nature portfolio

Peer Review File

Automatic recognition of Pd ions in high-resolution mass spectra of multicomponent samples without visible fine isotope structures

Corresponding Author: Professor Valentine Ananikov

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript, Ananikov and coworkers introduced new machine learning based method to identify various atoms including palladium ions combining with mass spectrometers. The authors have improved upon the existing deisotoping method (ref 76) by extending it beyond singly charged ions and significantly accelerating the analysis process. They have aimed to apply this approach across various fields. Among these, the ability to detect trace amounts of Pd is indeed a major advantage, as even very small quantities of Pd can have a significant impact on experimental outcomes (ref 22). However, it is not entirely clear how this method can be practically applied to mechanistic studies in the way the authors suggest. For example, it would be more compelling if the method could enable mass spectrometric detection of key intermediates in palladium-catalyzed reactions, such as Pd–H species. While the time-dependent observation of Pd-I species is interesting, it is not fully convincing how this alone provides substantial mechanistic insight into the catalytic processes discussed. Despite the focus on catalysis, the study offers no new mechanistic understanding of Pd species in solution. Without a strong connection to chemical reactivity or molecular structure, the work misses the opportunity to provide deeper insight into catalytic systems.
The manuscript reports the development of a machine-learning pipeline for identifying palladium (Pd) and several other metal ions in complex mass spectrometry (MS) data. While the combination of graph-based deisotoping and neural network classifiers is technically interesting, the work lacks sufficient innovation, rigorous benchmarking, and real-world applicability to justify publication in a high-impact journal.

Additional comment
(1) The method had shown that it could be also available to analyze various atoms including nickel, copper, silver, and even halogen ions. Is it possible to expand the method to analyze additional atoms?
(2) In page 17 to 18, check the figure number, S23 and S24. There is no Figure S24 in the supporting information.

Reviewer #2

(Remarks to the Author)
The authors present an end-to-end machine learning workflow that converts raw high-resolution ESI-MS data into element specific information, with particular emphasis on detecting Pd at ultra-low amounts in multicomponent mixtures. A graph-based deisotoping algorithm first assembles peaks belonging to the same multicharged isotopic distribution. Classical ensemble models and a dual-encoder CNN then classify the presence of Pd, Ni, Cu, Ag, Cl and Br, even when fine isotope structure is unresolved. Overall, the study addresses an analytical need in catalysis by transforming routine MS instruments into automated metal scanning. Following points are several conceptual and methodological issues that need to be addressed before the work is suitable for publication.
1. Benchmarking transparency: Current test relies heavily on synthetic data and a tiny real set. Could authors curate larger benchmark set with >100 independent spectra, covering different resolving powers, polarity modes, and charge-state distributions? For example, showing whether the graph algorithm still leads when only double charged ions are present would be helpful for readers.
2. False-positive control at ultra-low intensities: Figure 6 claims confident Pd detection near 10-11 g/mL, yet the likelihood of

misclassifying random noise peaks grows rapidly in this regime. Could authors construct a blank series in which solvent-only injections are measured under identical settings and processed through the full pipeline? Plotting true-positive rate against false-positive rate across the concentration gradient would be informative for practitioners choosing cut-offs.
3. Scalability to additional elements: Scalability is one of many advantages of the manuscript, yet the current version gives readers no concrete pathway for doing so. Could authors outline a reproducible element-onboarding protocol? Specifying at minimum the followings would be insightful: 1) the number and variety of labelled spectra required to distinguish a new element, 2) whether the dual-encoder CNN must be retrained end-to-end or whether its layers can be frozen and only the output layer fine-tuned, 3) any failure cases where envelope similarity or charge-state ambiguity hinders classification.

Reviewer #3

(Remarks to the Author)
The paper describes a high-quality innovative work implementing the machine learning approaches for facilitating the analysis of the high-resolution mass spectra of liquid samples containing Pd down to ppm/ppb levels. Routine detection of Pd-containing compounds, especially at low concentrations, is crucial for accelerating the progress of chemistry and catalytic science, considering that Pd is widely used in fine chemistry synthesis and the automotive industry. The challenge is increased by the inactivity of Pd in NMR, in contrast to other noble metals with similar applications, such as Pt and Rh, which makes the development of MS-based approaches particularly valuable. Detection of Pd-contamination at ppm/ppb levels is also important to avoid contamination issues while developing more sustainable chemistry routes based on abundant metals such as Ni, Cu, Ag, etc. The work demonstrates the high potential of the novel analytic approach. It also provides the software for the data analysis, which can be potentially used by a wide community of users, including non-experts. All of this makes the results of this work extremely valuable for the wider community of chemists across different fields. The comments below can help the author further improve their manuscript:

1. While the machine learning data analysis part of the work presented in great detail is convincing, the part related to the experimental validation seems too short to appreciate, especially for non-expert users from the neighboring fields. It would be useful to add full experimental details for operando experiments, including the setups and the concentrations of each reagent. It would also be helpful adding a discussion about the advantages, challenges, and possible limitations of the data analysis in each case.

2. Concerning the use of the developed software, it would be helpful to understand if the software works with specific data formats and has any other limitations.

There are also minor corrections and suggestions:
3. In Figure 1b : “Three” should probably be replaced by “Tree”

4. In Table 1, it would be helpful to explain what AP@k means and also give more details about the other abbreviations not explained in the main text

5. In the main text, Figure S24 is mentioned, but it does not exist in SI. The authors should double check to which figures they refer.

Overall, I recommend a major revision.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Thank you for your comprehensive and thoughtful responses to my comments. I appreciate the substantial additional work undertaken to address the raised concerns. The expansion to 100 real deisotoping test spectra and 143 additional MS spectra across multiple instruments significantly strengthens the validation and addresses my concerns. The systematic analysis of true-positive and false-positive rates across the concentration gradient provides the practitioner-oriented guidance I requested. Finally, the element-onboarding protocol is well-documented and reproducible. The manuscript is now suitable for publication.

Reviewer #3

(Remarks to the Author)
The authors successfully answered to all the questions. I recommend the publication of the paper as it is.

Automated recognition of Pd species in reaction mixtures and samples with machine learning

Authors: Pavel E. Gurevich,† Konstantin S. Kozlov,† Artem S. Silverstov,‡ Lev E. Nersesyan,‡ Nikita I. Kolomoets,‡ Valentina V. Ilyushenkova, Julia V. Burykina, Daniil A. Boiko, Artem P. Vorozhtsov, and Valentine P. Ananikov*

Point-by-point responses to the comments and descriptions of the changes made in the revised manuscript

Particular parts of the revised manuscript mentioned in the replies below are highlighted in yellow, and this highlighted manuscript file was uploaded upon submission of the revision (files: Manuscript-highlighted.pdf and Supporting-Information_highlighted.pdf).

The authors are grateful to the Editor and Reviewers for detailed manuscript reading and numerous important comments with deep insight into the subject of the manuscript!

Summary of the significant changes made in this revision:

1. Experimental data were expanded with new mass spectra with different resolving powers and polarity modes to evaluate model performance and robustness on real spectra.
2. Expanded section with real-world applicability using machine learning-based analysis of the spectra of 4 additional reactions, increasing our coverage to a total of 10 reactions. An example of how a presented spectra analysis approach can be used for mass spectrometric investigations of Pd species in catalytic systems was described. (see Manuscript page 17 paragraph 2, and Supporting Information section **SI4.4**).
3. A pipeline for the onboarding of a new element is added to our double encoder classifier. It includes code and detailed instructions in the GitHub repository of the project. Additionally, the results are described in the Supporting Information, section **SI3.8**.
4. Added information about the false-positive control at low intensities (section **SI4.3**).
5. The code in the GitHub repository of the project was polished, and detailed instructions for data generation, model training and inference, and end-to-end usage were added.
6. The title of the manuscript was changed to “Automated recognition of Pd species in reaction mixtures and samples with machine learning” to provide a more accurate description of the work.

