. Priel, D. A. Keller, A. Ginsburg, K. J. Rietwyk, S. Meir, A. Zaban, Process-function data mining for the discovery of solid-state iron-oxide pv, ACS Combinatorial Science 19 (12) (2017) 755–762, pMID: 29120164. arXiv:https://doi.org/10.1021/acscombsci.7b00121, doi:10.1021/acscombsci.7b00121.
URL https://doi.org/10.1021/acscombsci.7b00121
[14] E. A. Olivetti, J. M. Cole, E. Kim, O. Kononova, G. Ceder, T. Y.-J. Han, A. M. Hiszpanski, Data-driven materials research enabled by natural language processing and information extraction, Applied Physics Reviews 7 (4) (2020) 041317. arXiv:https://doi.org/10.1063/5.0021106, doi:10.1063/5.0021106.
URL https://doi.org/10.1063/5.0021106
[15] E. Kim, K. Huang, S. Jegelka, E. Olivetti, Virtual screening of inorganic materials synthesis parameters with deep learning, npj Computational Materials 3 (1) (2017) 53. doi:10.1038/s41524-017-0055-6.
URL https://doi.org/10.1038/s41524-017-0055-6
[16] E. Kim, Z. Jensen, A. van Grootel, K. Huang, M. Staib, S. Mysore, H.-S. Chang, E. Strubell, A. McCallum, S. Jegelka, E. Olivetti, Inorganic materials synthesis planning with literature-trained neural networks, Journal of Chemical Information and Modeling 60 (3) (2020) 1194–1201, pMID: 31909619. arXiv:https://doi.org/10.1021/acs.jcim.9b00995, doi:10.1021/acs.jcim.9b00995.
URL https://doi.org/10.1021/acs.jcim.9b00995
[17] C. Karpovich, Z. Jensen, V. Venugopal, E. Olivetti, Inorganic synthesis reaction condition prediction with generative machine learning (2021). arXiv:2112.09612.
[18] S. A. Malik, R. E. A. Goodall, A. A. Lee, Predicting the outcomes of material syntheses with deep learning, Chemistry of Materials 33 (2) (2021) 616–624. arXiv:https://doi.org/10.1021/acs.chemmater.0c03885, doi:10.1021/acs.chemmater.0c03885.
URL https://doi.org/10.1021/acs.chemmater.0c03885
[19] O. Wodo, S. Broderick, K. Rajan, Microstructural informatics for accelerating the discovery of processing-microstructure-property relationships, MRS Bulletin 41 (8) (2016) 603–609. doi:10.1557/mrs.2016.161.
URL https://www.cambridge.org/core/article/microstructural-informatics-for-accelerating-the-discovery-of-processing-mic
28DA8E333BF33856C9647FCD0F0983D1
[20] L. Xu, N. Hoffman, Z. Wang, H. Xu, Harnessing structural stochasticity in the computational discovery and design of microstructures, Materials Design 223 (2022) 111223. doi:https://doi.org/10.1016/j.matdes.2022.111223.
URL https://www.sciencedirect.com/science/article/pii/S0264127522008450
[21] L. C. O. Tiong, J. Kim, S. S. Han, D. Kim, Identification of crystal symmetry from noisy diffraction patterns by a shape analysis and deep learning, npj Computational Materials 6 (1) (2020) 196. doi:10.1038/s41524-020-00466-5.
URL https://doi.org/10.1038/s41524-020-00466-5
[22] S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, L. Hung, Random forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships, npj Computational Materials 6 (1) (2020) 109. doi:10.1038/s41524-020-00376-6.
URL https://doi.org/10.1038/s41524-020-00376-6
[23] A. Zunger, Inverse design in search of materials with target functionalities, Nature Reviews Chemistry 2 (4) (2018) 0121. doi:10.1038/s41570-018-0121.
URL https://doi.org/10.1038/s41570-018-0121
[24] J. Wang, Y. Wang, Y. Chen, Inverse design of materials by machine learning, Materials 15 (5). doi:10.3390/ma15051811.
URL https://www.mdpi.com/1996-1944/15/5/1811
[25] J. B. Goodenough, P. Singh, Review—solid electrolytes in rechargeable electrochemical cells, Journal of The Electrochemical Society 162 (14) (2015) A2387. doi:10.1149/2.0021514jes.
URL https://dx.doi.org/10.1149/2.0021514jes
[26] J. C. Bachman, S. Muy, A. Grimaud, H.-H. Chang, N. Pour, S. F. Lux, O. Paschos, F. Maglia, S. Lupart, P. Lamp, L. Giordano, Y. Shao-Horn, Inorganic solid-state electrolytes for lithium batteries: Mechanisms and properties governing ion conduction, Chemical Reviews 116 (1) (2016) 140–162. doi:10.1021/acs.chemrev.5b00563.
URL https://doi.org/10.1021/acs.chemrev.5b00563
[27] B. V. Lotsch, J. Maier, Relevance of solid electrolytes for lithium-based batteries: A realistic view, Journal of Electroceramics 38 (2) (2017) 128–141. doi:10.1007/s10832-017-0091-0.
URL https://doi.org/10.1007/s10832-017-0091-0
[28] Z. Zou, Y. Li, Z. Lu, D. Wang, Y. Cui, B. Guo, Y. Li, X. Liang, J. Feng, H. Li, C.-W. Nan, M. Armand, L. Chen, K. Xu, S. Shi, Mobile ions in composite solids, Chemical Reviews 120 (9) (2020) 4169–4221, pMID: 32267697. arXiv:https://doi.org/10.1021/acs.chemrev.9b00760, doi:10.1021/acs.chemrev.9b00760.
URL https://doi.org/10.1021/acs.chemrev.9b00760
[29] J. W. Fergus, Ceramic and polymeric solid electrolytes for lithium-ion batteries, Journal of Power Sources 195 (15) (2010) 4554–4569. doi:https://doi.org/10.1016/j.jpowsour.2010.01.076.
URL https://www.sciencedirect.com/science/article/pii/S037877531000234X
[30] J. F. M. Oudenhoven, L. Baggetto, P. H. L. Notten, All-solid-state lithium-ion microbatteries: A review of various three-dimensional concepts, Advanced Energy Materials 1 (1) (2011) 10–33. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.201000002, doi:https://doi.org/10.1002/aenm.201000002.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201000002
[31] T. S. Hernandez, M. Alshurafa, M. T. Strand, A. L. Yeang, M. G. Danner, C. J. Barile, M. D. McGehee, Electrolyte for improved durability

