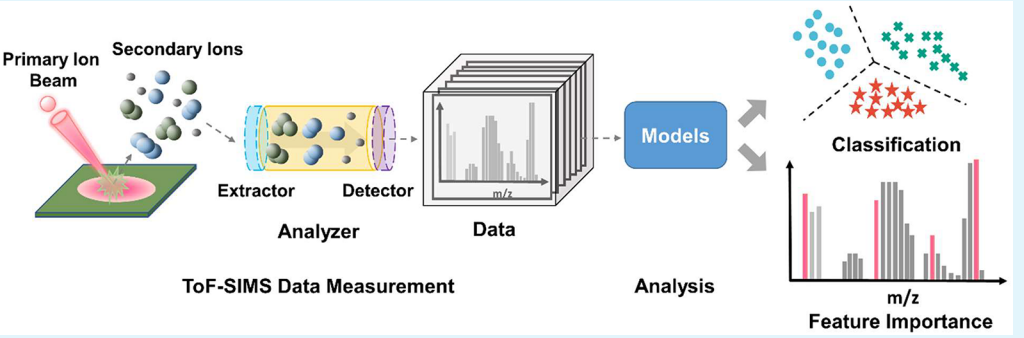
ACS APPLIED MATERIALS & INTERFACES

www.acsami.org
Research Article

Identification of Lithium Compounds on Surfaces of Lithium Metal Anode with Machine-Learning-Assisted Analysis of ToF-SIMS Spectra

Yinghan Zhao,[‖] Svenja-K. Otto,[‖] Teo Lombardo, Anja Henss, Arnd Koeppe,[*] Michael Selzer, Jürgen Janek, and Britta Nestler

Cite This: *ACS Appl. Mater. Interfaces* 2023, 15, 50469−50478



ABSTRACT: Detailed knowledge about contamination and passivation compounds on the surface of lithium metal anodes (LMAs) is essential to enable their use in all-solid-state batteries (ASSBs). Time-of-flight secondary ion mass spectrometry (ToF-SIMS), a highly surface-sensitive technique, can be used to reliably characterize the surface status of LMAs. However, as ToF-SIMS data are usually highly complex, manual data analysis can be difficult and time-consuming. In this study, machine learning techniques, especially logistic regression (LR), are used to identify the characteristic secondary ions of 5 different pure lithium compounds. Furthermore, these models are applied to the mixture and LMA samples to enable identification of their compositions based on the measured ToF-SIMS spectra. This machine-learning-based analysis approach shows good performance in identifying characteristic ions of the analyzed compounds that fit well with their chemical nature. Moreover, satisfying accuracy in identifying the compositions of unseen new samples is achieved. In addition, the scope and limitations of such a strategy in practical applications are discussed. This work presents a robust analytical method that can assist researchers in simplifying the analysis of the studied lithium compound samples, offering the potential for broader applications in other material systems.

KEYWORDS: *lithium metal anode, all-solid-state battery, ToF-SIMS, machine-learning-assisted analysis, data science*

1. INTRODUCTION

Lithium metal anodes (LMAs) are of great interest for future battery applications due to their high theoretical specific capacity and low redox potential. Particularly, LMAs have great potential for their use in all-solid-state batteries (ASSBs), because they are currently the sole option that could lead to ASSBs with a higher energy density than commercial lithium-ion batteries (LIBs) using liquid electrolytes and graphite anodes.[1] However, the application of LMAs still faces severe challenges, such as morphological instability and low Coulomb efficiency.[2,3] Different investigations carried out so far on this topic indicate that one critical aspect of LMAs is the degradation process at the surface. Hence, the precise knowledge about contaminations and passivation compounds on its surface is key to unlocking LMA-based ASSBs.[4−7]

In this context, we have previously reported how X-ray photoelectron spectroscopy (XPS) and time-of-flight secondary ion mass spectrometry (ToF-SIMS) can be used to reliably characterize lithium metal surfaces.[8] On the one hand, XPS allows one to obtain both quantitative information on the surface elements and compounds and their qualitative depth distributions (depth profiling). On the other hand, ToF-SIMS depth profiling can complement the XPS results with higher lateral resolution and quantitative depth information, even if

Received: July 4, 2023
Accepted: September 13, 2023
Published: October 18, 2023

ToF-SIMS results are not inherently compound-specific and the method is only semiquantitative due to matrix effects. Despite these limitations, ToF-SIMS has already been demonstrated to be a powerful analytical technique to probe LMA surfaces. In 2014, Karen et al. described how ToF-SIMS can help to identify reaction products of lithium−air battery electrodes by using pure lithium compounds as reference materials.[9] In particular, the authors used an argon gas cluster ion beam (GCIB) to stabilize the lithium compounds, which facilitated the identification of the characteristic secondary ions. Following a similar approach, we have previously identified characteristic secondary ions of different lithium compounds by comparing their signal intensity after cleaning through argon cluster ions sputtering.[8] Even if the peak interpretation is enhanced through this approach, it is still not enough to draw conclusions without any prior characterization, for instance, by XPS. To overcome this limitation, a more sophisticated data analysis method is needed, which is the main objective of the present work.

Manual analysis of ToF-SIMS spectra is particularly challenging and time-consuming due to the presence of several hundreds of peaks. Furthermore, important information may be overlooked or missed easily. Data science techniques, especially machine learning (ML) models, which possess the ability to explore the latent relationships behind data, are gaining more and more popularity in the analysis of ToF-SIMS data. Besides, a widely utilized approach to support ToF-SIMS data interpretation, the multivariate data analysis (MVA),[10−13] can indicate, for instance, the relationships between different peaks and their relative importance, as well as information on the peaks' variance. MVA also takes advantage of many unsupervised learning methods in the machine learning field, like principal component analysis (PCA),[14−16] multivariate curve resolution (MCR),[17] and non-negative matrix factorization (NMF).[18−23] For example, Heller et al. applied PCA and MCR to study the degradation products in lithium-ion batteries with only very limited prior knowledge,[24] while Schroder et al. used PCA to improve the understanding of ToF-SIMS depth profiles of LIB solid-electrolyte interphases (SEI),[25] and Higgins et al. developed a workflow with NMF to extract salient features of associated chemical changes at halide perovskite interfaces and to separate the light- and voltage-dependent dynamics.[26] Many other ML models, such as support vector machines, decision-trees-based models such as random forest model, hierarchical clustering analysis, neural networks and so on, have been successfully applied in the field of mass spectrometry for potential enhancement of data analysis and interpretation,[27−33] as well as many other battery related fields.[34] Among these algorithm architectures, logistic regression (LR) is a commonly used machine learning approach that has been effectively adopted in various research areas.[35] The main advantages of this approach are its easy implementation, efficient (lower computational cost) training, and low probability of overfitting.

In the presented work, the ToF-SIMS spectral data of various lithium metal anodes are studied by using the LR model. With the help of other machine learning techniques, the performance of the model is evaluated, and its behavior is explained from various aspects. To be specific, the LR model is first used to identify the characteristic ions for the pure compounds, and their corresponding chemical information is analyzed. Additionally, this model is further applied to mixtures of these compounds and real LMA samples for composition identification, which evaluates the extrapolative prediction ability of the models and the possible variety of scenarios that it can apply. Overall, the reported identification of specific ion species and the application to LMA samples demonstrate that the machine learning-assisted analysis methods have the potential to provide quick, automated, and accurate ToF-SIMS data analysis, which can help researchers to better distinguish the different components present in their sample(s) given little or no previous knowledge.