7. A new coauthor, Artem Vorozhtsov, who contributed to the code development and independently validated the GitHub repository for this project, was added. He also participated in the processing, labeling and analysis of new experimental data.

We believe that the manuscript has been significantly improved, and we thank the Reviewers for their extremely helpful feedback. A detailed description of the changes and replies to the comments are provided below.

Reviewer #1

In this manuscript, Ananikov and coworkers introduced new machine learning based method to identify various atoms including palladium ions combining with mass spectrometers. The authors have improved upon the existing deisotoping method (ref 76) by extending it beyond singly charged ions and significantly accelerating the analysis process. They have aimed to apply this approach across various fields. Among these, the ability to detect trace amounts of Pd is indeed a major advantage, as even very small quantities of Pd can have a significant impact on experimental outcomes (ref 22).

We are grateful to the Reviewer for the evaluation of our manuscript.

However, it is not entirely clear how this method can be practically applied to mechanistic studies in the way the authors suggest. For example, it would be more compelling if the method could enable mass spectrometric detection of key intermediates in palladium-catalyzed reactions, such as Pd–H species. While the time-dependent observation of Pd-I species is interesting, it is not fully convincing how this alone provides substantial mechanistic insight into the catalytic processes discussed.

We thank the Reviewer for this excellent point. One of the major challenges in mechanistic studies of catalytic reactions is the detection of Pd-containing intermediates, which are often present at low concentrations. As demonstrated in **Figure 6a** of the manuscript, our machine learning (ML) model is capable of detecting Pd species at submicromolar concentrations. This enables researchers to use our model as a sensitive and automated tool for Pd detection in reaction mixtures, thereby facilitating mechanistic investigations.

To demonstrate the effectiveness of the developed pipeline in the mass spectrometric detection of Pd—H species, mass spectra of three Pd–NHC complexes (BIMePdI_2Py, IPrPdBr_2Py, and IPrPdCl_2Py) in methanol solutions were prepared. As can be seen on **Figure R1**, Pd—H species can be reliably identified with the developed models across all three complexes. It should be pointed out that all of these hydride

complexes were formed *in situ* after coordination of the solvent to the precatalyst and subsequent rearrangement of (NHC)PdX_2H, followed by elimination of the NHC ligand, leading to the formation of [PdX_2H]- ions.

Indeed, the developed method can accelerate and simplify research by automatically processing and interpreting a large set of mass spectrometry data. The overall pipeline is presented in **Figure R2a**.

We added a discussion about Pd—H species in Section **SI4.4** along with a reference to the section in the manuscript (page 17, paragraph 2).

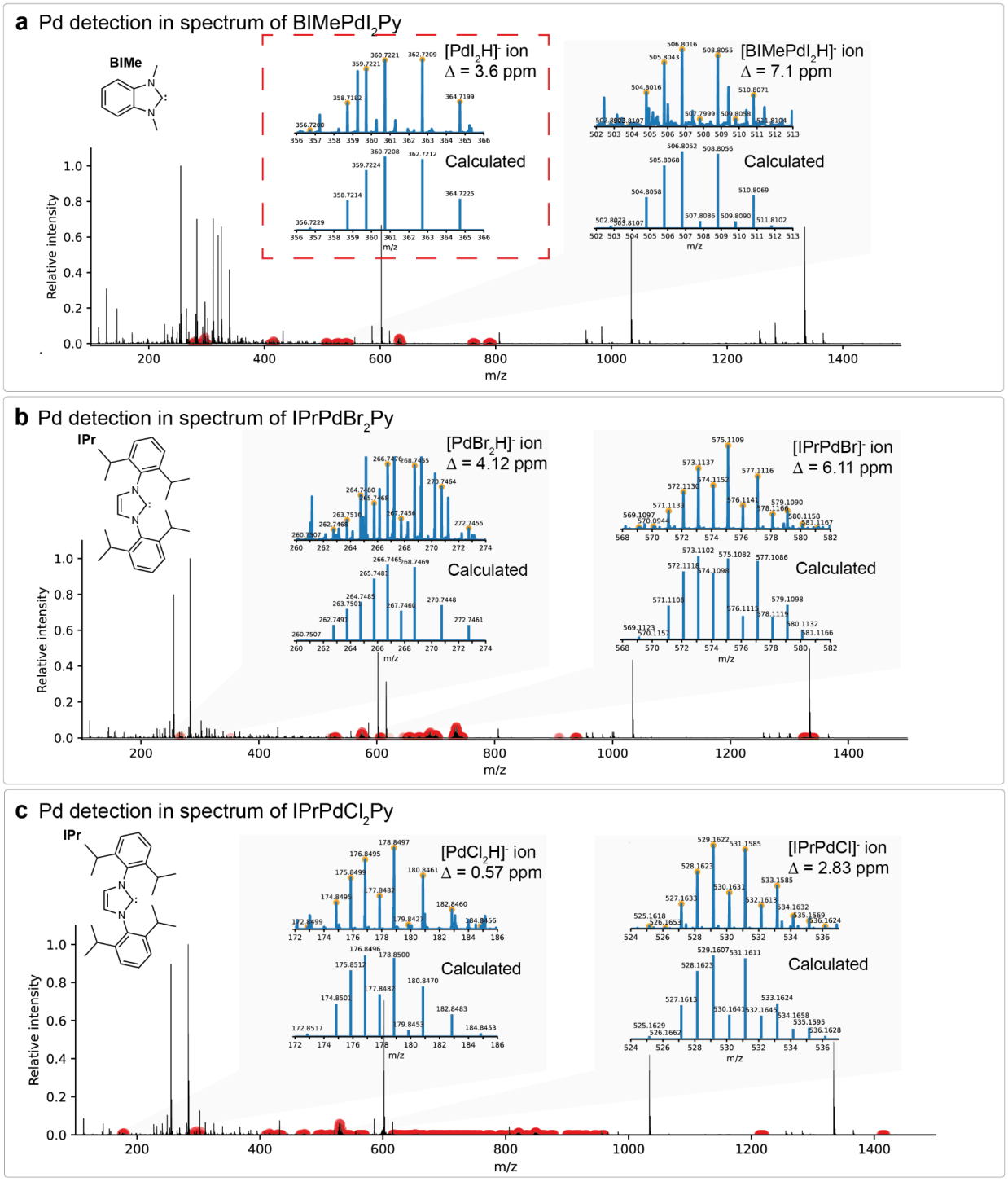

Figure R1. Automatic Pd detection in catalytic systems containing Pd—H species. a — spectrum of $BIMePdI_2Py$; b — spectrum of $IPrPdBr_2Py$; c — spectrum of $IPrPdCl_2Py$ (the figure is presented in the Supporting Information as Figure S27).

Despite the focus on catalysis, the study offers no new mechanistic understanding of Pd species in solution. Without a strong connection to chemical reactivity or molecular structure, the work misses the opportunity to provide deeper insight into catalytic systems.