of dynamic windows based on reversible metal electrodeposition, Joule 4 (7) (2020) 1501–1513. doi:https://doi.org/10.1016/j.joule.2020.05.008.
URL https://www.sciencedirect.com/science/article/pii/S2542435120301914

[32] I. Peters, C. Breyer, S. Jaffer, S. Kurtz, T. Reindl, R. Sinton, M. Vetter, The role of batteries in meeting the pv terawatt challenge, Joule 5 (6) (2021) 1353–1370. doi:https://doi.org/10.1016/j.joule.2021.03.023.
URL https://www.sciencedirect.com/science/article/pii/S254243512100146X

[33] K. J. Huang, G. Ceder, E. A. Olivetti, Manufacturing scalability implications of materials choice in inorganic solid-state batteries, Joule 5 (3) (2021) 564–580. doi:https://doi.org/10.1016/j.joule.2020.12.001.
URL https://www.sciencedirect.com/science/article/pii/S2542435120305699

[34] R. Murugan, V. Thangadurai, W. Weppner, Fast lithium ion conduction in garnet-type li7la3zr2o12, Angewandte Chemie International Edition 46 (41) (2007) 7778–7781. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.200701144, doi:https://doi.org/10.1002/anie.200701144.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200701144

[35] R. Murugan, V. Thangadurai, W. Weppner, Effect of lithium ion content on the lithium ion conductivity of the garnet-like structure li5+xbala2ta2o11.5+0.5x (x = 0–2), Applied Physics A 91 (4) (2008) 615–620. doi:10.1007/s00339-008-4494-2.
URL https://doi.org/10.1007/s00339-008-4494-2

[36] V. Thangadurai, H. Kaack, W. J. F. Weppner, Novel fast lithium ion conduction in garnet-type li5la3m2o12 (m = nb, ta), Journal of the American Ceramic Society 86 (3) (2003) 437–440. arXiv:https://ceramics.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1151-2916.2003.tb03318.x, doi:https://doi.org/10.1111/j.1151-2916.2003.tb03318.x.
URL https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/j.1151-2916.2003.tb03318.x

[37] V. Thangadurai, D. Pinzaru, S. Narayanan, A. K. Baral, Fast solid-state li ion conducting garnet-type structure metal oxides for energy storage, The Journal of Physical Chemistry Letters 6 (2) (2015) 292–299. doi:10.1021/jz501828v.
URL https://doi.org/10.1021/jz501828v