2. EXPERIMENTAL SECTION

2.1. Sample Preparation. 5 different lithium compounds, namely, Li_2CO_3, Li_2O, Li_3N, LiH, and LiOH, were pressed into pellets (3 t, 1.5 min) under Ar-atmosphere. Mixtures of Li_2CO_3+LiOH (marked as sample No. 1 in Table 2) and LiH +Li_2O (sample No. 2) were prepared by mixing equal weights of the selected lithium compounds in a vial by shaking and then following the same pressing. As reference LMA samples, one commercial lithium foil exposed to an N_2-plasma (marked as sample No. 3 in Table 2) and two untreated foils (samples No. 4 and 5, pieces of the same pure lithium metal foils, and they were stored in a glovebox for 6 months) were chosen. The LMA samples were cut into small pieces each with an area of about 1.1 cm^2. All samples were mounted on a LEICA sample holder by using electronically insulating tape. The subsequent transfer to the ToF-SIMS instrument was also done under Ar-atmosphere.

2.2. Data Acquisition and Management. The ToF-SIMS measurements were performed using a TOF.SIMS 5−100 instrument (ION-TOF GmbH, Muenster, Germany) equipped with a 25 kV Bicluster primary ion gun for analysis and with a 20 kV gas cluster ion beam (GCIB) for sputtering. For the pure lithium compounds, mixtures (samples No. 1 and 2), and N_2-plasma sample (sample No. 3), the corresponding surface regions were sputter-cleaned with (Ar_{1500}^+) using a fluence of 4×10^{15} ions/cm^2 (10 kV, 10 nA) before the analysis. For the lithium foil, different fluences (1 frame = 9×10^{13} ions/cm^2 for sample No. 4 or 150 s = 9.5×10^{15} ions/cm^2 for sample No. 5) were applied to reach certain specific sample regions. For analysis, areas of 100×100 μm^2 were measured in spectrometry mode, using Bi_3^+ (20 kV, 1 pA) as the primary ion by setting an ion dose density of 10^{12} ions/cm^2. Measurements were done in negative ion mode with a cycle time of 100 μs. A flood gun was used for charge compensation during the sputtering and analysis. For each material, 12 different spots were examined. Besides, the research data infrastructure Kadi4Mat was used to share and manage data for continuously developing the machine learning model.[36]

2.3. Data Preprocessing. The software SurfaceLab (version 7.1, IONTOF GmbH, Muenster, Germany) was used for the first processing of the ToF-SIMS data. All spectra were calibrated to the ion signals $^6Li^-$, Li^-, O^-, OH^- and O_2^-. Following that, an automated peak search was conducted up to mass-to-charge (m/z) = 120 a.u. and for peaks with a minimum of 100 counts, after which the borders of the found peaks were calibrated to uniform ranges for subsequent comparison. The peak area was exported for all selected peaks and spectra to be used for further analysis.

3. RESULTS AND DISCUSSION

The first aim of this study was to build machine learning models to identify the key secondary ions representing different lithium compounds based on their ToF-SIMS spectra. For this purpose, a ToF-SIMS data set was collected, which contained 60 spectra from 5 different pure lithium-containing materials (12 spectra each), namely, Li_2CO_3, Li_2O, Li_3N, LiH, and LiOH. These compounds are commonly expected to be present on LMAs. The second goal of the present work was to use these models to recognize the lithium compounds present on LMA surfaces based on their ToF-SIMS spectra. In this study, the logistic regression (LR) model was mainly used to

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

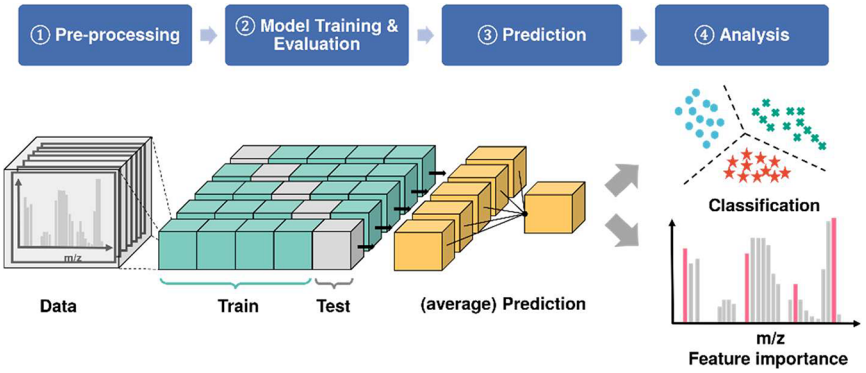


Figure 1. Schematic workflow adopted to build the LR model applied to the analysis of ToF-SIMS data, which consists of data preprocessing, model training, 5-fold cross-validation to evaluate its performance, application on unseen samples (predictions), and analysis of the classification results. More information about ToF-SIMS spectra, such as the importance of specific peaks or the composition of the mixture samples, can be obtained by a further analysis of the model results.

reach the aforementioned goals, whose accuracies were tested by analyzing two mixture samples (i.e., containing more than one substance, as expected by real-world LMAs) and three real LMAs samples. For each mixture and LMA sample, 12 spectra were collected. In addition to the LR model, principal component analysis (PCA), as one of the most widely used MVA techniques that can represent the basis for many other methods,[37] was also used to facilitate the analysis of data, and its results were compared with LR model in Supporting Information Section S1.

3.1. Logistic Regression Workflow. A logistic regression (LR) model was developed to distinguish between the different lithium compounds and determine the characteristic m/z peaks. Logistic regression (or logit regression), despite its name, is a kind of classification model in statistical learning. In the binary classification case, it is used to model the probability (p) that given data belong to (or do not belong) to a certain class. In the simplest case, it uses the following equation to model the probability:

$$p(\boldsymbol{x}) = \frac{1}{1 + e^{-(\boldsymbol{w}\cdot\boldsymbol{x}+b)}}$$

where $p(\boldsymbol{x})$ is the probability of being part of a certain class, e is the mathematical constant Euler's number, $\boldsymbol{x} \in R^n$ are the input variables, $\boldsymbol{w} \in R^n$ are the weights (coefficients), and the bias $b \in R$ is the intercept. As can be seen from this equation, the model attempts to learn a mathematical function that predicts by evaluating a combination of features $\boldsymbol{x}$. The parameter $\boldsymbol{w}$ can intuitively be expressed as the relative importance of each feature in the prediction, and it is therefore used to quantify the feature importance and infer the most important m/z peaks for the different compounds. Furthermore, since the model can predict the probability value of belonging to each class of compounds, such a prediction can be used to assist in recognizing the compositions of the mixture and LMA samples. A more detailed description of the LR model is given in Supporting Information Section S2. In this work, the Python machine learning library scikit-learn was used as the implementation of the LR model.[38]

The schematic diagram of the workflow followed to build the LR model is illustrated in Figure 1, and it mainly consists of four processing steps (from left to right in the figure):