Thank you for this important comment. We would like to clarify that the primary mechanistic contribution of the present work lies not in proposing new reaction pathways or assigning complete molecular structures, but in establishing, for the first time, a reliable, automated, and time-resolved method for profiling Pd species under real reaction conditions at ultralow concentrations. This capability itself represents a critical and previously missing layer of mechanistic information.
Importantly, our approach already delivers direct mechanistic observables, not merely abstract signal classification. For example, in the operando ESI-HRMS monitoring of the Pd/NHC system (Video 1), we demonstrate real-time transformation of Pd–Cl species into Pd–I species upon KI addition, accompanied by a complete suppression and re-emergence of Pd-containing distributions. This directly visualizes dynamic halide exchange, metal redistribution, and speciation shifts, which are important mechanistic events in Pd catalysis but are extremely difficult to capture at such low metal concentrations by conventional methods.

Furthermore, our approach enabled mechanistic insights in:

- discrimination between heterogeneous Pd reservoirs and leached homogeneous Pd species;
- time-dependent tracking of Pd–ligand ensembles; and
- quantitative monitoring of metal participation vs. metal scavenging,

all of which are key elements of modern cocktail-type and dynamic catalysis frameworks. In this sense, the work provides a fundamentally new type of mechanistic access to catalytic systems.

With respect to molecular structure: full structural assignment of each detected ion is indeed a subsequent, higher-level step that requires integration with reaction-specific chemical knowledge and additional fragmentation experiments. However, such structure-based analysis is critically dependent on the prior, unbiased identification of Pd-containing ions, which is exactly the unresolved bottleneck that our work addresses. Without this step, mechanistic MS studies at low metal levels are inherently incomplete or biased.

Finally, we stress that the absence of automated Pd recognition has been a long-standing barrier in mechanistic catalysis, contaminant catalysis, and environmental Pd monitoring. By removing this barrier, the present work does not replace structural or reactivity-based mechanistic analysis, but makes it technically feasible at metal-detection regimes that were previously inaccessible. We have clarified this point in the revised manuscript and added a comment to the Conclusions (page 21).

The manuscript reports the development of a machine-learning pipeline for identifying palladium (Pd) and several other metal ions in complex mass spectrometry (MS) data. While the combination of graph-based deisotoping and neural network classifiers is technically interesting, the work lacks sufficient innovation, rigorous benchmarking, and real-world applicability to justify publication in a high-impact journal.

Thank you for the evaluation of the technical implementation of the analysis approach. We enhanced the experimental dataset and added examples of real-world applicability for the detection of Pd—H species in the mass spectra of Pd/NHC complexes (Section **SI4.4**).

Additional examples of how developed detection models can be integrated into research pipelines are described with an analysis of 4 different catalytic systems:

1. Disulfide addition to alkynes
2. Oxidative addition
3. Buchwald–Hartwig reaction
4. Ethynyl-NHC coupling

Using the approach developed in the present study, determination of the presence of ions in the spectrum can be performed. These algorithms and corresponding code are included in our package. The pipeline is schematically depicted in **Figure R2a**.

As an example of real-world applicability, we show how Pd species can be searched in reaction spectra (**Figures R2, R3**). All found ion isotopic distributions were labeled as Pd-containing. The results of the ion search with Pd detection models can be summarized as follows.

In the disulfide addition to alkyne reaction, **a1** (m/z = 731.2974) and **a2** (m/z = 631.1597) ions were found. In this experiment, **a2** is the ionization product of a palladium complex, which is the result of the insertion of palladium through an S-S disulfide bond with a phosphine ligand.

In the oxidative addition reaction, **a3** (m/z = 409.0904) is a product of intramolecular palladium insertion via the C–H bond of the NHC ligand with the formation of a Pd(II) complex that has lost one ligand during ionization. Ion **a4** (m/z = 445.0668) is an ionization product of the Pd(II) complex with the NHC ligand. The ions **a3** and **a4** found are not specific to oxidative addition.

The formation of **a5** (m/z = 487.1387) in the Buchwald–Hartwig reaction is the result of ionization (with bromine anion elimination) of the product of the oxidative addition of bromobenzene to the NHC–Pd(0) complex.

The mechanism of ethynyl-NHC coupling has been debated in the literature. In our study of this reaction, **a6** (m/z = 409.0884) is a product of intramolecular palladium insertion via the C-H bond of the NHC ligand with the formation of a Pd(II) complex that lost one ligand during ionization. Ion **a7** (m/z = 511.1397) is either the result of palladium complex ionization prior to reductive elimination, containing an NHC ligand and an acetylene fragment, or the result of palladium coordination in the product of an ethynyl-NHC coupling. The mechanism of formation of **a8** (m/z = 613.1819) is identical to that of **a7** but involves the addition of an extra phenylacetylene molecule via a triple bond of the ethynyl-NHC combination product. We also registered an ethynyl-NHC coupling product ($C_{29}H_{29}N_2^+$, *m/z* 405.2347).

We added a discussion of the reaction spectra analysis in Section **SI4.5** along with a reference to the section in the manuscript (**page 17, paragraph 2**). Plausible structures of the detected ions **a1 – a8** are shown in the Supporting Information as Table S9.