[38] A. Miura, C. J. Bartel, Y. Goto, Y. Mizuguchi, C. Moriyoshi, Y. Kuroiwa, Y. Wang, T. Yaguchi, M. Shirai, M. Nagao, N. C. Rosero-Navarro, K. Tadanaga, G. Ceder, W. Sun, Observing and modeling the sequential pairwise reactions that drive solid-state ceramic synthesis, Advanced Materials 33 (24) (2021) 2100312. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202100312, doi:https://doi.org/10.1002/adma.202100312.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202100312

[39] J. R. Chamorro, T. M. McQueen, Progress toward solid state synthesis by design, Accounts of Chemical Research 51 (11) (2018) 2918–2925. arXiv:https://doi.org/10.1021/acs.accounts.8b00382, doi:10.1021/acs.accounts.8b00382.
URL https://doi.org/10.1021/acs.accounts.8b00382

[40] Z. Cai, Y.-Q. Zhang, Z. Lun, B. Ouyang, L. C. Gallington, Y. Sun, H.-M. Hau, Y. Chen, M. C. Scott, G. Ceder, Thermodynamically driven synthetic optimization for cation-disordered rock salt cathodes, Advanced Energy Materials 12 (21) (2022) 2103923. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/aenm.202103923, doi:https://doi.org/10.1002/aenm.202103923.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.202103923

[41] H. C. Kolb, M. G. Finn, K. B. Sharpless, Click chemistry: Diverse chemical function from a few good reactions, Angewandte Chemie International Edition.

[42] J. Maier, Physical Chemistry of Ionic Materials: Ions and Electrons in Solids, 1st Edition, Wiley, 2004.

[43] N. Hamao, K. Hamamoto, N. Taguchi, S. Tanaka, J. Akimoto, Synthesis and crystal structure of fluorite-type la2.4zr1.2ta0.4o7: A precursor oxide for low temperature formation of garnet-type li6.5la3zr1.5ta0.5o12, Solid State Ionics 357 (2020) 115460. doi:https://doi.org/10.1016/j.ssi.2020.115460.
URL https://www.sciencedirect.com/science/article/pii/S0167273820305142

[44] T. Gensch, G. dos Passos Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner-D'Addario, M. S. Sigman, A. Aspuru-Guzik, A comprehensive discovery platform for organophosphorus ligands for catalysis, Journal of the American Chemical Society 144 (3) (2022) 1205–1217, pMID: 35020383. arXiv:https://doi.org/10.1021/jacs.1c09718, doi:10.1021/jacs.1c09718.
URL https://doi.org/10.1021/jacs.1c09718

[45] T. A. Lee, A. Navrotsky, I. Molodetsky, Enthalpy of formation of cubic yttria-stabilized zirconia, Journal of Materials Research 18 (4) (2003) 908–918. doi:10.1557/JMR.2003.0125.
URL https://doi.org/10.1557/JMR.2003.0125

[46] A. Bogicevic, C. Wolverton, G. M. Crosbie, E. B. Stechel, Defect ordering in aliovalently doped cubic zirconia from first principles, PRB 64 (1) (2001) 014106. doi:10.1103/PhysRevB.64.014106.
URL https://link.aps.org/doi/10.1103/PhysRevB.64.014106

[47] J. Maier, Ionic conduction in space charge regions, Progress in Solid State Chemistry 23 (3) (1995) 171–263. doi:https://doi.org/10.1016/0079-6786(95)00004-E.
URL https://www.sciencedirect.com/science/article/pii/007967869500004E

[48] R. H. Castro, Interfacial energies in nanocrystalline complex oxides, Current Opinion in Solid State and Materials Science 25 (3) (2021) 100911. doi:https://doi.org/10.1016/j.cossms.2021.100911.
URL https://www.sciencedirect.com/science/article/pii/S1359028621000140

[49] S. V. Ushakov, A. Navrotsky, Direct measurements of water adsorption enthalpy on hafnia and zirconia, Appl. Phys. Lett. 87 (16) (2005) 164103. doi:10.1063/1.2108113.
URL https://doi.org/10.1063/1.2108113

[50] B. T. B. J. Schrier J, Norquist A, In pursuit of the exceptional: Research directions for machine learning in chemical and materials science, Chemrxiv.

[51] Z. Zhang, J. Brgoch, Treating superhard materials as anomalies, Journal of the American Chemical Society 144 (39) (2022) 18075–18080, pMID: 36136594. arXiv:https://doi.org/10.1021/jacs.2c07957, doi:10.1021/jacs.2c07957.

