

ROBERT: Bridging the Gap between Machine Learning and Chemistry

David Dalmau, Juan V. Alegre-Requena*

Departamento de Química Inorgánica, Instituto de Síntesis Química y Catálisis Homogénea (ISQCH), CSIC-Universidad de Zaragoza, C/ Pedro Cerbuna 12, 50009 Zaragoza, Spain.

*Correspondence: jv.alegre@csic.es

ABSTRACT

Beyond addressing technological demands, the integration of machine learning (ML) into human societies has also promoted sustainability through the adoption of digitalized protocols. Despite these advantages and the abundance of available toolkits, a substantial implementation gap is preventing the widespread incorporation of ML protocols into the computational and experimental chemistry communities. In this work, we introduce ROBERT, a software carefully crafted to make ML more accessible to chemists of all programming skill levels, while achieving results comparable to those of field experts. We conducted benchmarking using six recent ML studies in chemistry containing 18–4,149 entries. Furthermore, we demonstrated the program's ability to initiate workflows directly from SMILES strings, which simplifies the generation of ML predictors for common chemistry problems. To assess ROBERT's practicality in real-life scenarios, we employed it to discover new luminescent Pd complexes with a modest dataset of 23 points, a frequently encountered scenario in experimental studies.

INTRODUCTION

The world is witnessing a growing interest in applying machine learning (ML) to everyday tasks, primarily driven by substantial savings in time, effort, and resources. Its integration not only fulfills technological needs but also fosters sustainability through the adoption of digitalized procedures, yielding important benefits for a more environmentally conscious future. In particular, the integration of ML in chemistry has opened up new avenues for exploring chemical space and predicting properties of molecules and reaction outcomes. This has led to significant advances in fields such as drug discovery,^{1–5} materials science,^{6–9} chemical synthesis,^{10–17} and catalyst discovery,^{18–24} among others.

Automated ML workflows in chemistry have also become increasingly popular, enabling researchers to predict outcomes efficiently and accurately.^{25–28} Numerous packages and tools are available, including molSimplify,²⁹ pycaret,³⁰ ChemML,³¹ Chemprop,³² DeepChem,³³ Chainer Chemistry,³⁴ Lazy Predict,³⁵ PREFER,³⁶ and PyChemFlow³⁷. However, despite the increasing interest in this field, there is a vast community of chemists that possess little experience in cheminformatics and find these ML tools impossible to include as part of their research.

Another major problem that chemical ML faces is the lack of reproducibility and transparency of the studies in this field.³⁸ Despite prior attempts to establish quality standards, there is an unexpectedly high rate of publications that lack information regarding the database used, the model parameters, or the source code.³⁹ The problem extends further as there are no standard protocols for most ML techniques, causing that multiple groups use different code and workflows to carry out the same tasks.

All these arguments stress that there is a need for a program in chemical ML that produces expert-quality results and grants easy use for researchers independently on their programming experience, while meeting strict reproducibility and transparency standards. In this work, we present ROBERT, a free and open-source software that performs automated ML workflows particularly designed for supervised regression and classification problems in chemistry. ROBERT automates protocols commonly conducted in chemical ML (Figure 1) and generates results directly from a CSV file with a single command, yielding performance comparable to that achieved by experts in the field.

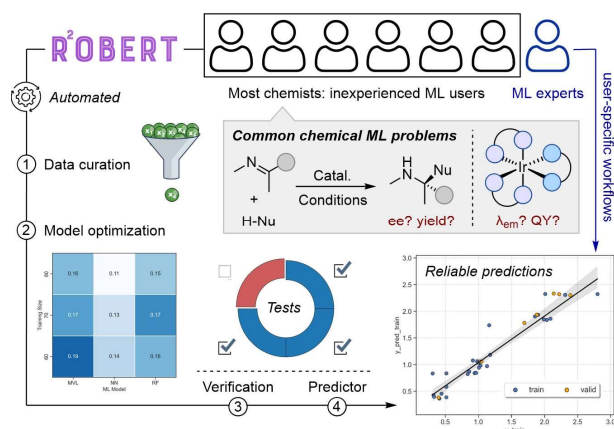


Figure 1. ROBERT automated protocols and two examples of common chemical problems.