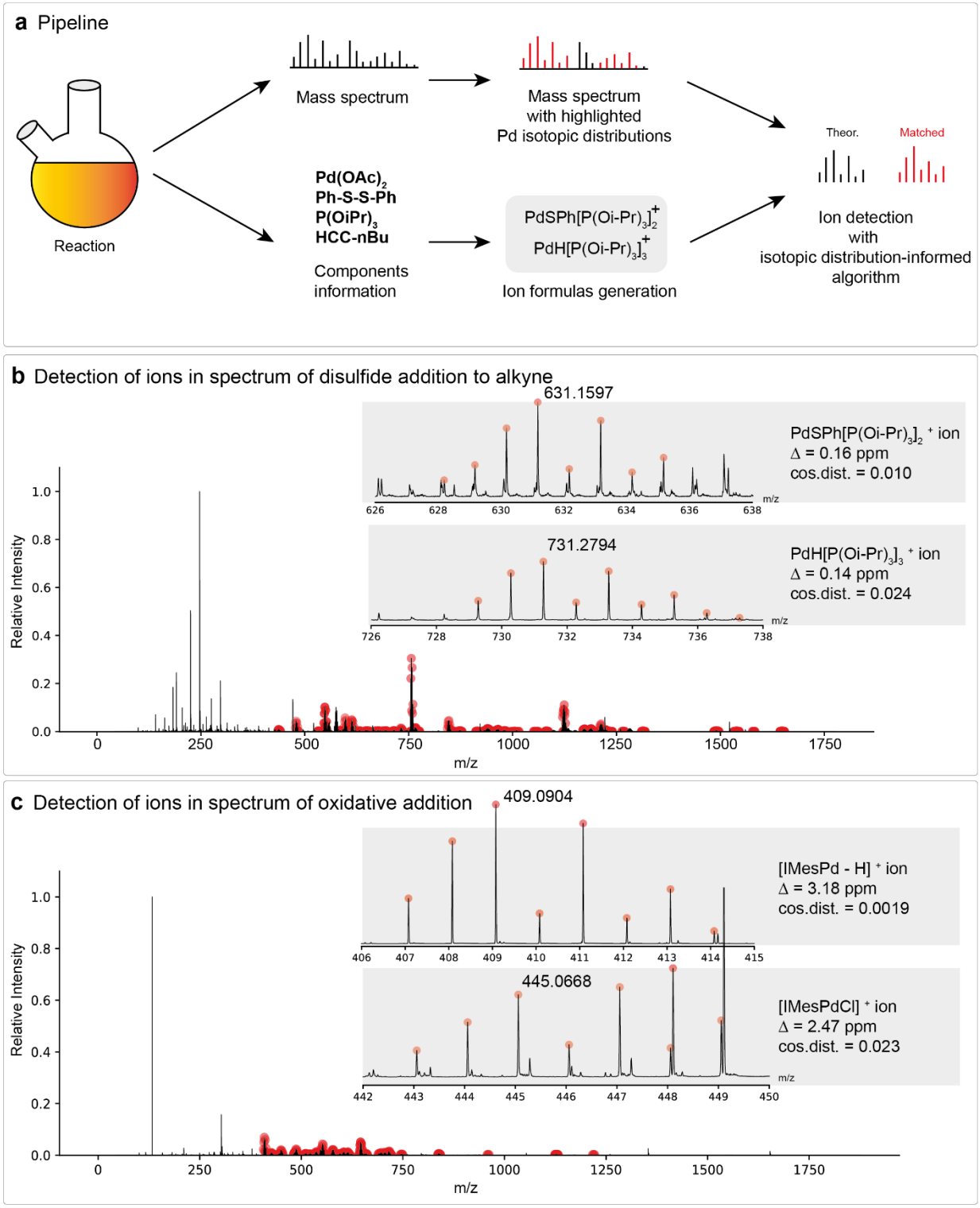

Figure R2. Detection of palladium-containing ions in the mass spectra of different reactions.

a — Pipeline for automated detection; b — Disulfide addition reaction; c — Oxidative addition reaction (the figure is presented in the Supporting Information as Figure S28).

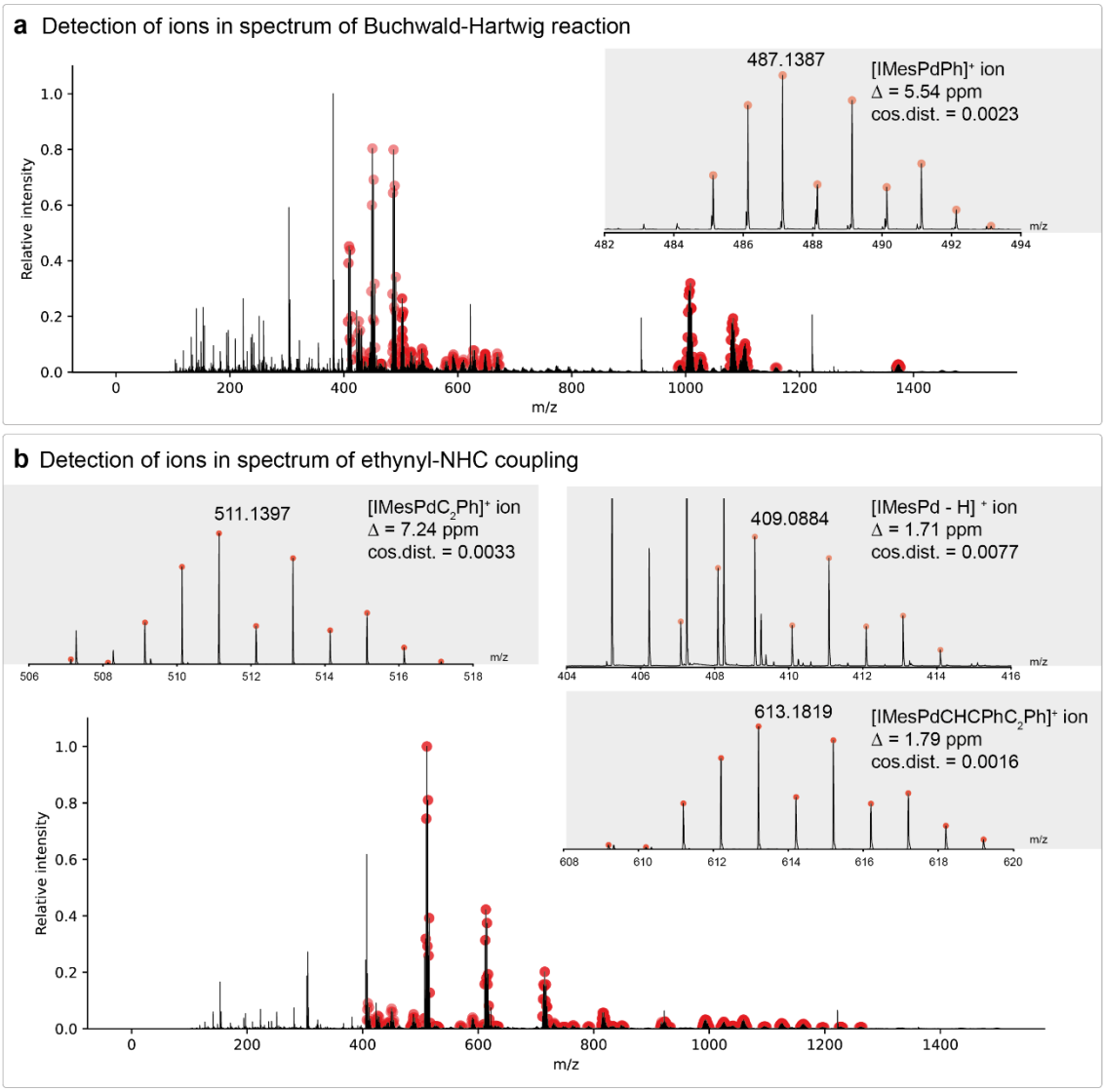

Figure R3. Detection of palladium ions in the mass spectra of different reactions.

a — Buchwald-Hartwig reaction; b —ethynyl-NHC coupling (the figure is presented in the Supporting Information as Figure S29).

Additional comment

(1) The method had shown that it could be also available to analyze various atoms including nickel, copper, silver, and even halogen ions. Is it possible to expand the method to analyze additional atoms?

We appreciate this important question. The possibility of extending the method to additional elements is now discussed in the “Elements extension” subsection of the “Results and Discussion” section (p. 14). For further clarification, we have prepared all necessary experiments, code, and detailed instructions for incorporating new elements, which are now available in the project’s GitHub repository and in the Supporting Information, section **SI3.8**. Users can either retrain from scratch or fine-tune the original double-encoder model on a single CPU in approximately 15 minutes.

As Figure **S23** shows, even 500 additional labeled ions containing a new element are sufficient for the model to predict it efficiently. We also observed that adding a new element generally improved the prediction quality for the original elements (**Table S6**).

However, in some cases, the addition of a new element may be challenging or even impossible, such as monoisotopic elements. Without recognizable isotopic patterns, neural network will not able to detect such elements. Another challenge is a pair of elements with very similar isotopic distributions. For example, silver and bromine both have two isotopes with nearly identical mass differences and similar relative intensities. This similarity led to occasional misclassification, where silver-containing ions were incorrectly identified as bromine-containing ions. Notably, the test ROC AUC decreased exactly for silver after Br onboarding. Additional support for this explanation is provided by the dimensionality-reduction visualizations of the model's hidden representations (Manuscript, **Fig. 4**): both the PCA and t-SNE plots show bromine- and silver-related ions forming a shared cluster, indicating that the model encodes these ion types very similarly.

(2) In page 17 to 18, check the figure number, S23 and S24. There is no Figure S24 in the supporting information.

We apologize for this oversight and thank the Reviewer for bringing it to our attention. We have carefully reviewed all the figure references in both the manuscript and the supporting information and have corrected all the inaccuracies. We are now confident that all the figure references are accurate and consistent.