URL https://doi.org/10.1021/jacs.2c07957
[52] R. Wagner, G. J. Redhammer, D. Rettenwander, A. Senyshyn, W. Schmidt, M. Wilkening, G. Amthauer, Crystal structure of garnet-related li-ion conductor li7-3xgaxla3zr2o12: Fast li-ion conduction caused by a different cubic modification?, Chemistry of Materials 28 (6) (2016) 1861–1871, pMID: 27019548. arXiv:https://doi.org/10.1021/acs.chemmater.6b00038, doi:10.1021/acs.chemmater.6b00038.
URL https://doi.org/10.1021/acs.chemmater.6b00038
[53] D. Rettenwander, J. Langer, W. Schmidt, C. Arrer, K. J. Harris, V. Terskikh, G. R. Goward, M. Wilkening, G. Amthauer, Site occupation of ga and al in stabilized cubic li7–3(x+y)gaxalyla3zr2o12 garnets as deduced from 27al and 71ga mas nmr at ultrahigh magnetic fields, Chemistry of Materials 27 (8) (2015) 3135–3142. arXiv:https://doi.org/10.1021/acs.chemmater.5b00684, doi:10.1021/acs.chemmater.5b00684.
URL https://doi.org/10.1021/acs.chemmater.5b00684
[54] D. Rettenwander, G. Redhammer, F. Preishuber-Pflügl, L. Cheng, L. Miara, R. Wagner, A. Welzl, E. Suard, M. M. Doeff, M. Wilkening, J. Fleig, G. Amthauer, Structural and electrochemical consequences of al and ga cosubstitution in li7la3zr2o12 solid electrolytes, Chemistry of Materials 28 (7) (2016) 2384–2392, pMID: 27110064. arXiv:https://doi.org/10.1021/acs.chemmater.6b00579, doi:10.1021/acs.chemmater.6b00579.
URL https://doi.org/10.1021/acs.chemmater.6b00579
[55] R. Wagner, G. J. Redhammer, D. Rettenwander, G. Tippelt, A. Welzl, S. Taibl, J. Fleig, A. Franz, W. Lottermoser, G. Amthauer, Fast li-ion-conducting garnet-related li7-3xfexla3zr2o12 with uncommon i43d structure, Chemistry of Materials 28 (16) (2016) 5943–5951, pMID: 27570369. arXiv:https://doi.org/10.1021/acs.chemmater.6b02516, doi:10.1021/acs.chemmater.6b02516.
URL https://doi.org/10.1021/acs.chemmater.6b02516
[56] N. Bernstein, M. D. Johannes, K. Hoang, Origin of the structural phase transition in $li_7la_3zr_2o_{12}$, Phys. Rev. Lett. 109 (2012) 205702. doi:10.1103/PhysRevLett.109.205702.
URL https://link.aps.org/doi/10.1103/PhysRevLett.109.205702
[57] S. Ohno, T. Bernges, J. Buchheim, M. Duchardt, A.-K. Hatz, M. Kraft, H. Kwak, A. Santhosha, Z. Liu, N. Minafra, F. Tsuji, A. Sakuda, R. Schlem, S. Xiong, Z. Zhenggang, P. Adelhelm, H. Chen, A. Hayashi, Y. S. Jung, W. Zeier, How certain are the reported ionic conductivities of thiophosphate-based solid electrolytes? an interlaboratory study, ACS Energy Letters 5 (2020) 910–915. doi:10.1021/acsenergylett.9b02764.
[58] R. Inada, K. Kusakabe, T. Tanaka, S. kudo, Y. Sakurai, Synthesis and properties of al-free li7 - xla3zr 2 - xtaxo12 garnet related oxides, Solid State Ionics 262 (2014) 568–572. doi:10.1016/j.ssi.2013.09.008.
[59] S. M. Alizadeh, I. Moghim, M. Golmohammad, Synthesis and characterization of highly conductive ga/y co-doped llzo by facile combustion sol-gel method, Solid State Ionics 397 (2023) 116260. doi:https://doi.org/10.1016/j.ssi.2023.116260.
URL https://www.sciencedirect.com/science/article/pii/S0167273823001182