1. *Preprocessing*: The raw ToF-SIMS data were denoised and scaled to have comparable values between different variables, which is one of the key steps preceding any ML-based approach. In the field of ToF-SIMS analysis, there are many commonly used preprocessing methods. However, no general method was found yet to be able to fit all the situations.[39−42] In this study, we found that the min−max scaling method performed well on our data set (see Supporting Information Section S3 for detailed discussion). For the min−max scaling, each intensity value of spectra is first subtracted from the minimum value of the corresponding m/z peak and is then divided by its intensity range (difference between the original maximum and minimum intensity of that m/z peak). In this way, the intensity in each m/z peak is scaled to a range between 0 and 1. This kind of scaling approach was found to be beneficial for the identification of the specific secondary ions, allowing the distinction between the samples.
2. *Model training and evaluation*: The LR model was trained to classify the 5 pure lithium compounds based on their measured ToF-SIMS spectra. An accurate model should not only be able to classify samples it has been trained on, but it should also perform well on new samples that it has not seen before. The K-fold cross-validation (CV) strategy was used to take full advantage of the limited amount of ToF-SIMS data and to assess the predictive performance of the LR model in a fair manner.[43] In K-fold cross-validation, the original samples are randomly partitioned into K complementary subsets, where only one single subset is kept as the test set for evaluating the performance of the model, while the remaining $K - 1$ are used to train the ML model. This process is iterated K times, where in each iteration a different subset is used for the model evaluation and the rest for the training. The prediction accuracy is quantified as the number of correct predictions divided by the total one, and through the F-score. The K evaluation results are then averaged to assess the overall performance of the model. Besides, in many cases, an optional extra unseen test data set can also be utilized to further evaluate the model performance. In this study, there was no such kind of additional test set due to the limited number of measured samples. A 5-fold cross-validation scheme was adopted, and in each cross-validation step, the min−max scaling was performed. The scaling parameters are defined as a function of the training set, and these same parameters

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

Table 1. Results of Identified Specific Peaks Presented in Different Font Styles for Better Illustration[a]

	LiH		Li_3N		Li_2CO_3		LiOH		Li_2O	
no.	m/z	Ion	m/z	Ion	m/z	Ion	m/z	Ion	m/z	Ion
1	9.03	**LiH_2^-**	27.00	**CHN^-**	67.00	**$LiCO_3^-$**	65.04	**$Li_2O_3H_3^-$**	48.00	C_4^-
2	8.03	**$^6LiH_2^-$**	26.00	**CN^-**	66.00	**$^6LiCO_3^-$**	64.04	**$Li_2O_3H_2^-$**	49.01	C_4H^-
3	17.05	**$Li_2H_3^-$**	21.02	**LiN^-**	43.99	**CO_2^-**	49.05	**$Li_2O_2H_3^-$**	49.05	$Li_2O_2H_3^-$
4	25.08	**$Li_3H_4^-$**	22.02	**$LiNH^-$**	59.99	**CO_3^-**	63.03	**$Li_2O_3H^-$**	8.02	LiH^-
5	16.05	**$Li_2H_2^-$**	15.01	**NH^-**	12.00	**C^-**	41.02	**$LiO_2H_2^-$**	31.99	O_2^-
6	24.09	**$^6LiLi_2H_4^-$**	23.04	**$LiNH_2^-$**	68.01	**$LiCO_3H^-$**	34.00	**$H_2O_2^-$**	65.04	$Li_2O_3H_3^-$
7	16.05	**$^6LiLiH_3^-$**	30.00	NO^-	54.00	**$^6LiO_3^-$**	33.00	**O_2H^-**	15.04	Li_2H^-
8	15.04	**Li_2H^-**	34.03	**$CHNLi^-$**	51.01	$H_3O_3^-$	56.01	**LiO_3H^-**	56.01	LiO_3H^-
9	33.10	**$Li_4H_5^-$**	16.02	**NH_2^-**	60.99	**CO_3H^-**	40.01	**LiO_2H^-**	60.99	CO_3H^-
10	8.02	**LiH^-**	33.02	$CNLi^-$	55.00	**LiO_3^-**	18.01	H_2O^-	36.00	C_3^-
11	7.02	**$^6LiH^-$**	24.04	**$LiNH_3^-$**	61.99	$CO_3H_2^-$	19.01	H_3O^-	64.04	$Li_2O_3H_2^-$
12	22.06	**Li_3H^-**	21.05	**Li_3^-**	39.00	**LiO_2^-**	33.99	**$H_2O_2^-$**	12.00	C^-
13	33.05	$Li_2OH_3^-$	38.04	CH_3OLi^-	62.02	**$Li_2O_3^-$**	31.99	**O_2^-**	63.03	$Li_2O_3H^-$
14	14.04	**$^6LiLiH^-$**	14.03	**Li_2^-**	38.01	**$^6LiO_2^-$**	47.03	**$Li_2O_2H^-$**	33.99	*PH_3^-*
15	24.00	C_2^-	13.03	**$^6LiLi^-$**	46.03	**$Li_2O_2^-$**	63.96	O_4^-	55.00	LiO_3^-

[a]Ion fragments marked in bold show the highest intensities in the peak area plots for the corresponding compound and that the chemical information on the identified ions fits well with the compound. Ion fragments marked in normal font do not show the highest intensities for the corresponding compound, but the chemical information may still fit. Ion fragments marked in italics indicate that the chemical information does not fit with the compound.

are applied to the associated test set. This allows keeping the test set as "unseen" for the model, which ensures the fairness of the model evaluation procedure. Considering the limited amount of training data (60 spectra from 5 pure compounds) and the large variance within the data, this process is critical to ensure robust results.

3. *Prediction*: As mentioned above, the fitted LR model was evaluated on 5 test sets (12 spectra each) of pure Li compounds through 5-fold CV, to ensure its correctness. Once the LR model accuracy for pure Li compounds was assessed, it was further applied to compound mixtures, to test its capability of identifying their compositions, and to real LMA samples. In this case, the probability of each sample belonging to any pure Li compound was computed through the trained LR model, and it was considered that a given Li compound was present when the associated probability was high enough to suggest this.
4. *Analysis*: To better analyze and interpret the results of the LR model, the data of ToF-SIMS spectra were reduced to low dimensions with the aid of PCA to visualize the classification results from the LR model. Moreover, feature importance obtained from the LR model's parameters was utilized to infer which were the specific ions allowing to better distinguish among different compounds. The final identification of the key specific ions was based on the average feature importance calculated during the 5-fold CV.

3.2. Identification of Specific Secondary Ions. The trained LR model was able to make accurate (>99%) classifications of the measured 5 kinds of pure Li compound samples. Its detailed performance evaluation and comparison with other common machine learning classification models are summarized in Supporting Information Section S4. Based on the feature importance given by the LR model, the characteristic peaks are identified for the 5 lithium compounds samples, and they are listed in Table 1, where the 15 top m/z values are reported in the order of importance. For Li_2CO_3, Li_3N, LiH, and LiOH, the LR model has identified at least 12 out of 15 characteristic peaks (marked in bold font) that have the highest intensities in the peak area plots and show good compatibility with their corresponding chemical nature. A few peaks also match chemical information despite their relatively low intensity (marked in normal font). Moreover, almost no secondary ions are found to mismatch their chemical information (marked in italics). Furthermore, for a better comparison, these data are also analyzed by the PCA method and are discussed in Supporting Information Section S1.