The authors thank the Reviewer for the careful reading and comments, which helped to improve the manuscript!

Reviewer #2

The authors present an end-to-end machine learning workflow that converts raw high-resolution ESI-MS data into element specific information, with particular emphasis on detecting Pd at ultra-low amounts in multicomponent mixtures. A graph-based deisotoping algorithm first assembles peaks belonging to the same multicharged isotopic distribution. Classical ensemble models and a dual-encoder CNN then classify the presence of Pd, Ni, Cu, Ag, Cl and Br, even when fine isotope structure is unresolved. Overall, the study addresses an analytical need in catalysis by transforming routine MS instruments into automated metal scanning.

We are grateful to the Reviewer for the evaluation of our manuscript.

Following points are several conceptual and methodological issues that need to be addressed before the work is suitable for publication.

1. Benchmarking transparency: Current test relies heavily on synthetic data and a tiny real set. Could authors curate larger benchmark set with >100 independent spectra, covering different resolving powers, polarity modes, and charge-state distributions? For example, showing whether the graph algorithm still leads when only double charged ions are present would be helpful for readers.

Thank you for highlighting this point! To address your concern, we enhanced both the deisotoping algorithm and neural network benchmarking.

Deisotoping algorithm

We enhanced the real test set for deisotoping. The actual test size is equal to 100 spectra of ions of various elemental compositions (Pd, Ni, Ir, Si, Br, Cl, N, C, H, O) and charges (88 single and 12 double charged ions). The algorithm successfully merges prominent peaks in all test ions without unexpected behavior. In some cases, low-intensity peaks are not labeled because of the imperfection of automated peak picking strategies (**Figure R4**). However, their absence is not crucial because elemental recognition with isotopic distributions can still be successfully achieved. Additionally, we implemented manual peak picking threshold settings directly into the deisotoping algorithm to increase the flexibility of deisotoping for users. We made necessary corrections in the manuscript (**page 6, paragraph 1**) and added information about the real test set for deisotoping in **Section SI1.5**.

We also performed a computational experiment on graph deisotoping algorithm behavior in cases where all ions in multicomponent mass spectra are double charged. Unfortunately, obtaining such a spectrum experimentally if practically infeasible because the vast majority of ions in experiments are single charged. However, the creation of such a synthetic spectrum was performed, and the deisotoping algorithm keeps working correctly on spectra where all ions are double charged (**Figure R5**). We added the results of the experiment to **Section SI 1.6**.

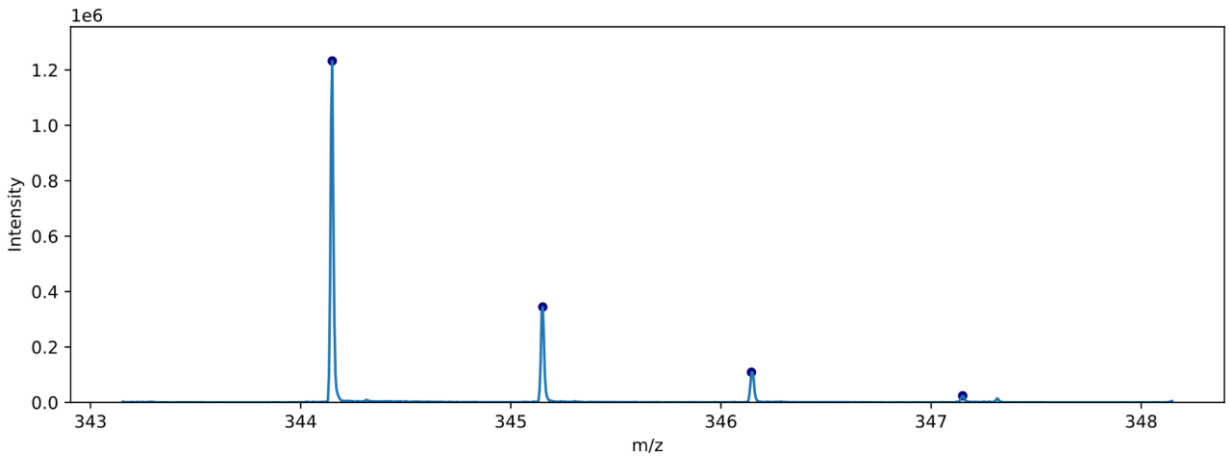

Figure R4. An example of the data sample ($C_{18}H_{25}NO_2Si_2H^+$ ion) from the enhanced real set. All these spectra figures can be downloaded from Figshare data repository.

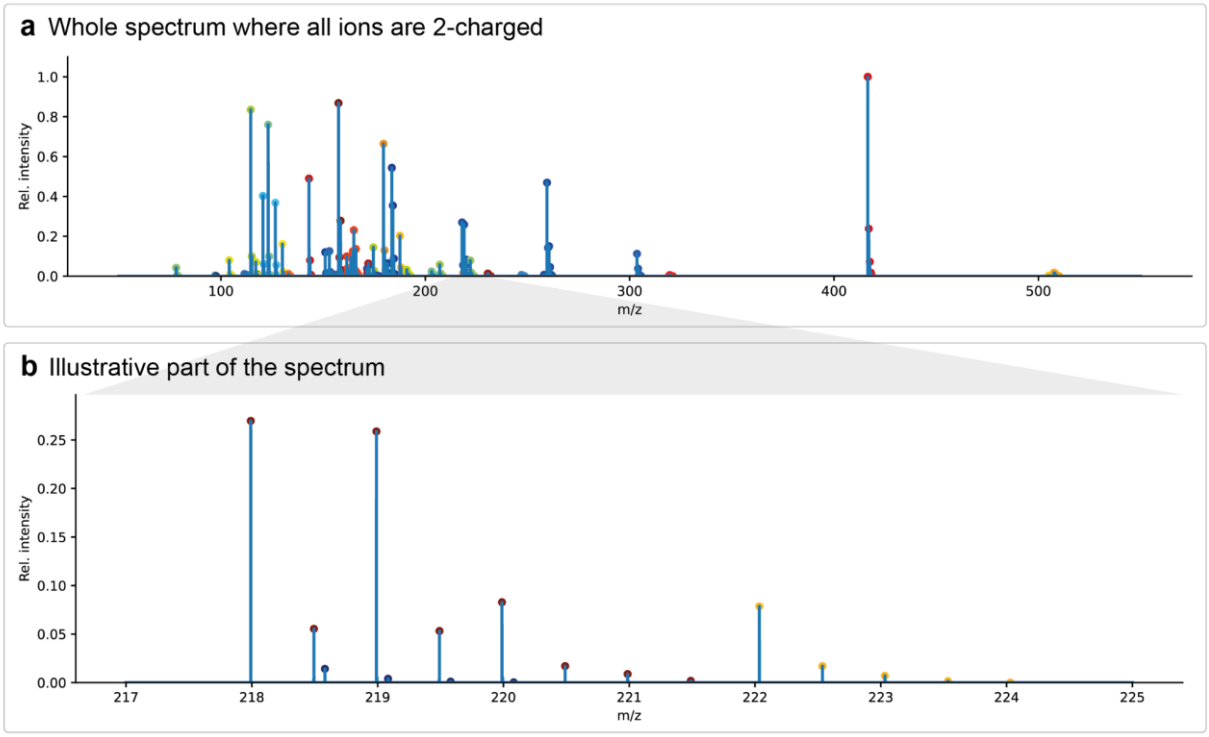

Figure R5. Example of graph deisotoping algorithm performance where all ions are double charged (the figure is presented in the Supporting Information as Figure S13).