[60] J. Kosir, S. Mousavihashemi, B. P. Wilson, E.-L. Rautama, T. Kallio, Comparative analysis on the thermal, structural, and electrochemical properties of al-doped li7la3zr2o12 solid electrolytes through solid state and sol-gel routes, Solid State Ionics 380 (2022) 115943. doi:https://doi.org/10.1016/j.ssi.2022.115943.
URL https://www.sciencedirect.com/science/article/pii/S0167273822000923
[61] O. Brylev, O. Shlyakhtin, T. Kulova, A. Skundin, Y. Tretyakov, Influence of chemical prehistory on the phase formation and electrochemical performance of licoo2 materials, Solid State Ionics 156 (3) (2003) 291–299. doi:https://doi.org/10.1016/S0167-2738(02)00686-0.
URL https://www.sciencedirect.com/science/article/pii/S0167273802006860
[62] C. Korte, B. Franz, Reaction kinetics in the system y2o3/al2o3 – use of an external electric field to control the product phase formation in a system forming multiple product phases, Solid State Ionics 383 (2022) 115978. doi:https://doi.org/10.1016/j.ssi.2022.115978.
URL https://www.sciencedirect.com/science/article/pii/S0167273822001278
[63] A. Sazvar, H. Sarpoolaky, M. Golmohammad, The effects of electric field on physical properties of llzo made by flash sintering method, Solid State Ionics 386 (2022) 116054. doi:https://doi.org/10.1016/j.ssi.2022.116054.
URL https://www.sciencedirect.com/science/article/pii/S016727382200203X
[64] A. M. Belenguer, A. A. L. Michalchuk, G. I. Lampronti, J. K. M. Sanders, Understanding the unexpected effect of frequency on the kinetics of a covalent reaction under ball-milling conditions., Beilstein journal of organic chemistry 15 (2019) 1226–1235.
[65] T. Scheiber, M. Gombotz, K. Hogrefe, H. M. R. Wilkening, Fluoride ion dynamics in nanocrystalline a-pbf2: On the tremendous impact of structural disorder on f- anion hopping in poor ion conductors, Solid State Ionics 387 (2022) 116077. doi:https://doi.org/10.1016/j.ssi.2022.116077.
URL https://www.sciencedirect.com/science/article/pii/S0167273822002260
[66] A. A. Shindrov, Increasing sinterability and ionic conductivity of na3zr2si2po12 ceramics by high energy ball-milling, Solid State Ionics 391 (2023) 116139. doi:https://doi.org/10.1016/j.ssi.2022.116139.
URL https://www.sciencedirect.com/science/article/pii/S0167273822002880
[67] J. Maier, Pushing nanoionics to the limits: Charge carrier chemistry in extremely small systems, Chemistry of Materials 26 (1) (2014) 348–360. arXiv:https://doi.org/10.1021/cm4021657, doi:10.1021/cm4021657.
URL https://doi.org/10.1021/cm4021657
[68] Y. Zhu, M. Chon, C. V. Thompson, J. L. M. Rupp, Time-temperature-transformation (ttt) diagram of battery-grade li-garnet electrolytes for low-temperature sustainable synthesis, Angewandte Chemie International Edition 62 (45) (2023) e202304581. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.202304581, doi:https://doi.org/10.1002/anie.202304581.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202304581
[69] T. Kimura, Y. Yamada, K. Yamamoto, T. Matsuda, H. Nomura, T. Hirayama, Rapid low-temperature synthesis of tetragonal single-phase li7la3zr2o12, Journal of the American Ceramic Society 100 (4) (2017) 1313–1319. arXiv:https://ceramics.onlinelibrary.wiley.com/doi/pdf/10.1111/jace.14633, doi:https://doi.org/10.1111/jace.14633.