RESULTS

Overview of ROBERT

ROBERT is a Python program designed to meet the demands of predictive ML that many members of the computational and experimental chemistry communities have, especially those who lack experience in this field. The software can be used to create predictive models for supervised regression and classification problems related to chemistry, such as experimental yield and selectivity, activation barriers calculated with density functional theory (DFT), and luminescent properties, among others. Using simple databases as starting points, ROBERT allows users with limited or no ML expertise to generate expert-level results within timeframes that allow its integration into modern research workflows.

The installation of this toolkit is performed using conda, taking less than two minutes to set up from scratch on a pool of nine testers with various operating systems and different levels of expertise in Python (Table S1). ROBERT then initiates workflows from CSV databases that contain the target values (y values) along with either X values or SMILES strings of compounds (Figure 2, top). In the second case, the SMILES strings undergo an AQME⁴⁰ workflow to automatically generate atomic and molecular descriptors, which are then used as X values.

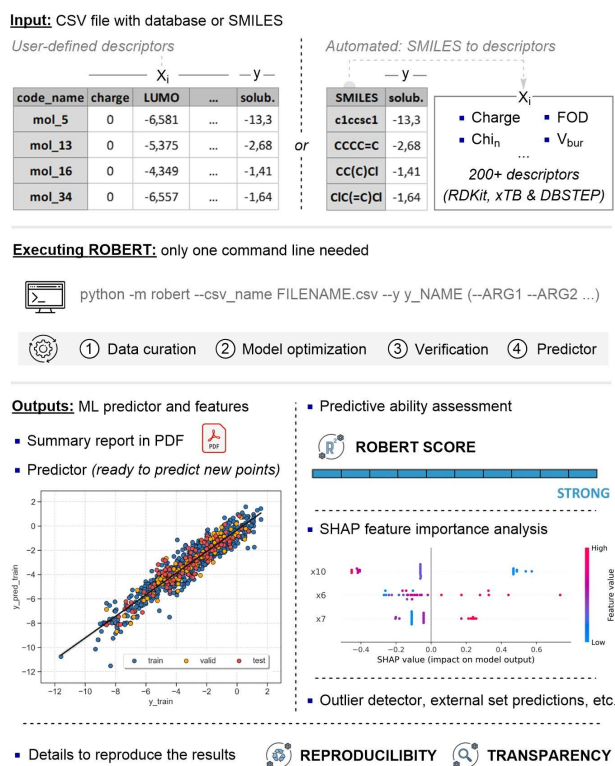


Figure 2. Overview of ROBERT: inputs required, how to run the program, and outputs generated.

The user only needs to execute a single command line to couple multiple protocols, including data curation, scikit-learn-based model screening,^{41,42} assessment of predictive ability, and predictor generation (Figure 2, middle). The resulting models can be readily used to predict the y values of new data points. The program also provides feature importance analyses, enabling users to identify the most influential parameters affecting the target values and facilitating the identification of data trends (Figure 2, bottom). Additionally, the outcomes are supplemented with an outlier detector, allowing chemists to recognize potential inaccuracies associated with data flaws, such as experimental measurement errors or imprecise computational calculations. Different ready-to-use examples are accessible in our online documentation,⁴³ and a comprehensive explanation of the protocols performed in each module is included in the *Robert modules* section of the ESI.

ROBERT incorporates a range of innovative features specifically designed for challenges in chemical ML. For instance, the program generates a comprehensive report in PDF format that assesses the reliability of the resulting models' predictive capabilities, extending beyond conventional metrics such as coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). This is particularly valuable for researchers with limited experience in the field, as these metrics may not accurately reflect the models' predictive performance. The software also performs end-to-end workflows that originate from databases of SMILES strings, including the generation of molecular and atomic descriptors. This functionality offers automated discovery workflows to both experimental and computational chemists directly from routine laboratory data.

This toolkit also seeks to enhance reproducibility and transparency of chemical ML protocols, addressing two long-standing concerns within the scientific community working in this field when publishing in peer-reviewed journals. Multiple authors have emphasized the critical need to establish and enforce standards in this regard³⁸ since, currently, a significant proportion of publications are challenging or even impossible to replicate.³⁹ The generated ROBERT_report.pdf files contain a reproducibility section that guides authors on which files they should upload as supporting information and provides other researchers with the exact programs, versions and commands necessary to replicate the results (Figure 2, bottom). Furthermore, the PDF reports contain comprehensive information regarding the ML models employed and other pertinent details regarding ROBERT protocols, including the data partition scheme and the type of ML problem.