Neural network

To enhance and improve the study, we recorded 143 additional mass spectra, including Pd complexes in different concentration ranges, spectra recorded with different mass analyzers (Q-TOF, FT/ICR MS, TOF), and spectra of different Pd-catalyzed reaction mixtures (Heck, Suzuki, and Sonogashira reactions; oxidative addition reaction; reaction with Pd-H species formation). From these spectra, we obtained 25 Pd isotopic distributions and included them in our test set. The Pd test ROC AUC value changed from **0.9819** to **0.9991** for the double-encoder architecture. The other metrics slightly increased or remained at approximately the same value. This demonstrates the robustness of our model and the generality of the existing test set. The test evaluation data were updated in the manuscript and Supporting Information.

Additionally, we compared our pipeline performance for mass analyzers with different resolving powers: time-of-flight (TOF), quadrupole time-of-flight (Q-TOF) and Fourier transform ion cyclotron resonance (FT/ICR); the resolving powers are ~ 10 000, ~26 000 and > 100 000, respectively. Note that the difference between the TOF and Q-TOF instruments should be attributed to specific instrument models, as the quadrupole and collision cells do not affect the resolving power.

We recorded the spectra of 5.6×10^{-6} mol/L $PdCl_2$ solutions via these instruments and analyzed them with our end-to-end pipeline (**Fig. R6**). Key results of our comparison are as follows:

1. Spectral analysis process starts with selecting initial signals for consideration. As different methods have different noise profiles, the peak threshold becomes one of the most important

hyperparameters. Specifically, after optimization, we choose the quantile threshold determination algorithm with 98.5, 99.5 and 99.98 quantiles for the TOF, Q-TOF and FT/ICR, respectively.

2. After threshold tuning, the deisotoping algorithm performs well with all three mass analyzers. This demonstrates the robustness of our deisotoping–classification pipeline across different mass analyzers.
3. Molecular ions were detected in all three spectra. However, additional Pd distributions were not found in the spectrum obtained from the TOF instrument. The most “balanced” spectrum was recorded with a Q-TOF instrument: additional Pd distributions are distinguishable and are not mixed with noise peaks.
4. Several Pd distributions were identified by the deisotoping algorithm with minor errors, such as missed peaks or substitution with noise peaks. Nevertheless, our double-encoder classifier, trained with augmentations, was still able to correctly detect Pd in these “corrupted” distributions. These findings demonstrate that our neural network is highly robust and can effectively compensate for imperfections in the deisotoping algorithm.
5. Overall, these additional studies confirmed the wide applicability of the described approach and emphasized that future readers should consider instrument-specific features.

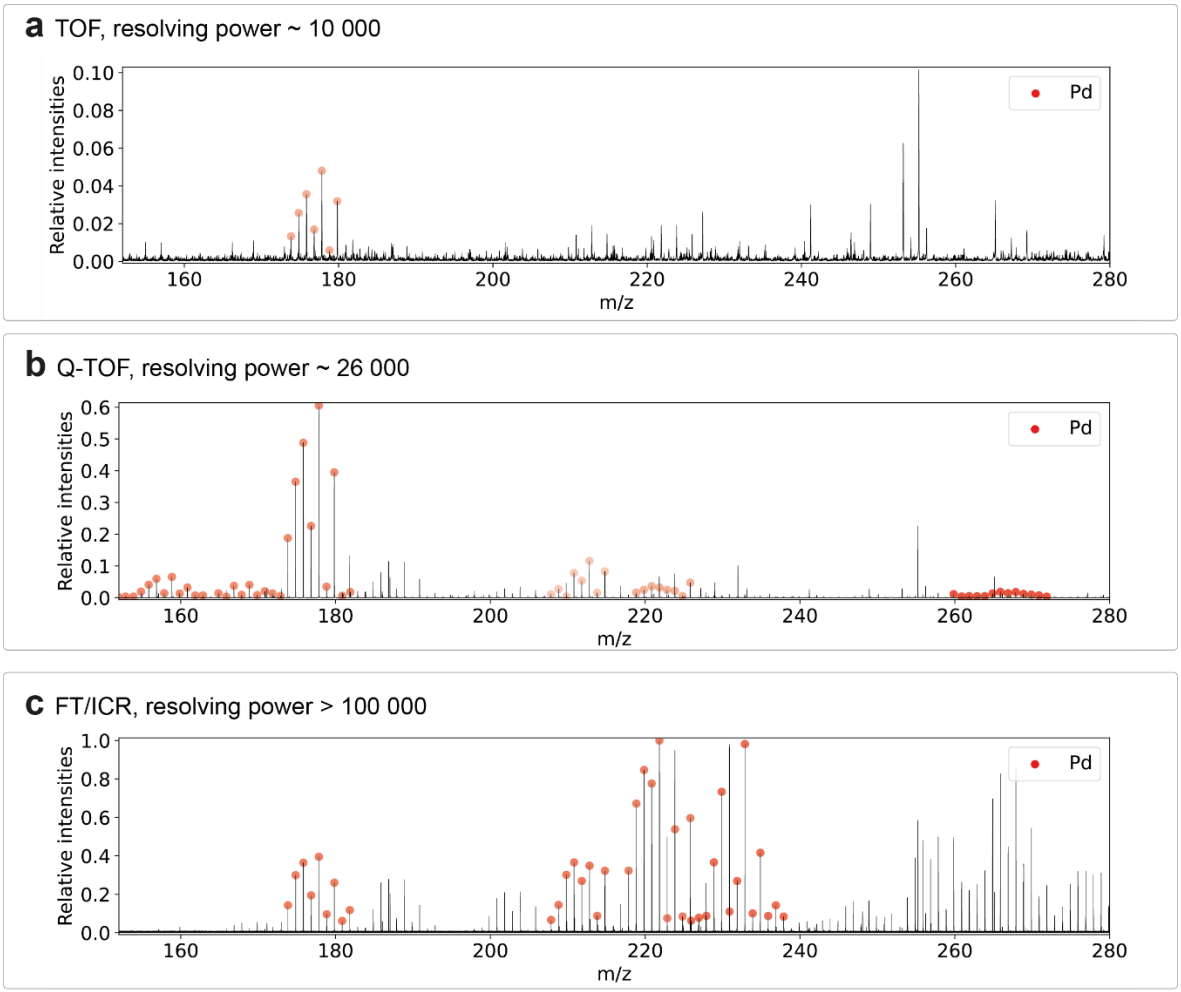

Figure R6. Comparison of Pd detection pipeline performance for spectra acquired using mass analyzers with different resolving powers: TOF (~10,000), Q-TOF (~26,000), and FT/ICR (>100,000). As a result, we have considerably improved our benchmarking strategy, including validation across multiple instruments, showing applicability of the method across a broad set of possible resolving powers. Results were added to the Supporting Information, section **SI4.6** (the figure is presented in the Supporting Information as Figure S30).

2. False-positive control at ultra-low intensities: Figure 6 claims confident Pd detection near 10-11 g/mL, yet the likelihood of misclassifying random noise peaks grows rapidly in this regime. Could authors construct a blank series in which solvent-only injections are measured under identical settings and processed through the full pipeline? Plotting true-positive rate against false-positive rate across the concentration gradient would be informative for practitioners choosing cut-offs.