URL https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/jace.14633
[70] E. Anderson, A. Jonderian, R. Z. Khaliullin, E. McCalla, Combinatorial study of the li-la-zr-o system, Solid State Ionics 388 (2022) 116087. doi:https://doi.org/10.1016/j.ssi.2022.116087.
URL https://www.sciencedirect.com/science/article/pii/S0167273822002363
[71] F. Rahmawati, B. Musyarofah, K. D. Nugrahaningtyas, A. Prasetyo, V. Suendo, H. Haeruddin, M. F. Handaka, H. Nilasari, H. Nursukatmo, A different zirconia precursor for li7la3zr2o12 synthesis, Journal of Materials Research and Technology 15 (2021) 2725–2734. doi:https://doi.org/10.1016/j.jmrt.2021.09.064.
URL https://www.sciencedirect.com/science/article/pii/S2238785421010450
[72] A. J. G. Ellison, A. Navrotsky, Enthalpy of formation of zircon, Journal of the American Ceramic Society 75 (6) (1992) 1430–1433. arXiv:https://ceramics.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1151-2916.1992.tb04205.x, doi:https://doi.org/10.1111/j.1151-2916.1992.tb04205.x.
URL https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/j.1151-2916.1992.tb04205.x
[73] F. González, R. Khadka, R. López-Juárez, J. Collins, B. Di Bartolo, Emission of white-light in cubic y4zr3o12:yb3+ induced by a continuous infrared laser, Journal of Luminescence 198 (2018) 320–326. doi:https://doi.org/10.1016/j.jlumin.2018.02.053.
URL https://www.sciencedirect.com/science/article/pii/S0022231317320537
[74] N. Hamao, J. Akimoto, A novel synthetic route of garnet-type li6.5la3zr1.5ta0.5o12 using pyrochlore-type la2zr2o7 and weberite-type la3tao7 as starting materials, Journal of the Ceramic Society of Japan 127 (2019) 374–377. doi:10.2109/jcersj2.19014.
[75] C. Deviannapoorani, S. Ramakumar, N. Janani, R. Murugan, Synthesis of lithium garnets from la2zr2o7 pyrochlore, Solid State Ionics 283 (2015) 123–130. doi:https://doi.org/10.1016/j.ssi.2015.10.006.
URL https://www.sciencedirect.com/science/article/pii/S0167273815003847
[76] G. T. Hitz, E. D. Wachsman, V. Thangadurai, Highly li-stuffed garnet-type li7+xla3zr2-xyxo12, Journal of The Electrochemical Society 160 (8) (2013) A1248. doi:10.1149/2.088308jes.
URL https://dx.doi.org/10.1149/2.088308jes
[77] H. Huo, C. J. Bartel, T. He, A. Trewartha, A. Dunn, B. Ouyang, A. Jain, G. Ceder, Machine-learning rationalization and prediction of solid-state synthesis conditions, Chemistry of Materials 34 (16) (2022) 7323–7336. arXiv:https://doi.org/10.1021/acs.chemmater.2c01293, doi:10.1021/acs.chemmater.2c01293.
URL https://doi.org/10.1021/acs.chemmater.2c01293
[78] K. Cruse, A. Trewartha, S. Lee, Z. Wang, H. Huo, T. He, O. Kononova, A. Jain, G. Ceder, Text-mined dataset of gold nanoparticle synthesis procedures, morphologies, and size entities, Scientific Data 9 (1) (2022) 234. doi:10.1038/s41597-022-01321-6.
URL https://doi.org/10.1038/s41597-022-01321-6
[79] T. He, H. Huo, C. J. Bartel, Z. Wang, K. Cruse, G. Ceder, Precursor recommendation for inorganic synthesis by machine learning materials similarity from scientific literature, Science Advances 9 (23) (2023) eadg8180. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.adg8180, doi:10.1126/sciadv.adg8180.
URL https://www.science.org/doi/abs/10.1126/sciadv.adg8180
[80] F. Lalère, V. Seznec, M. Courty, J. N. Chotard, C. Masquelier, Coupled x-ray diffraction and electrochemical studies of the mixed ti/v-containing nasicon: Na2tiv(po4)3, J. Mater. Chem. A 6 (2018) 6654–6659. doi:10.1039/C7TA10689K.
URL http://dx.doi.org/10.1039/C7TA10689K
[81] N. J. Szymanski, P. Nevatia, C. J. Bartel, Y. Zeng, G. Ceder, Autonomous and dynamic precursor selection for solid-state materials synthesis, Nature Communications 14 (1) (2023) 6956. doi:10.1038/s41467-023-42329-9.
URL https://doi.org/10.1038/s41467-023-42329-9
[82] N. J. Szymanski, C. J. Bartel, Y. Zeng, Q. Tu, G. Ceder, Probabilistic deep learning approach to automate the interpretation of multi-phase diffraction spectra, Chemistry of Materials 33 (11) (2021) 4204–4215. arXiv:https://doi.org/10.1021/acs.chemmater.1c01071, doi:10.1021/acs.chemmater.1c01071.
URL https://doi.org/10.1021/acs.chemmater.1c01071
[83] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, in: S. Solla, T. Leen, K. Müller (Eds.), Advances in Neural Information Processing Systems, Vol. 12, MIT Press, 1999.
URL https://proceedings.neurips.cc/paper_files/paper/1999/file/8725fb777f25776ffa9076e44fcfd776-Paper.pdf
[84] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[85] E. Nguyen, M. Seo, S. J. Oh, A bayesian perspective on training data attribution (05 2023).
[86] G. Ritter, M. T. Gallegos, Outliers in statistical pattern recognition and an application to automatic chromosome classification, Pattern Recognition Letters 18 (6) (1997) 525–539. doi:https://doi.org/10.1016/S0167-8655(97)00049-4.
URL https://www.sciencedirect.com/science/article/pii/S0167865597000494
[87] D. M. Tax, R. P. Duin, Support vector domain description, Pattern Recognition Letters 20 (11) (1999) 1191–1199. doi:https://doi.org/10.1016/S0167-8655(99)00087-2.
URL https://www.sciencedirect.com/science/article/pii/S0167865599000872
[88] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, Vol. 80 of Proceedings of Machine Learning Research, PMLR, 2018, pp. 4393–4402.
URL https://proceedings.mlr.press/v80/ruff18a.html
[89] R. Chalapathy, A. Krishna Menon, S. Chawla, Anomaly Detection using One-Class Neural Networks, arXiv e-prints (2018) arXiv:1802.06360arXiv:1802.06360, doi:10.48550/arXiv.1802.06360.
[90] S. Ben-David, M. Lindenbaum, Learning distributions by their density levels: A paradigm for learning without a teacher, Journal of Computer and System Sciences 55 (1) (1997) 171–182. doi:https://doi.org/10.1006/jcss.1997.1507.
URL https://www.sciencedirect.com/science/article/pii/S0022000097915075