When following the provided instructions, the results starting from databases of descriptors were correctly reproduced by the nine program testers across multiple operating systems.⁴⁴ When repeating end-to-end

workflows starting from SMILES strings, it may not be possible to exactly reproduce the results due to subtle changes in xTB optimized geometries. Nevertheless, the resulting ROBERT scores and model accuracies are very similar between different ROBERT runs (Table S4). In such cases, the PDF report recommends that authors upload the descriptor database created (i.e., AQME-ROBERT_FILENAME.csv) to facilitate the reproduction of results by other researchers.

ROBERT's Performance in Real Examples

We ensured that the program is capable of producing publication-quality results by comparing its outcomes with examples from recent literature in chemical ML, spanning various areas such as catalysis and photophysics. This benchmarking required only a single initial CSV database per example, either downloaded from the manuscripts or obtained directly from the authors. Despite the diversity of conditions included in the benchmarking, the command line used in ROBERT remained identical in all cases (i.e., `python -m robert --y VALUE --csv_name FILENAME [--ARG]`, see section *Benchmarking with Six Examples* in the ESI). To increase the complexity of the challenge, the examples included models designed by research groups that are highly respected in the field, such as the Liu,⁴⁵ Doyle,⁴⁶ Sunoj,⁴⁷ Kulik,⁴⁸ Aspuru-Guzik and Balcells,⁴⁹ and Ess⁵⁰ groups. We also introduced other parameters to make the benchmarking more demanding, including a wide and realistic range of data points spanning more than two orders of magnitude (from 18 to 4149 data points), the presence of different ML algorithms, and the inclusion of both regression and classification supervised problems.

The results demonstrated that ROBERT performs similarly to experts in the field in three cases (−1.9%, +0.9% and +0.3% scaled error for examples C, D and E, respectively), while achieving lower errors in the other instances (−3.3%, −7.0% and −4.1% for examples A, B and F, respectively) (Figure 3 and Table S5). Furthermore, the exhaustive algorithm screening conducted by ROBERT led to models that employed similar training sizes ($\pm 10\%$ of the data) in five of the examples (B-F). Another advantage of the program is its permutation feature importance (PFI) filter, which generates simpler and more interpretable protocols. In five examples, ROBERT models contained fewer descriptors than the original examples, with some noteworthy reductions, such as in examples A (from three to one descriptor), C (from 101 to two), and E (from 135 to 32). Despite the various protocols integrated into the workflows, the execution times on eight processors ranged

from less than a minute to a few hours (Figure 3). These execution times make it feasible to run typical ROBERT processes on personal computers.