Thank you for this helpful idea. To implement this, we obtained all the model predictions for the spectra related to the low-concentration experiments from the manuscript. Next, specialist in mass spectrometry validated the results and calculated the number of false positive and false negative predictions. On the basis of these estimations, we observed that false positive predictions occur only for

Pd concentration of 10^{-8} mol/L (**Table R1, Figure R7**). We added these results to the manuscript and Supporting Information.

Table R1. Pd detection pipeline performance for spectra with different Pd concentrations (the table is presented in the Supporting Information as Table S8).

Pd concentration, mol/L	# of manually detected Pd distributions	TP	TN	FP	FN	TPR	FPR	Accuracy
10^{-3}	27	18	16	0	9	0.6667	0	0.79
10^{-4}	21	15	17	0	6	0.7143	0	0.84
10^{-5}	19	10	22	0	9	0.5263	0	0.78
10^{-6}	13	7	56	0	6	0.5385	0	0.91
10^{-7}	9	4	77	0	5	0.4444	0	0.94
10^{-8}	2	0	83	4	2	0	0.0460	0.93
10^{-9}	0	0	103	0	0	-	0	1.00

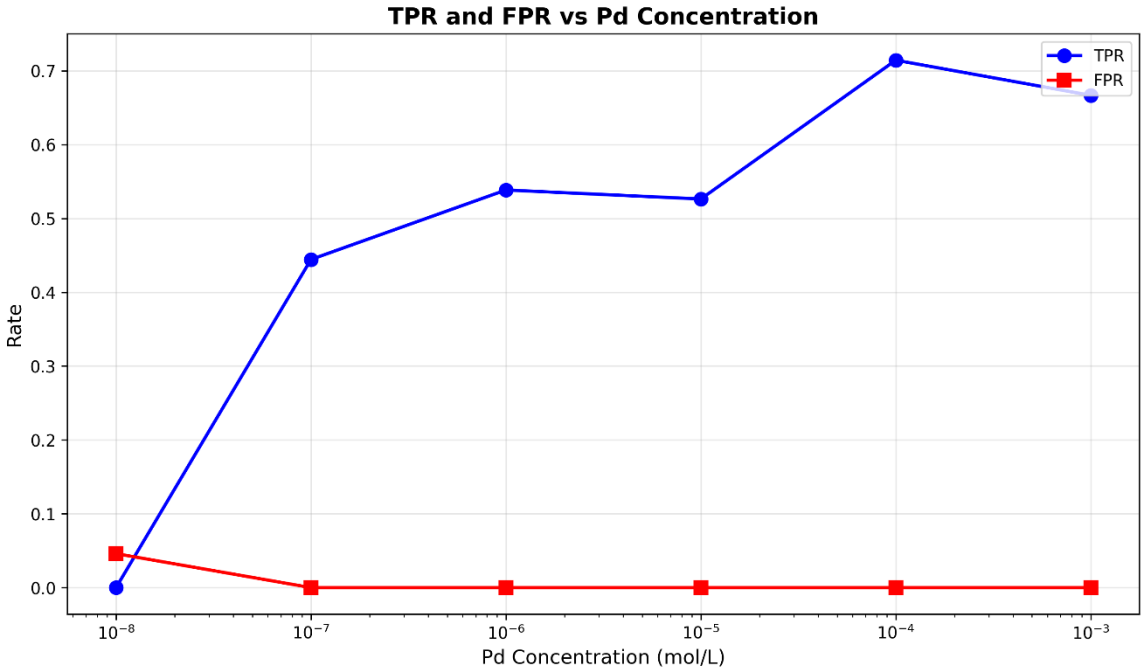

Figure R7. TPR and FPR values for different Pd concentrations (the figure is presented in the Supporting Information as Figure S26).

3. Scalability to additional elements: Scalability is one of many advantages of the manuscript, yet the current version gives readers no concrete pathway for doing so. Could authors outline a reproducible element-onboarding protocol?

Thank you for highlighting this essential topic. To address your request, we conducted a dedicated experiment to systematically evaluate and document the process of onboarding a new element:

1. We trained our double-encoder model from scratch on the original training dataset, excluding all labeled bromine (Br) distributions.
2. We then added 100, 500, 1,000, 5,000, and 8,000 Br samples to the original dataset. For each case, we fine-tuned the pretrained double-encoder model (with frozen convolutional layers) and, for comparison, also trained a model from scratch.
3. Final performance was evaluated on the experimental test set (**Figure R8**).

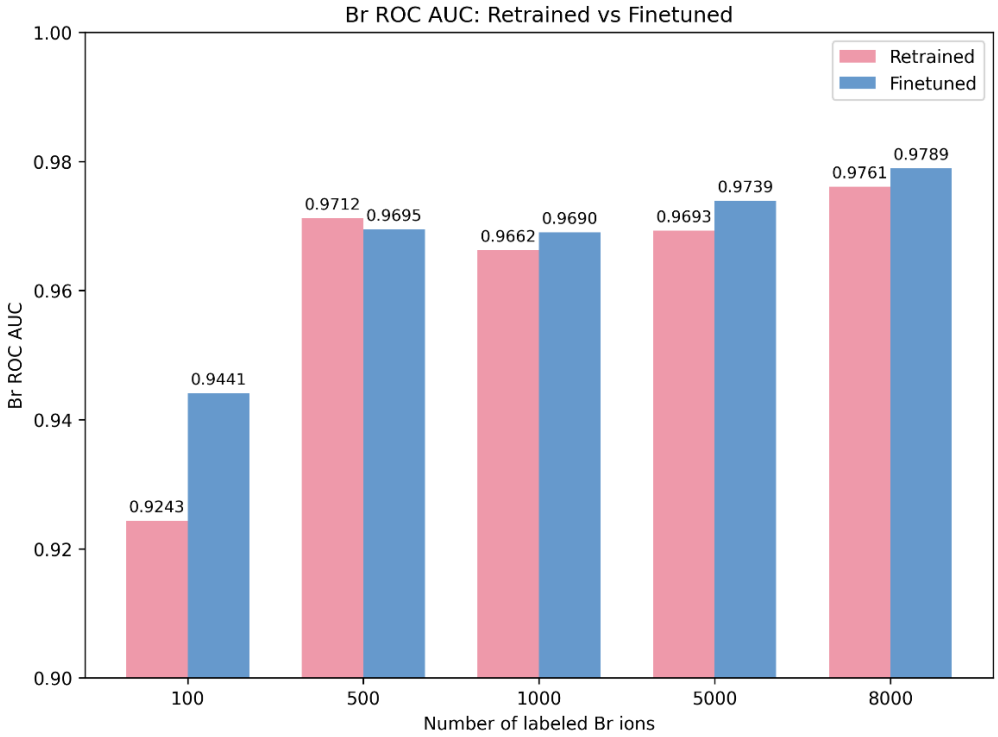

Figure R8. Br test ROC AUC for models trained (or finetuned) with different # of Br-containing ions added to the training dataset (the figure is presented in the Supporting Information as Figure S23).

We also observed that adding a new element generally improves prediction quality for the original elements. For this purpose, we fixed results with 8 000 Br ions and evaluated the results for other target elements (**Table R2**):

Table R2. Test ROC AUC values for target elements obtained using a model trained on data without Br, and models either retrained or finetuned on data with 8,000 Br ions added. The best results for each element are highlighted in bold, and the corresponding rank is shown in parentheses (the table is presented in the Supporting Information as Table S6).