Besides, it is worth noting that the result for Li_2O is unsatisfying, as the selective identification of Li_2O through ToF-SIMS can be particularly difficult due to its similarities with Li_2CO_3. For example, some peaks, such as LiO^- and LiO_2^-, which could be intuitively regarded as specific for Li_2O, appear to be even more intense in the Li_2CO_3 samples. As a result, no prominent secondary ions solely specific to Li_2O can be determined by the LR model within our sample data set, and a specific analysis is illustrated in Supporting Information Section S5.

3.3. Identification of Compositions for Mixtures and LMAs Samples. The identified specific peaks for the different lithium compounds can help researchers recognize these materials from their ToF-SIMS spectra, especially for those having unique characteristic peak(s). In addition, it is desirable to be able to identify chemical species with no particular characteristic peak, such as Li_2O in the present samples. However, realistic samples may contain several different compounds, from which secondary ions could originate that overlap on the characteristic peaks identified for the pure Li compounds under consideration. This could significantly complicate the spectrometric analysis. To identify the compositions on a mixture sample, the obtained LR model is used to infer the possible components based on the probability values it outputs for different substances. Additionally, other MVA techniques, namely, PCA, NMF, and random forest model, are also tested and their results are provided as a comparison and complement to the LR model (discussed in Supporting Information Sections S1, S8, and S11).

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

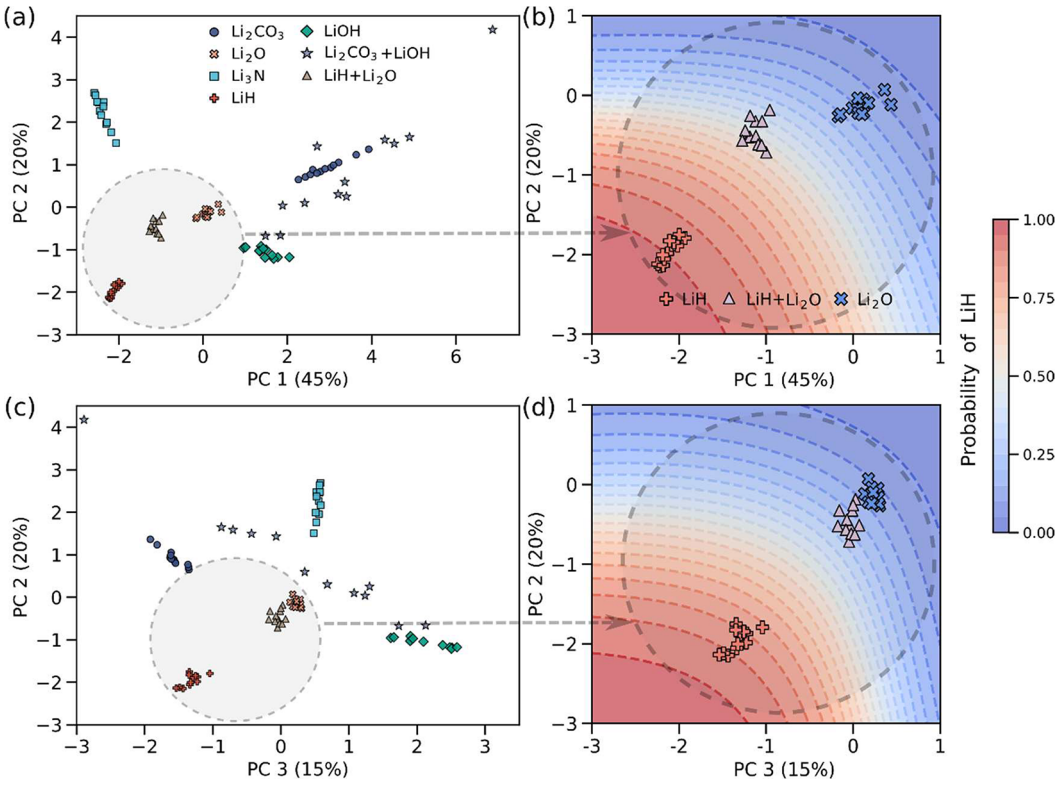


Figure 2. 2D scores plots (a) and (c) of the first 3 PCs (the corresponding proportions of variances are listed in parentheses) for the five pure lithium compounds and two mixture samples (Li_2CO_3+LiOH and LiH+Li_2O). Schematic diagrams of the decision boundary obtained from the trained LR model between LiH and Li_2O are shown next to the corresponding PCs in parts b and (d), where a higher probability of being LiH is colored in warmer color.

Table 2. Predicted Probability for the Mixture and LMA Samples from the LR Model

				composition probability				
no.	compositions (expected)	category	remark	Li_2CO_3	Li_2O	Li_3N	LiH	LiOH
1	Li_2CO_3+LiOH	mixture	1:1 mixed	**0.54**	0.00	0.00	0.00	**0.46**
2	LiH+Li_2O	mixture	1:1 mixed	0.00	**0.83**	0.03	**0.14**	0.00
3	Li_3N	LMA	N_2-plasma treated	0.01	**0.12**	**0.87**	0.00	0.00
4	Li_2CO_3+LiOH	LMA	smaller sputter dose	**0.72**	0.00	0.02	0.00	**0.26**
5	Li_2O+LiH	LMA	double sputter dose	0.00	**0.40**	0.05	**0.55**	0.00

3.3.1. Identification of Compositions for Mixture Samples Using LR. The obtained LR model not only performs classification but also outputs the probability of belonging to a given class, which can be used to infer the possible components in mixtures without much need for manual interpretation. To better understand the predicted probability values from the LR model and the work process of the model, 5 pure lithium compounds and 2 mixture samples (Li_2CO_3+LiOH and LiH+Li_2O) are projected to the low-dimension space with PCA for better visualization (shown in Figure 2a,c). It can be seen that both mixtures lie in the position between their corresponding two constituents. Besides, the decision boundary plots of LiH are also shown in Figure 2b,d. They serve as the schematic diagram to illustrate how the compositions of a Li_2O+LiH mixture sample can be identified by the LR model, i.e., the probability associated with its presence. These decision boundary plots were obtained as follows: the PC1−3 spaces were uniformly gridded into dense data points, which were then inversely mapped back to the original feature space (m/z peaks), and their corresponding probabilities of being LiH were calculated by the LR model. The inverse mapping is performed in this way: PCA scores $\boldsymbol{S}$ can be given by $\boldsymbol{S} = \boldsymbol{XW}$, where $\boldsymbol{X}$ is the raw (centered) data matrix and $\boldsymbol{W}$ is the matrix consisting of eigenvectors, therefore, the inverse reconstruction of data $\hat{\boldsymbol{X}}$ to the original feature space is calculated by $\hat{\boldsymbol{X}} = \boldsymbol{SW}^T$. In the schematic diagram, only pure LiH (red star), Li_2O (blue cross) substances, and the associated mixture samples (purple triangle) are shown for the convenience of visualization. The warmer the color, the higher the probability that LiH is present according to the model, while a cold color indicates a low probability. It can be clearly seen that the mixtures lie in between the two reference compounds and are biased toward Li_2O (especially in PC2−3), which is consistent with the predicted probability values (0.83 for Li_2O and 0.14 for LiH) for this mixture sample (sample No. 2 in Table 2). In conclusion, for mixture samples, their corresponding compositions can be estimated qualitatively based on the LR model's predictions.