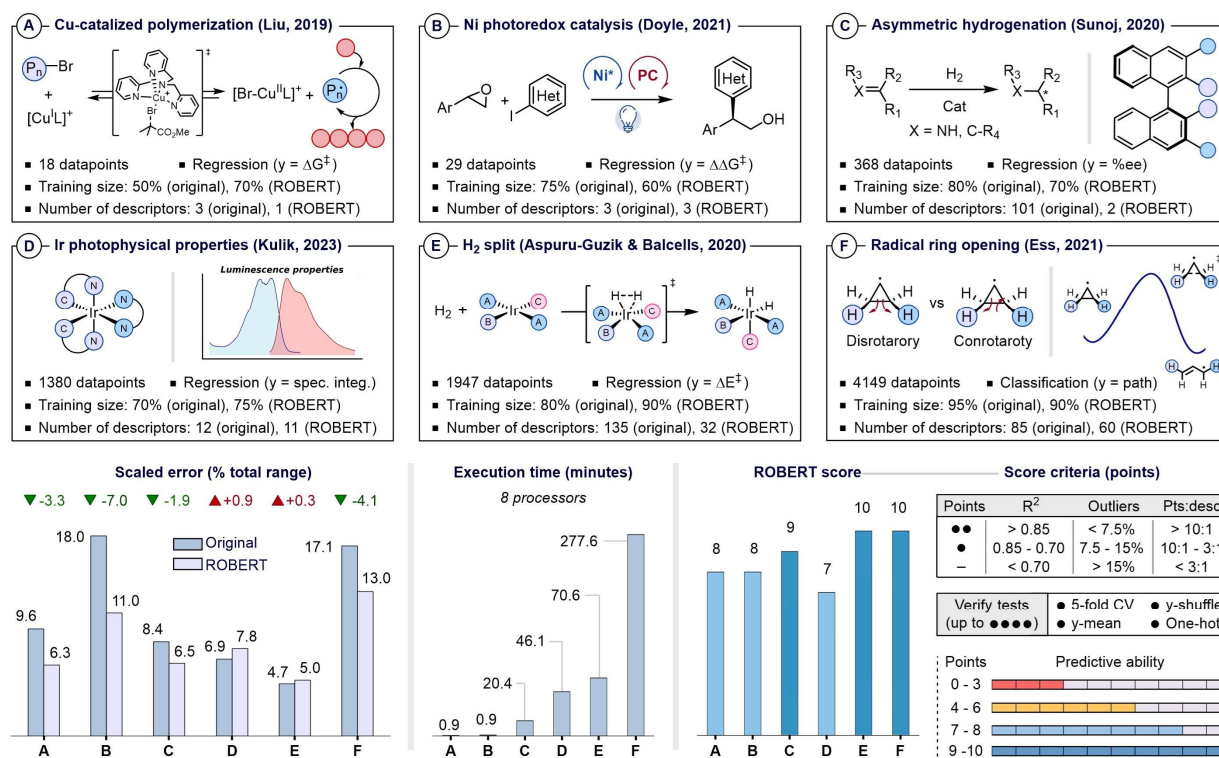


Figure 3. Benchmarking of scaled errors (RMSE or MAE, see the ESI for more details) obtained with ROBERT against six relevant examples from the chemical ML literature, along with the execution time required when using eight processors. Furthermore, the criteria used in ROBERT scores and the scores obtained for the six examples are also provided.

Next, we pondered whether conventional metrics employed for evaluating ML algorithms, such as R², MAE, and RMSE, sufficed for creating a versatile, general-purpose tool. While experienced ML users can assess whether an ML workflow has yielded favorable results for valid reasons, inexperienced users may encounter difficulties in gauging the reliability of its predictive proficiency. It is widely recognized that ML algorithms can exhibit good metrics in the validation set while harboring questionable predictive ability. For example, a user might assume that a model with good metrics is proficient at prediction, but that very model could yield comparably low errors when the y values are shuffled⁵¹ or when using random numbers as descriptors.⁵²

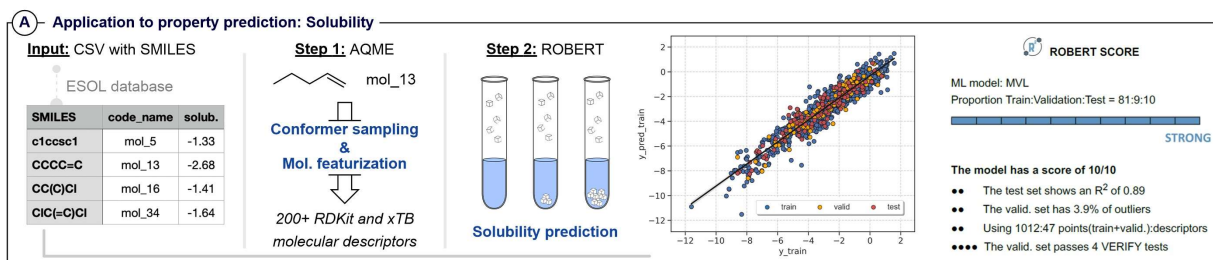
For this reason, we have included a section in the PDF files generated by the program that contains ROBERT scores. The ROBERT score is a rating out of ten, designed to offer users insights into the predictive capabilities of the protocols selected by ROBERT. We attempted to rank the models following guidelines in line with modern

ML research³⁸ which encompass factors such as the correlation between predictions and measurements, human interpretability, error distribution, sensitivity to features, and avoidance of overfitting and underfitting (Figure 3 and section *ROBERT Score* of the ESI). Across the six examples examined, the scores ranged from seven to ten, indicating varying degrees of predictive ability from moderate to strong.