[91] C. Zhou, R. C. Paffenroth, Anomaly detection with robust deep autoencoders, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 665–674. doi:10.1145/3097983.3098052.
URL https://doi.org/10.1145/3097983.3098052
[92] S. Zhai, Y. Cheng, W. Lu, Z. Zhang, Deep structured energy based models for anomaly detection.
[93] D. P. Kingma, M. Welling, An introduction to variational autoencoders, Foundations and Trends® in Machine Learning 12 (4) (2019) 307–392. doi:10.1561/2200000056.
URL https://doi.org/10.1561%2F2200000056
[94] S. Rifai, P. Vincent, X. Muller, X. Glorot, Y. Bengio, Contractive auto-encoders: Explicit invariance during feature extraction, in: Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, Omnipress, Madison, WI, USA, 2011, p. 833–840.
[95] S. Pidhorskyi, R. Almohsen, D. A. Adjeroh, G. Doretto, Generative probabilistic novelty detection with adversarial autoencoders (2018). arXiv:1807.02588.
[96] M. Sakurada, T. Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, MLSDA'14, Association for Computing Machinery, New York, NY, USA, 2014, p. 4–11. doi:10.1145/2689746.2689747.
URL https://doi.org/10.1145/2689746.2689747
[97] M. Mirza, S. Osindero, Conditional generative adversarial nets (2014). arXiv:1411.1784.
[98] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. ki Cho, H. Chen, Deep autoencoding gaussian mixture model for unsupervised anomaly detection, in: International Conference on Learning Representations, 2018.
[99] I. Golan, R. El-Yaniv, Deep anomaly detection using geometric transformations, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018.
URL https://proceedings.neurips.cc/paper_files/paper/2018/file/5e62d03aec0d17facfc5355dd90d441c-Paper.pdf
[100] E. Parzen, On Estimation of a Probability Density Function and Mode, The Annals of Mathematical Statistics 33 (3) (1962) 1065 – 1076. doi:10.1214/aoms/1177704472.
URL https://doi.org/10.1214/aoms/1177704472
[101] J. Kim, C. D. Scott, Robust kernel density estimation, J. Mach. Learn. Res. 13 (1) (2012) 2529–2565.
[102] E. J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J. ACM 58 (3). doi:10.1145/1970392.1970395.
URL https://doi.org/10.1145/1970392.1970395
[103] T. Defard, A. Setkov, A. Loesch, R. Audigier, Padim: A patch distribution modeling framework for anomaly detection and localization, in: A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, R. Vezzani (Eds.), Pattern Recognition. ICPR International Workshops and Challenges, Springer International Publishing, Cham, 2021, pp. 475–489.
[104] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413–422. doi:10.1109/ICDM.2008.17.
[105] M. Breunig, P. Kröger, R. Ng, J. Sander, Lof: Identifying density-based local outliers., Vol. 29, 2000, pp. 93–104. doi:10.1145/342009.335388.
[106] T. Kohonen, Self-Organizing Maps, Springer Berlin, Heidelberg, 2012.
[107] A. Smola, R. Williamson, S. Mika, B. Schölkopf, Regularized principal manifolds, Vol. 1, 1999, pp. 67–67. doi:10.1007/3-540-49097-3_17.
[108] C. Bishop, M. Svensen, C. Williams, Gtm: The generative topographic mapping, Neural Computation 10 (1997) 215–234. doi:10.1162/089976698300017953.
[109] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, K.-R. Müller, A unifying review of deep and shallow anomaly detection, Proceedings of the IEEE 109 (5) (2021) 756–795. doi:10.1109/JPROC.2021.3052449.
[110] P. Baldi, K. Hornik, Neural networks and principal component analysis: Learning from examples without local minima, Neural Networks 2 (1) (1989) 53–58. doi:https://doi.org/10.1016/0893-6080(89)90014-2.
URL https://www.sciencedirect.com/science/article/pii/0893608089900142
[111] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507. arXiv:https://www.science.org/doi/pdf/10.1126/science.1127647, doi:10.1126/science.1127647.
URL https://www.science.org/doi/abs/10.1126/science.1127647
[112] L. van der Maaten, K. Weinberger, Stochastic triplet embedding, in: 2012 IEEE International Workshop on Machine Learning for Signal Processing, 2012, pp. 1–6. doi:10.1109/MLSP.2012.6349720.
[113] C.-C. Chang, C.-J. Lin, Libsvm : a library for support vector machines, LIBSVM : a library for support vector machines Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.
[114] J. M. Hernández-Lobato, R. P. Adams, Probabilistic backpropagation for scalable learning of bayesian neural networks (2015). arXiv:1502.05336.
[115] A. Damianou, N. D. Lawrence, Deep Gaussian processes, in: C. M. Carvalho, P. Ravikumar (Eds.), Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, Vol. 31 of Proceedings of Machine Learning Research, PMLR, Scottsdale, Arizona, USA, 2013, pp. 207–215.
URL https://proceedings.mlr.press/v31/damianou13a.html
[116] Database of properties of chemical elements.
URL http://phases.imet-db.ru/elements/main.aspx
[117] P. Villars, J. DAAMS, Y. SHIKATA, K. Rajan, S. Iwata, A new approach to describe elemental-property parameters.
[118] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, Nature 533 (2016) 73 EP –.