Table 2 reports the probabilities predicted by the LR model for the two prepared mixtures (sample Nos. 1 and 2). For these mixtures, the model can correctly predict their

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

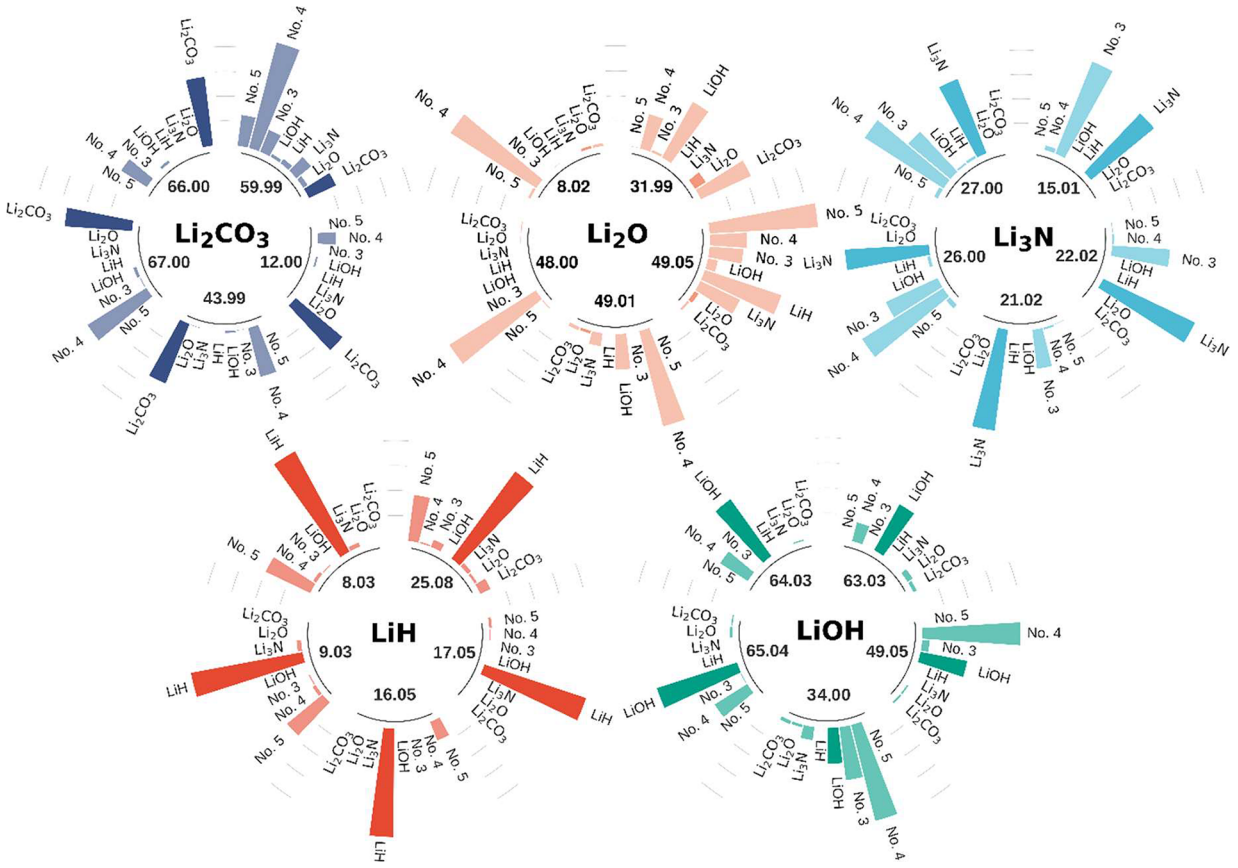


Figure 3. Scaled intensity values for the five top specific m/z peaks (labeled in each subplot) for five pure lithium compounds and three LMA samples in these peaks. Each circular area plot (in polar coordinates) represents one pure compound, where its corresponding intensity in each m/z peak is marked in a darker color, while other materials are marked in a lighter color.

compositions as it yields the highest probabilities for these corresponding compounds. Before examining the predicted values in detail, it should be noted that during the measurement of SIMS spectra, the secondary ion currents are ideally proportional to the amount of a species, but it usually could not strictly reflect the quantitative composition of the measured sample due to varying ionization probabilities caused by matrix effects. Therefore, the predicted values from the model are used only as a qualitative indicator rather than a quantitative metric. For the mixture sample No. 1, a mixture of Li_2CO_3 and LiOH in 1:1 weight proportion (molar ratio is about 1:3.1), the model predicts a probability value of 0.54 for Li_2CO_3 and 0.46 for LiOH, which match well the position of these mixture samples in Figure 2: They are indeed located in the middle of pure Li_2CO_3 and LiOH compounds, explaining why the model makes such a prediction of roughly equal proportions. On the other hand, for mixture No. 2 (1:1 weight mixed from Li_2O and LiH), the model predicts a probability of 0.83 for Li_2O and only 0.14 for LiH, implying that they may contain more Li_2O. This can be supported by the fact that these mixture samples lie toward Li_2O in decision boundary plots, indicating they are more similar to Li_2O. Furthermore, it is found that the model also predicts a very small value of 0.03 for Li_3N, which is not expected on this sample. Such a prediction is presumably due to the interference from noisy peaks, and a closer examination is given in Supporting Information Section S9. Overall, the analysis of the above results reveals that the possible compounds in a mixture can be qualitatively identified according to the predicted probability values from the LR model, although these values do not strictly correspond to their expected molar or weight ratios. Furthermore, the model can predict compounds that do not show specific peaks, namely, Li_2O, which could be helpful for an accurate determination.

3.3.2. Identification of the Compositions for LMA Samples. Based on the analysis above, the LR model was applied to the real LMA samples (Nos. 3−5 in Table 2). Besides, to better investigate and understand the reasons for the model to make such predictions, the intensity values of the LMA ToF-SIMS spectra in the top 5 characteristic m/z peaks of the five examined pure compounds, are exhibited in Figure 3. For better comparison, all spectra data are scaled together using the min−max scaling method. Furthermore, these samples are projected to low-dimensional space with PCA to facilitate intuitive inspection, which is shown and discussed in Supporting Information, Section S7.

Sample No. 3 is a lithium metal foil that was exposed to a N_2-plasma to produce Li_3N on its surface. As can be read from Table 2, the model predicts the highest probability value for Li_3N (0.87), which is in line with the expectation of nitride formation and can also be demonstrated by the large intensity values of these samples in the characteristic m/z peaks of Li_3N (shown in Figure 3). It is worth noting that the model outputs a small probability value for Li_2O (0.12), suggesting the possible presence of Li_2O. To investigate this inference, XPS measurements are performed on this sample, whose results not only confirm the formation of Li_3N but also reveal that there are indeed some oxide residuals on this sample (Supporting

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

Information Section S6). These are presumed to be introduced by either oxidation processes during sample transport or other contamination. Thus, the model successfully helps to identify the component that may be overlooked otherwise, and it reminds researchers to process the samples carefully to reduce possible influencing impurities.

In the following, particular attention will be given to the other two LMA samples, as they allow discussion of potential difficulties in the interpretation of the LR results and therefore its limitations. It should be noted that during the training process, the LR model has only seen the data of pure lithium compounds, but it is now applied to unseen samples that may contain multiple components. Before the ToF-SIMS spectra were measured, different sputter times were applied. The longer the sputtering time is, the more sample volume is ablated, and the subsequent analysis should give information about deeper parts of that sample.