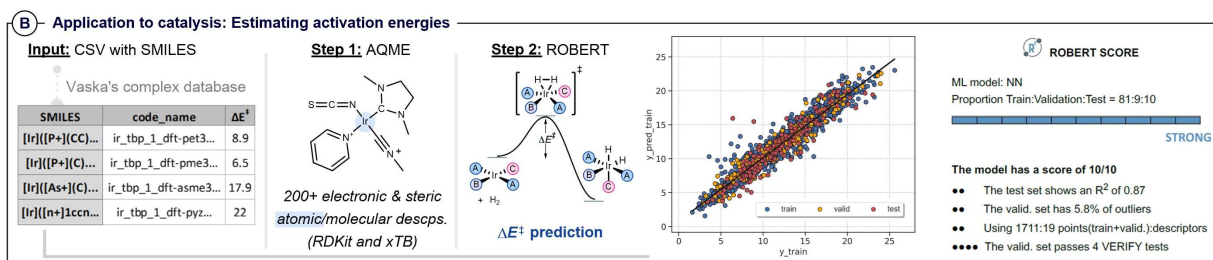
Workflows from SMILES Strings

After assessing ROBERT's capability to generate results of a quality comparable to experts in chemical ML, we aimed to advance the program's implementation in real-life scenarios. One significant challenge that chemists often face when integrating ML into their work routines is the limited experience in creating useful databases. This is important because choosing the right descriptors is much more crucial than including many irrelevant ones in the database. ROBERT introduces an innovative workflow enabling users to generate meaningful descriptors from SMILES strings.⁵³ This 1D representation type has gained popularity in the field of chemical ML and is readily available in most chemical databases or easily generated using chemical drawing software such as ChemDraw.⁵⁴ Using AQME workflows,⁴⁰ the program can produce over 200 Boltzmann-weighted descriptors with methods significantly faster than DFT. These workflows include RDKit⁵⁵ conformer samplings and xTB⁵⁶ geometry optimizations, followed by the generation of descriptors with RDKit, xTB and DBSTEP.⁵⁷ The resulting descriptors encompass electronic and steric properties and can be calculated for the entire molecule and for specific relevant atoms. After the PFI filter, only the most influencing descriptors remain.

We tested the SMILES-to-predictions workflow in two examples from fields that are particularly interesting for ML predictions: property calculation and catalysis. First, we employed ROBERT to create a predictor for the solubility of organic molecules in water. To do this, we accessed the 1124 entries of the ESOL database, which is available in the supporting information of the corresponding publication,⁵⁸ and executed the command line shown in Figure 4A to generate molecular descriptors. Using eight processors, the predictor was ready in approximately three hours. The resulting multivariate lineal (MVL) model used 80% training size and 47 descriptors, and can accurately predict the solubility of new organic molecules (ROBERT score of 10). Furthermore, the generated PDF report contains valuable information that can aid in the rational design of solutes. This includes feature importance analysis to understand the key molecular properties influencing solubility and outlier analysis to identify potential measurement errors, among other parameters.



Only one command line: `python -m robert --csv_name solubility.csv --y solubility --aqme --qdesc_keywords "--qdesc_solvent h2o"`



Only one command line: `python -m robert --csv_name vaskas.csv --y barrier --aqme --qdesc_keywords "--qdesc_atoms [Ir]"`

Figure 4. SMILES workflows used in aqueous solubility (A) and activation energy (B) predictions.

Subsequently, we employed ROBERT to predict the activation barriers of hydrogenations using different Vaska's catalysts (example E from Figure 3).⁴⁹ In this case, the database contained 1901 SMILES strings of the catalyst-substrate complexes displayed in Figure 4B. We executed the command line depicted in the figure using 16 processors and the predictor was generated within nine hours. In this case, the optimal model was a neural network (NN) that used 80% training size and 19 descriptors. This algorithm exhibited robust predictive performance (ROBERT score of 10) and is ready for inclusion within a catalyst discovery workflow, where users can input SMILES representation of new catalyst candidates. It's noteworthy that in this example, ROBERT incorporated atomic properties of the Ir metal center. This feature enables the addition of relevant local electronic and steric environments alongside the default molecular properties.

Overall, these findings indicate that employing non-DFT descriptors can yield highly predictive models while utilizing a reasonable number of descriptors (47 and 19 descriptors for 1124 and 1901 entries, respectively). The use of these cost-effective descriptors enables ROBERT to generate predictors in timeframes that result suitable for integrating ML protocols into real-life research workflows. Even in the two examples involving over 1,000 SMILES entries, the predictors were generated in just three and nine hours.