	Without Br	**Retrain**	**Finetune**
Pd	0.9851 (2)	**0.9863 (1)**	**0.9863 (1)**
Ni	0.9923 (2)	0.9902 (3)	**0.9942 (1)**
Cu	0.8881 (3)	**0.9219 (1)**	0.8946 (2)

Ag	**0.9632 (1)**	0.9362 (2)	0.9352 (3)
Cl	0.9653 (3)	**0.9707 (1)**	0.9683 (2)
Average rank	2.2	**1.6**	1.8
Average ROC AUC	0.9588	**0.96106**	0.95572

As **Table R2** shows, for all the cases except silver, the performance increases after Br onboarding.
A detailed description of the experimental procedure is provided in the Supporting Information, section **SI3.8** and in the GitHub repository of our project.

Below, we address your specific questions:

Specifying at minimum the followings would be insightful: 1) the number and variety of labelled spectra required to distinguish a new element,

As shown in the table above, 100 labeled Br distributions were already sufficient to achieve robust performance on the test set. However, the required number of samples depends on the isotopic pattern of the element. Elements with more distinctive isotopic patterns may require fewer samples. Conversely, elements with less informative or more ambiguous isotopic distributions may require more data to reach comparable performance. At its extreme, prediction for monoisotopic elements would purely rely on element co-occurrence in the training dataset and satisfactory performance is unlikely to be obtained.

2) whether the dual-encoder CNN must be retrained end-to-end or whether its layers can be frozen and only the output layer fine-tuned,

Our experiments demonstrate that fine-tuning the model with frozen convolutional layers is not only feasible but also preferable in many cases. Compared with retraining from scratch, fine-tuning requires significantly fewer training steps and achieves higher performance. Notably, we fine-tuned the model using the entire training set, augmented with the new Br-containing distributions, rather than only the new data. Importantly, our double-encoder model has only approximately 76000 parameters, which allows us to train it even on a CPU for 15–20 minutes (100 epochs, ≈50 000 train distributions). Therefore, we can easily both retrain or finetune the model on the whole training set without any computing issues.

3) any failure cases where envelope similarity or charge-state ambiguity hinders classification.

The most obvious failure case is that of monoisotopic elements. Without recognizable isotopic patterns, a neural network will not able to detect such elements. Another challenge is a pair of elements with very similar isotopic distributions. For instance, silver and bromine both have two isotopes with nearly identical mass differences and similar relative intensities. This similarity led to occasional misclassification, where silver-containing ions were incorrectly identified as bromine-containing ions. Notably, the test ROC AUC decreased exactly for silver after Br onboarding. Additional support for this explanation is provided by the dimensionality-reduction visualizations of the model's hidden representations (Manuscript, **Fig. 4**): both the PCA and t-SNE plots show bromine- and silver-related ions forming a shared cluster, indicating that the model encodes these ion types very similarly.

For elements with even greater isotopic overlap, the misclassification rate may increase further.

We hope that this detailed protocol and discussion clarify the scalability and limitations of our approach.

The authors thank the Reviewer for the careful reading and comments, which helped to improve the manuscript!

Reviewer #3

The paper describes a high-quality innovative work implementing the machine learning approaches for facilitating the analysis of the high-resolution mass spectra of liquid samples containing Pd down to ppm/ppb levels. Routine detection of Pd-containing compounds, especially at low concentrations, is crucial for accelerating the progress of chemistry and catalytic science, considering that Pd is widely used in fine chemistry synthesis and the automotive industry. The challenge is increased by the inactivity of Pd in NMR, in contrast to other noble metals with similar applications, such as Pt and Rh, which makes the development of MS-based approaches particularly valuable. Detection of Pd-contamination at ppm/ppb levels is also important to avoid contamination issues while developing more sustainable chemistry routes based on abundant metals such as Ni, Cu, Ag, etc. The work demonstrates the high potential of the novel analytic approach. It also provides the software for the data analysis, which can be potentially used by a wide community of users, including non-experts. All of this makes the results of this work extremely valuable for the wider community of chemists across different fields.

We are grateful to the Reviewer for the evaluation of our manuscript.

The comments below can help the author further improve their manuscript:

1. While the machine learning data analysis part of the work presented in great detail is convincing, the part related to the experimental validation seems too short to appreciate, especially for non-expert users from the neighboring fields. It would be useful to add full experimental details for operando experiments, including the setups and the concentrations of each reagent. It would also be helpful

adding a discussion about the advantages, challenges, and possible limitations of the data analysis in each case.

Thank you for the comment. All experimental details were added to Supporting Information, section **SI5**. Our method indeed has several advantages in terms of experimental data analysis. For instance, Section SI3.8 shows that our approach allows new target elements to be incorporated with minimal fine-tuning (see also “Elements extension” subsection of the manuscript, page 14). In addition, it is sufficiently robust across different mass spectrometers with such analyzers as (FT-ICR, Q-TOF, TOF), it was demonstrated in Section SI4.6. Moreover, it can be successfully applied to Pd detection in the submicromolar concentration range with high precision (see the subsection “Trace contamination. Pd detection in the low-concentration range” on page 17 of the manuscript and Section SI4.3).
However, there are also several challenges and limitations. The isotopic composition of the element is an intrinsic limitation: without considering element cooccurrence monoisotopic elements cannot be used as targets because they do not influence the shape of the isotopic distribution. Resolving power is another limitation: while we obtained very good results for high-resolution spectra, prediction quality may degrade significantly for low-resolution spectra. An additional challenge is the manual tuning of the threshold hyperparameter in the deisotoping algorithm for spectra obtained from different mass spectrometers (see Section SI4.6). In such cases, users may need to adjust this parameter for a new instrument rather than simply loading a spectrum into the pipeline and running the analysis.
Thank you again for your comment. Discussions of the mentioned advantages, limitations, and challenges have been added to the corresponding sections of the manuscript and Supporting Information. Paragraph about limitations has been added to the Conclusions (page 20).

2. Concerning the use of the developed software, it would be helpful to understand if the software works with specific data formats and has any other limitations.

Thank you for raising this important point. To facilitate the use of our software, we have prepared comprehensive instructions covering installation, training data generation, model training, inference, and evaluation. These instructions are available in the README file of our project’s GitHub repository. The software currently supports .mzXML spectra and was tested on high-resolution data; its performance on low-resolution spectra (resolution power < 10 000) may be reduced. Data format requirements and limitations are described in the GitHub repository.

There are also minor corrections and suggestions:

3. In Figure 1b : “Three” should probably be replaced by “Tree”

Thank you for mentioning this. We intended to refer to the number of neural network architectures. We have four architectures: multilayer perceptron, single-encoder CNN, double-channel CNN and double-encoder CNN. We have corrected the figure to state “four” architectures. Thank you again for your careful reading.

4. In Table 1, it would be helpful to explain what AP@k means and also give more details about the other abbreviations not explained in the main text

Thank you for this suggestion. A detailed explanation of AP@k is provided in the Supporting Information, section **SI1.4,** and we now refer to this section in the manuscript. We have also reviewed the main text to ensure that all abbreviations are properly explained.

5. In the main text, Figure S24 is mentioned, but it does not exist in SI. The authors should double check to which figures they refer.

We apologize for this oversight and thank the Reviewer for bringing it to our attention. We have carefully reviewed all figure references in both the manuscript and the supporting information and have corrected all inaccuracies. We are now confident that all figure references are accurate and consistent.

Overall, I recommend a major revision.

The authors thank the Reviewer for the careful reading and comments, which helped to improve the manuscript!