For sample No. 4, only 1 frame (= 9×10^{13} ions/cm^2, smaller sputter dose) was sputtered to remove contaminants from the very outermost surface layer. The model correctly predicts the presence of Li_2CO_3 (0.72) and LiOH (0.26) on this sample, which is consistent with the XPS analysis of lithium metal foil, showing the existence of Li_2CO_3 and LiOH as published previously.[8] However, a direct comparison between their results is not possible, as ToF-SIMS is more surface sensitive (1−3 upmost atom layers) than XPS (about 10 nm). Therefore, these results could indicate that Li_2CO_3 and LiOH are the dominating components on the very top of the sample surface and consequently mainly detected with ToF-SIMS. In addition, it should be noted that an unusual data distribution of this sample can be observed in Figure 3, as intensity values in some peaks are even higher than that of the pure samples. For example, in the characteristic peak of Li_2CO_3 where m/z = 59.99 (CO_3^-), the intensity value of sample No. 4 is even ∼3.5 times higher than that of the pure Li_2CO_3 sample. This phenomenon is understandable as this sample was measured after a shorter sputter time to obtain information on the outmost layer. Such a large value, though this signal may come from additional contaminations like hydrocarbons, could dominate the model's behavior and may lead to misleading results. Hence, this sample is regarded as an outlier and to mitigate the possible negative impact, a data treatment is performed to truncate the particularly high peak values (detailed in Supporting Information Section S10).

For sample No. 5, a sputter time of 150 s was applied prior to the ToF-SIMS measurements, which allows for the investigation of a deeper part of the lithium metal foil. The model predicts the presence of Li_2O (0.40), and XPS analysis also indicates the presence of Li_2O mainly after longer sputtering, while going even deeper in the sample can lead to the detection of lithium metal.[8] Additionally, the model also predicts LiH (0.55) on this sample, which is difficult to determine with XPS measurements, as it can only rule out the presence of large amounts of LiH, making the potential identification of LiH through ToF-SIMS especially interesting. To further study this, the original ToF-SIMS spectra are investigated. Considering the longer sputtering time applied to sample No. 5, its outermost layers should have already been removed completely before the analysis. This agrees with the lack of peaks associated with Li_2CO_3 (illustrated in Figure 3), explaining why the model predicts a probability value of 0.0 for this compound. On the contrary, certain signal intensity is detected in the characteristic m/z peaks of LiH, which further confirms the accuracy of the model.

Besides, the very small probability values associated with Li_3N for samples No. 4 (0.02) and No. 5 (0.05) attract attention and raise concerns about the model's prediction. Similar to the case of mixture sample No. 2, the original peak profiles of this sample (especially characteristic peaks for Li_3N) are carefully examined to investigate whether the model has misclassified species. From the peak profiles (detailed in Supporting Information Section S9) it is known that certain strong intensities are detected in some characteristic peaks for Li_3N, which suggests a possible presence of Li_3N on these samples, showing that the model's predictions are reasonable. However, these peaks seem to be very noisy, which may interfere with the model's prediction, explaining why the LR model predicts a very low probability value of Li_3N. However, it is not particularly clear where these peaks originate from. Possible reasons could be that a small portion of particles was brought on these samples when sputtering the Li_3N LMA sample as the samples were in the analysis chamber at the same time. Alternatively, another option is that these two LMA samples were brought into contact inadvertently. On the one hand, the short period of sputtering time on sample No. 4 was not sufficient to remove all contaminations on the sample surface, which might lead to different charging of the LMA sample. This can result in a different data distribution in the measured spectrum of this sample compared to those of pure samples, for which a longer sputter time was applied. On the other hand, for sample No. 5, after a longer sputtering time, it has possibly reached the metallic lithium area, the signal of which may add a different intensity distribution to the measured spectra, which may interfere with the model to some degree, making an accurate prediction challenging. However, as the correlation of sputter times in ToF-SIMS and XPS is difficult and the lithium metal does not show any specific signals in ToF-SIMS analysis itself, from the available results alone it is impossible to infer whether the lithium metal is present or not. When this is the case, researchers need to examine the raw data to ensure the reliability of the model. Nevertheless, such samples can still be analyzed more easily and quickly with the help of the characteristic peaks previously identified by the model. Last, we emphasize the importance of considering comparable sputter times and the need to reduce surface contamination, which is critical to ensure the accuracy of machine learning models applied to ToF-SIMS analysis.

Finally, it should also be noted once more that the LR model has seen only the data of pure lithium compounds but it is now applied to unseen mixture samples with unusual data distribution (extrapolation), which is a considerably challenging task. A possible improvement would be to add mixture samples to the training process. However, the scarcity of mixture samples in this work hinders this realization. To overcome this problem, a data augmentation method is used to synthesize pseudomixture samples, with which the original problem of identifying composition is transformed as a regression problem, and a random forest regression model is fitted to predict the quantity of substance (detailed in Supporting Information Section S11). Such a model can better predict the presence of Li_3N, which can serve as a supplement and comparison to the LR model. In addition to this, increasing the number of data samples, e.g., performing ToF-SIMS imaging to record richer information, or preparing more samples of various pure substances to enrich the

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

database, is expected to improve the models' performance and is considered as a future work.

In summary, the LR model performs satisfactorily on the three tested LMA samples despite the challenge of extrapolation. With the help of the characteristic peaks obtained from the model and the predicted probability associated with the different compounds, the compositions of the mixture and LMA samples can be identified qualitatively in a faster and simplified way. Besides, it is worth noting that detecting and handling the outlier samples are necessary steps to offer a more robust result. Hence, a more elaborate measuring procedure would need to be designed to ensure as much as possible a consistent distribution of data across the sample to mitigate the possible negative effects. Moreover, depth profiles can be measured to investigate probabilities change as a function of sputter time utilizing the high sensitivity and lateral resolution of ToF-SIMS. This could be helpful in identifying any artifact introduced by sputtering and assist in making more confident interpretations about oxidation, surface segregation, and other related activities such as the matrix effects.[44,45]

4. CONCLUSIONS

In this work, machine learning methods, especially the logistic regression (LR) model, were used to analyze ToF-SIMS spectra of 5 different lithium compounds that are expected to appear on lithium metal anodes, namely, Li_2CO_3, Li_2O, Li_3N, LiH, and LiOH. The LR model was utilized to identify the characteristic ions describing each of these compounds based on their measured ToF-SIMS spectra, which match well with their chemical nature and can be used as reference values for facilitating related studies. The LR model was further applied to the compound mixtures and real lithium metal anode samples for composition identification. After preprocessing, the results are more than satisfying, as the model can qualitatively identify compositions with the help of the predicted probability of individual compounds. This is valuable since the model was trained only on pure compounds, but it was shown to be also applicable to unseen samples that may contain multiple components. In addition, the model has also helped the researcher to find previously overlooked substances on some samples, though they are inferred to be most likely introduced accidentally through contamination, hence reminding researchers to process samples carefully to reduce possible influencing impurities. Moreover, special attention was given to the samples obtained under different measurement conditions, whose different data distributions may mislead model judgments. Its original profile was carefully inspected to accurately identify the components. However, this process can be simplified by previous identifications of the most characteristic ion fragments associated with each pure Li compound.