URL https://doi.org/10.1038/nature17439
[119] N. Kireeva, V. S. Pervov, Materials space of solid-state electrolytes: unraveling chemical composition–structure–ionic conductivity relationships in garnet-type metal oxides using cheminformatics virtual screening approaches, Phys. Chem. Chem. Phys. 19 (2017) 20904–20918. doi:10.1039/C7CP00518K.
URL http://dx.doi.org/10.1039/C7CP00518K
[120] N. Kireeva, V. S. Pervov, Materials informatics screening of li-rich layered oxide cathode materials with enhanced characteristics using synthesis data, Batteries & Supercaps 3 (5) (2020) 427–438. arXiv:https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/batt.201900186, doi:https://doi.org/10.1002/batt.201900186.
URL https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/batt.201900186
[121] N. V. Kireeva, A. Y. Tsivadze, V. S. Pervov, Modeling ionic conductivity and activation energy in garnet-structured solid electrolytes: The role of composition, grain boundaries and processing, Solid State Ionics 399 (2023) 116293. doi:https://doi.org/10.1016/j.ssi.2023.116293.
URL https://www.sciencedirect.com/science/article/pii/S0167273823001510
[122] N. Kireeva, A. Y. Tsivadze, V. S. Pervov, Predicting ionic conductivity in thin films of garnet electrolytes using machine learning, Batteries 9 (9). doi:10.3390/batteries9090430.
URL https://www.mdpi.com/2313-0105/9/9/430
[123] R. Guha, J. H. Van Drie, Structure-activity landscape index: Identifying and quantifying activity cliffs, Journal of Chemical Information and Modeling 48 (3) (2008) 646–658, pMID: 18303878. arXiv:https://doi.org/10.1021/ci7004093, doi:10.1021/ci7004093.
URL https://doi.org/10.1021/ci7004093
[124] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering 9 (3) (2007) 90–95. doi:10.1109/MCSE.2007.55.
[125] M. Waskom, O. Botvinnik, D. Kane, P. Hobson, S. Lukauskas, D. Gemperline, et al., mwaskom/seaborn: v0.8.1 (September 2017). Zenodo. (2017).