Data-driven Discovery of New Pd Luminescent Probes

We then decided to focus on a specific research line within our group, which involves the discovery of Pd-based emitters.⁵⁹ The synthesis and design of luminescent compounds have experienced significant growth and recognition in recent years, with applications spanning diverse areas such as sensors, catalysts, biomarkers and light-emitting devices.⁶⁰ However, Pd exhibits poor photophysical properties compared to other transition metals due to readily available non-radiative relaxation pathways. In our previous studies,^{59,61–63} we were only able to synthesize a very limited number of oxazolone-Pd probes with good luminescence using our chemical intuition, as there were no clear trends in the results. Given that only five structural modifications resulted in quantum yields (QYs) above 10% (**1a-e**, Figure 5A), our objective was to employ ROBERT for the identification of new Pd complexes with QY values exceeding this threshold.

First, we compiled a dataset of molecules in ChemDraw containing all the oxazolone-Pd complexes reported in our previous studies, albeit this database was still limited to 20 entries. Next, we generated SMILES strings of the compounds within minutes and executed ROBERT using one command line. The automated workflow generated molecular descriptors, as well as atomic features for the Pd centers. One of ROBERT's unique procedures is to include an evaluation of the predictive capability of the models at the beginning of the PDF report, quantified as the ROBERT score. With the initial 20 data points, the model exhibited weak performance, primarily attributed to an imbalance resulting from the limited number of data points in the region where the QY exceeded 10% (ROBERT score 5, Figure S9, left).

To address this issue, we decided to compensate the dataset with three additional molecules that were chosen for their rapid synthetic availability, with logical human decisions guiding the selection process (Figure S9, middle). For instance, we anticipated that adding a Me group in the *para* position of a Py ligand should maintain the favorable QY observed in points with more than 10% QY. Subsequently, we reran ROBERT with the expanded database of 23 points and observed a significant improvement in the ROBERT score, reaching 7, which represents a decent level of predictive ability for that dataset size (Figure 5B, top, and Figure S9, right).

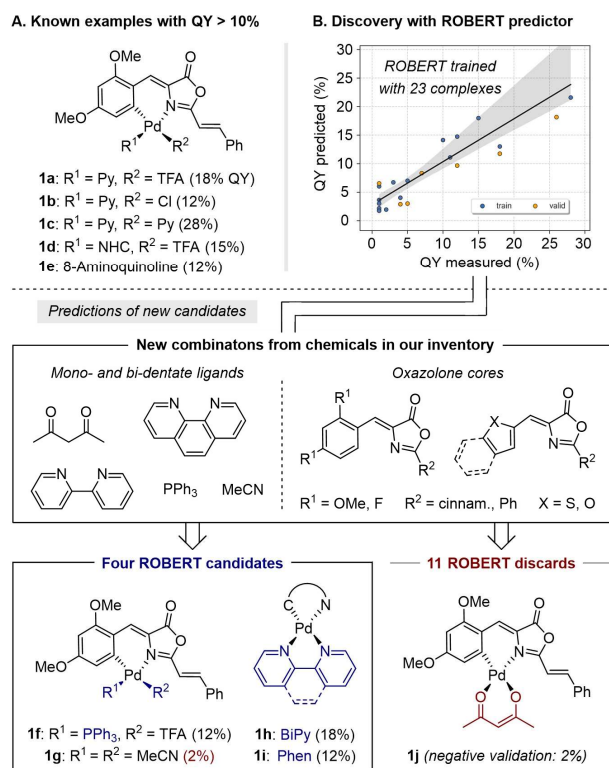


Figure 5. (A) Previously reported examples of oxazolone-Pd complexes with QY higher than 10%. (B) Top: results of the ML predictor model. Middle: ligands and oxazolone cores used in the screening. Bottom: experimental validation of predictions. NHC = 1,3-dimesityl-1H-imidazol-3-ium-2-ide. TFA = trifluoroacetate.

We then initiated the data-driven discovery of Pd probes, where the candidates featured structural modifications whose effects could not be reliably determined using human chemical intuition. The modifications were driven by the availability of ligands and oxazolone cores in our laboratory (Figure 5B, middle), which is arguably one of the most common practices used to expand substrate scopes in experimental chemistry studies. This combinatorial process generated 15 candidates with entirely unknown QYs (Figure S8). We used the SMILES strings of these new 15 compounds to generate descriptors and predict QY, resulting in four predictions (**1f-i**) with QY values exceeding 10%. Subsequently, we conducted experimental measurements of the QY for these positively predicted candidates, and also assessed the QY of a negative result (**1j**) for further experimental validation. Despite the limited number of data points in the initial training database, the predictor showed promising results, with four out of the five measured QYs yielding the expected range (80% accuracy, Figure 5B, bottom).