In conclusion, the strategies presented in this work are promising for an intuitive, mostly automatized, and robust ToF-SIMS data analysis, which can assist the researcher in recognizing characteristic ions of compounds and identifying possible compositions of mixtures and LMA samples in a fast and intuitive way with few or no previous knowledge. In future work, the proposed method can be adapted to more complex samples, for example, samples with multicomponents and real solid-electrolyte interphases (SEIs), to further extend its applicability.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.3c09643.

Short introduction of logistic regression; comparison of different machine learning models; analysis of feature importance; XPS data of one studied sample; detailed analysis of results from machine learning models (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Arnd Koeppe − *Institute for Applied Materials − Microstructure Modelling and Simulation, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany;* orcid.org/0000-0002-4833-1306; Email: arnd.koeppe@kit.edu

Authors

Yinghan Zhao − *Institute for Applied Materials − Microstructure Modelling and Simulation, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany;* orcid.org/0000-0001-9440-138X

Svenja-K. Otto − *Institute of Physical Chemistry, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany*

Teo Lombardo − *Institute of Physical Chemistry, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany*

Anja Henss − *Institute of Physical Chemistry, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany;* orcid.org/0000-0001-5009-6512

Michael Selzer − *Institute for Applied Materials − Microstructure Modelling and Simulation, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany; Institute for Digital Materials Science, Karlsruhe University of Applied Sciences, D-76133 Karlsruhe, Germany*

Jürgen Janek − *Institute of Physical Chemistry, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany;* orcid.org/0000-0002-9221-4756

Britta Nestler − *Institute for Applied Materials − Microstructure Modelling and Simulation, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany; Institute for Digital Materials Science, Karlsruhe University of Applied Sciences, D-76133 Karlsruhe, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.3c09643

Author Contributions

‖Y.Z. and S.-K.O. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support by the Federal Ministry of Education and Research (BMBF, Bundesministerium für Bildung und Forschung) within the FestBatt-Cluster of Competence for Solid-state Batteries (Grant Nos. 03XP0433D and 03XP0435D), within the ProGraL project (Grant No. 03XP0427), and the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) under Germany's Excellence Strategy − EXC 2154 − Project number 390874152. The authors also acknowledge the support funded

by the BMBF and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK-BW) as part of the Excellence Strategy of the German Federal and State Governments in the project ExU-Kadi4X. In addition, S.-K.O. acknowledges the financial support (Kékulé scholarship) by the Funds of the Chemical Industry (FCI).

REFERENCES

(1) Janek, J.; Zeier, W. G. A solid future for battery development. *Nature Energy* **2016**, *1* (9), 1167.
(2) Cheng, X.-B.; Zhang, R.; Zhao, C.-Z.; Zhang, Q. Toward Safe Lithium Metal Anode in Rechargeable Batteries: A Review. *Chem. Rev.* **2017**, *117* (15), 10403−10473.
(3) Lin, D.; Liu, Y.; Cui, Y. Reviving the lithium metal anode for high-energy batteries. *Nature Nanotechnol.* **2017**, *12* (3), 194−206.
(4) Becking, J.; Gröbmeyer, A.; Kolek, M.; Rodehorst, U.; Schulze, S.; Winter, M.; Bieker, P.; Stan, M. C. Lithium-Metal Foil Surface Modification: An Effective Method to Improve the Cycling Performance of Lithium-Metal Batteries. *Advanced Materials Interfaces* **2017**, *4* (16), No. 1700166.
(5) Harry, K. J.; Hallinan, D. T.; Parkinson, D. Y.; MacDowell, A. A.; Balsara, N. P. Detection of subsurface structures underneath dendrites formed on cycled lithium metal electrodes. *Nat. Mater.* **2014**, *13* (1), 69.
(6) Maslyn, J. A.; Frenck, L.; Loo, W. S.; Parkinson, D. Y.; Balsara, N. P. Extended Cycling through Rigid Block Copolymer Electrolytes Enabled by Reducing Impurities in Lithium Metal Electrodes. *ACS Applied Energy Materials* **2019**, *2* (11), 8197−8206.
(7) Meyerson, M. L.; Sheavly, J. K.; Dolocan, A.; Griffin, M. P.; Pandit, A. H.; Rodriguez, R.; Stephens, R. M.; Vanden Bout, D. A.; Heller, A.; Mullins, C. B. The effect of local lithium surface chemistry and topography on solid electrolyte interphase composition and dendrite nucleation. *J. Mater. Chem. A* **2019**, *7* (24), 14882.
(8) Otto, S.-K.; Moryson, Y.; Krauskopf, T.; Peppler, K.; Sann, J.; Janek, J.; Henss, A. In-Depth Characterization of Lithium-Metal Surfaces with XPS and ToF-SIMS: Toward Better Understanding of the Passivation Layer. *Chem. Mater.* **2021**, *33* (3), 859−867.
(9) Karen, A.; Ito, K.; Kubo, Y. TOF-SIMS analysis of lithium air battery discharge products utilizing gas cluster ion beam sputtering for surface stabilization. *Surf. Interface Anal.* **2014**, *46* (S1), 344−347.
(10) Heller, D.; ter Veen, R.; Hagenhoff, B.; Engelhard, C. Hidden information in principal component analysis of ToF-SIMS data: On the use of correlation loadings for the identification of significant signals and structure elucidation. *Surf. Interface Anal.* **2017**, *49* (10), 1028−1038.
(11) Wagner, M. S.; Graham, D. J.; Ratner, B. D.; Castner, D. G. Maximizing information obtained from secondary ion mass spectra of organic thin films using multivariate analysis. *Surf. Sci.* **2004**, *570* (1), 78−97.
(12) Graham, D. J.; Castner, D. G. Image and Spectral Processing for ToF-SIMS Analysis of Biological Materials. *Mass Spectrom.* **2013**, *2*, No. S0014.
(13) van Nuffel, S.; Parmenter, C.; Scurr, D. J.; Russell, N. A.; Zelzer, M. Multivariate analysis of 3D ToF-SIMS images: method validation and application to cultured neuronal networks. *Analyst* **2016**, *141* (1), 90−95.
(14) Jackson, J. E. Principal Components and Factor Analysis: Part I—Principal Components. *Journal of Quality Technology* **1980**, *12* (4), 201−213.
(15) Jackson, J. E. *A User's Guide to Principal Components*; John Wiley & Sons, Inc, 1991.
(16) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom Intell Lab Syst.* **1987**, *2*, 37−52.
(17) de Juan, A.; Jaumot, J.; Tauler, R. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods* **2014**, *6* (14), 4964−4976.
(18) Trindade, G. F.; Abel, M.-L.; Lowe, C.; Tshulu, R.; Watts, J. F. A Time-of-Flight Secondary Ion Mass Spectrometry/Multivariate Analysis (ToF-SIMS/MVA) Approach To Identify Phase Segregation in Blends of Incompatible but Extremely Similar Resins. *Analytical chemistry* **2018**, *90* (6), 3936−3941.
(19) Trindade, G. F.; Bañuls-Ciscar, J.; Ezeh, C. K.; Abel, M.-L.; Watts, J. F. Characterisation of wood growth regions by multivariate analysis of ToF-SIMS data. *Surf. Interface Anal.* **2016**, *48* (7), 584−588.
(20) Marcoen, K.; Visser, P.; Trindade, G. F.; Abel, M.-L.; Watts, J. F.; Mol, J.; Terryn, H.; Hauffman, T. Compositional study of a corrosion protective layer formed by leachable lithium salts in a coating defect on AA2024-T3 aluminium alloys. *Prog. Org. Coat.* **2018**, *119*, 65−75.
(21) Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5* (2), 111−126.
(22) Lee, D. D.; Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401* (6755), 788−791.
(23) Visser, P.; Marcoen, K.; Trindade, G. F.; Abel, M.-L.; Watts, J. F.; Hauffman, T.; Mol, J.; Terryn, H. The chemical throwing power of lithium-based inhibitors from organic coatings on AA2024-T3. *Corros. Sci.* **2019**, *150*, 194−206.
(24) Heller, D.; Hagenhoff, B.; Engelhard, C. Time-of-flight secondary ion mass spectrometry as a screening method for the identification of degradation products in lithium-ion batteries—A multivariate data analysis approach. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* **2016**, *34* (3), No. 03H138.
(25) Schroder, K. W.; Dylla, A. G.; Harris, S. J.; Webb, L. J.; Stevenson, K. J. Role of surface oxides in the formation of solid-electrolyte interphases at silicon electrodes for lithium-ion batteries. *ACS Appl. Mater. Interfaces* **2014**, *6* (23), 21510−21524.
(26) Higgins, K.; Lorenz, M.; Ziatdinov, M.; Vasudevan, R. K.; Ievlev, A. V.; Lukosi, E. D.; Ovchinnikova, O. S.; Kalinin, S. V.; Ahmadi, M. Exploration of Electrochemical Reactions at Organic−Inorganic Halide Perovskite Interfaces via Machine Learning in In Situ Time-of-Flight Secondary Ion Mass Spectrometry. *Adv. Funct. Mater.* **2020**, *30* (36), No. 2001995.
(27) Madiona, R. M.; Winkler, D. A.; Muir, B. W.; Pigram, P. J. Optimal machine learning models for robust materials classification using ToF-SIMS data. *Appl. Surf. Sci.* **2019**, *487*, 773−783.
(28) Madiona, R. M.; Welch, N. G.; Russell, S. B.; Winkler, D. A.; Scoble, J. A.; Muir, B. W.; Pigram, P. J. Multivariate analysis of ToF-SIMS data using mass segmented peak lists. *Surf. Interface Anal.* **2018**, *50* (7), 713−728.
(29) Tuccitto, N.; Capizzi, G.; Torrisi, A.; Licciardello, A. Unsupervised Analysis of Big ToF-SIMS Data Sets: a Statistical Pattern Recognition Approach. *Analytical chemistry* **2018**, *90* (4), 2860−2866.
(30) Gardner, W.; Cutts, S. M.; Muir, B. W.; Jones, R. T.; Pigram, P. J. Visualizing ToF-SIMS Hyperspectral Imaging Data Using Color-Tagged Toroidal Self-Organizing Maps. *Analytical chemistry* **2019**, *91* (21), 13855−13865.
(31) Gardner, W.; Hook, A. L.; Alexander, M. R.; Ballabio, D.; Cutts, S. M.; Muir, B. W.; Pigram, P. J. ToF-SIMS and machine learning for single-pixel molecular discrimination of an acrylate polymer microarray. *Analytical chemistry* **2020**, *92*, 6587.
(32) Matsuda, K.; Aoyagi, S. Time-of-flight secondary ion mass spectrometry analysis of hair samples using unsupervised artificial neural network. *Biointerphases* **2020**, *15* (2), 21013.
(33) Aoyagi, S.; Fujiwara, Y.; Takano, A.; Vorng, J.-L.; Gilmore, I. S.; Wang, Y.-C.; Tallarek, E.; Hagenhoff, B.; Iida, S.-I.; Luch, A.; Jungnickel, H.; Lang, Y.; Shon, H. K.; Lee, T. G.; Li, Z.; Matsuda, K.; Mihara, I.; Miisho, A.; Murayama, Y.; Nagatomi, T.; Ikeda, R.; Okamoto, M.; Saiga, K.; Tsuchiya, T.; Uemura, S. Evaluation of Time-of-Flight Secondary Ion Mass Spectrometry Spectra of Peptides by Random Forest with Amino Acid Labels: Results from a Versailles Project on Advanced Materials and Standards Interlaboratory Study. *Analytical chemistry* **2021**, *93* (9), 4191−4197.

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478

(34) Lombardo, T.; Duquesnoy, M.; El-Bouysidy, H.; Årén, F.; Gallo-Bueno, A.; Jørgensen, P. B.; Bhowmik, A.; Demortière, A.; Ayerbe, E.; Alcaide, F. Artificial Intelligence Applied to Battery Research: Hype or Reality? *Chem. Rev.* **2022**, *122*, 10899.
(35) Hosmer, D. W.; Lemeshow, S.; Sturdivant, R. X. *Applied Logistic Regression*; Wiley, 2013.
(36) Brandt, N.; Griem, L.; Herrmann, C.; Schoof, E.; Tosato, G.; Zhao, Y.; Zschumme, P.; Selzer, M. Kadi4Mat: A Research Data Infrastructure for Materials Science. *Data Sci. J.* **2021**, *20*, 8.
(37) Graham, D. J.; Castner, D. G. Multivariate analysis of ToF-SIMS data from multicomponent systems: the why, when, and how. *Biointerphases* **2012**, *7* (1−4), 49.
(38) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825−2830.
(39) Graham, D. J.; Wagner, M. S.; Castner, D. G. Information from complexity: Challenges of TOF-SIMS data interpretation. *Appl. Surf. Sci.* **2006**, *252* (19), 6860−6868.
(40) Lee, J. L. S.; Gilmore, I. S.; Seah, M. P. Quantification and methodology issues in multivariate analysis of ToF-SIMS data for mixed organic systems. *Surf. Interface Anal.* **2008**, *40* (1), 1−14.
(41) Lee, J. L. S.; Gilmore, I. S.; Fletcher, I. W.; Seah, M. P. Multivariate image analysis strategies for ToF-SIMS images with topography. *Surf. Interface Anal.* **2009**, *41* (8), 653−665.
(42) Gilmore, I. *A Guide to the Practical Use of Multivariate Analysis in SIMS*; National Physical Laboratory, 2011.
(43) Efron, B.; Gong, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *American Statistician* **1983**, *37* (1), 36.
(44) Aoyagi, S.; Matsuda, K. Quantitative analysis of ToF-SIMS data of a two organic compound mixture using an autoencoder and simple artificial neural networks. *Rapid communications in mass spectrometry: RCM* **2023**, *37* (4), No. e9445.
(45) Shard, A. G.; Miisho, A.; Vorng, J.-L.; Havelund, R.; Gilmore, I. S.; Aoyagi, S. A two-point calibration method for quantifying organic binary mixtures using secondary ion mass spectrometry in the presence of matrix effects. *Surf. Interface Anal.* **2022**, *54* (4), 363−373.

https://doi.org/10.1021/acsami.3c09643
ACS Appl. Mater. Interfaces 2023, 15, 50469−50478