This digitalized approach enabled us to discover three new luminescent emitters with QYs exceeding our 10% threshold (**1f**, **1h** and **1i**) from a pool of over 15 potential candidates while conducting a minimal number of experiments. Finding three new probes marks a significant progress in the field of oxazolone-Pd-based emitters, especially considering that only five such examples were previously known. These positive outcomes highlight the potential of ROBERT to facilitate data-driven approaches, even in situations with limited available data.

CONCLUSION

ROBERT has been designed to make ML more accessible to chemists, irrespective of their programming expertise, while producing results comparable to those achieved by experts in the field with just one command line. Another primary objective of ROBERT is to address the long-standing issues of reproducibility and transparency that have plagued chemical ML. Upon executing the program, users receive a comprehensive PDF report evaluating the reliability of the generated models, accompanied by a step-by-step guide for replicating the outcomes.

To validate the effectiveness of ROBERT, we conducted benchmarking using six recent studies in chemical ML. These examples spanned a wide spectrum of databases, from modest datasets with just 18 entries to extensive ones including more than 4,000 entries. The results demonstrated ROBERT's capacity to produce results on par with domain experts, all with the simplicity of a single command line.

Furthermore, this software can initiate workflows directly from SMILES strings, bypassing complications tied to database generation and simplifying the generation of ML predictors for chemistry problems. We illustrated how databases containing only SMILES representations and target outcomes can produce precise predictions for properties and reactivity.

Lastly, we sought to evaluate ROBERT's practicality in real-life scenarios. To this end, we employed the program for discovering new luminescent Pd complexes. Starting with a rather limited dataset of only 23 data points, a situation frequently encountered in experimental studies, ROBERT demonstrated its utility by proposing several luminescent emitters from a pool of potential candidates.

DATA AND CODE AVAILABILITY

The ROBERT program is freely available on GitHub (<https://github.com/jvalegre/robert>), along with comprehensive documentation provided on Read the Docs (<https://robert.readthedocs.io>) and hands-on tutorials on YouTube (<https://www.youtube.com/@thealegregroup4964>). All the ROBERT PDF reports and CSV databases used in the examples from the paper are provided sorted in folders along with the ESI. These PDF reports encompass detailed instructions for replicating the results (Reproducibility section), information about the models (Transparency section), ROBERT scores, R², MAE and RMSE metrics of all datasets (ROBERT score section), feature importance, outlier analysis, and other relevant data (Predict section). Additionally, all the raw data containing the outputs of the ROBERT jobs can be accessed on Zenodo (DOI: 10.5281/zenodo.10057934).

ACKNOWLEDGEMENTS

The acronym ROBERT is dedicated to Prof. ROBERT Paton, who was a mentor to J.V.A.R. throughout his years at Colorado State University and who introduced him to the field of cheminformatics. D.D. thanks Oliver Lee (University of St Andrews) for assisting with the style of the ROBERT_report.pdf file. D.D. and J.V.A.R. thank the seven program testers who meticulously evaluated the installation, usage, and reproducibility of the program (listed in chronological order): Prof. David Valiente (Miguel Hernández University), Íñigo Iribarren (Trinity College Dublin), Dr. Heidi Klem (Colorado State University), Dr. Guilian Luchini (Bristol Myers Squibb), Alex Platt (Colorado State University), Xinchun Ran (Vanderbilt University), and Oliver Lee (University of St. Andrews). The authors also thank Ignacio Funes-Ardoiz (University of La Rioja) and Dr. Esteban Urriolabeitia (CSIC) for pre-reviewing the manuscript.

FUNDING

J.V.A.-R. and D.D. acknowledge Gobierno de Aragón-Fondo Social Europeo (Research Groups E07_23R and E17_23R) and the State Research Agency of Spain (MCIN/ AEI/ 10.13039/501100011033/ FEDER, UE) for financial support (IJC2020-044217-I, PID2022-140159NA-I00 and PID2019-106394GB-I00). J.V.A.-R. and D.D. acknowledge the computing resources at the Galicia Supercomputing Center, CESGA, including access to the

FinisTerae supercomputer, the Red Española de Supercomputación (grant number QH-2023-2-0003) and the Drago cluster facility of SGAI-CSIC.

REFERENCES

1. Klambauer, G., Hochreiter, S., and Rarey, M. (2019). Machine Learning in Drug Discovery. *J. Chem. Inf. Model.* *59*, 945–946. 10.1021/acs.jcim.9b00136.
2. Batra, K., Zorn, K.M., Foil, D.H., Minerali, E., Gawriljuk, V.O., Lane, T.R., and Ekins, S. (2021). Quantum Machine Learning Algorithms for Drug Discovery Applications. *J. Chem. Inf. Model.* *61*, 2641–2647. 10.1021/acs.jcim.1c00166.
3. Zhang, Y., Wang, Y., Zhou, W., Fan, Y., Zhao, J., Zhu, L., Lu, S., Lu, T., Chen, Y., and Liu, H. (2019). A combined drug discovery strategy based on machine learning and molecular docking. *Chem. Biol. Drug. Des.* *93*, 685–699. 10.1111/cbdd.13494.
4. Courtois, É., Nguyen, T.T.H., Fournier, A., Carcaillon-Bentata, L., Moutengou, É., Escolano, S., Tubert-Bitter, P., Elbaz, A., Thiébaud, A.C.M., and Ahmed, I. (2022). Identifying Protective Drugs for Parkinson's Disease in Health-Care Databases Using Machine Learning. *Mov. Disord.* *37*, 2376–2385. 10.1002/mds.29205.
5. Peña-Guerrero, J., Nguewa, P.A., and García-Sosa, A.T. (2021). Machine learning, artificial intelligence, and data science breaking into drug design and neglected diseases. *WIREs Comput. Mol. Sci.* *11*, e1513. 10.1002/wcms.1513.
6. Zhong, X., Gallagher, B., Liu, S., Kailkhura, B., Hiszpanski, A., and Han, T.Y.-J. (2022). Explainable machine learning in materials science. *npj Comput. Mater.* *8*, 204. 10.1038/s41524-022-00884-7.
7. Karande, P., Gallagher, B., and Han, T.Y.-J. (2022). A Strategic Approach to Machine Learning for Material Science: How to Tackle Real-World Challenges and Avoid Pitfalls. *Chem. Mater.* *34*, 7650–7665. 10.1021/acs.chemmater.2c01333.

8. Collins, C.M., Daniels, L.M., Gibson, Q., Gaultois, M.W., Moran, M., Feetham, R., Pitcher, M.J., Dyer, M.S., Delacotte, C., Zanella, M., et al. (2021). Discovery of a Low Thermal Conductivity Oxide Guided by Probe Structure Prediction and Machine Learning. *Angew. Chem. Int. Ed.* **60**, 16457–16465. 10.1002/anie.202102073.
9. Merte, L.R., Bisbo, M.K., Sokolović, I., Setvín, M., Hagman, B., Shipilin, M., Schmid, M., Diebold, U., Lundgren, E., and Hammer, B. (2022). Structure of an Ultrathin Oxide on Pt₃Sn(111) Solved by Machine Learning Enhanced Global Optimization. *Angew. Chem. Int. Ed.* **61**, e202204244. 10.1002/anie.202204244.
10. Zhang, Z., Jiao, Z., Shen, R., Song, P., and Wang, Q. (2023). Accelerated Design of Flame Retardant Polymeric Nanocomposites via Machine Learning Prediction. *ACS Appl. Eng. Mater.* **1**, 596–605. 10.1021/acsaenm.2c00145.
11. Zhang, S., Xu, L., Li, S., Oliveira, J.C.A., Li, X., Ackermann, L., and Hong, X. (2023). Bridging Chemical Knowledge and Machine Learning for Performance Prediction of Organic Synthesis. *Chem. Eur. J.* **29**, e202202834. 10.1002/chem.202202834.
12. Chen, J., Luo, J.B., Hu, M.Y., Zhou, J., Huang, C.Z., and Liu, H. (2023). Controlled Synthesis of Multicolor Carbon Dots Assisted by Machine Learning. *Adv. Funct. Mater.* **33**, 2210095. 10.1002/adfm.202210095.
13. Xie, Y., Zhang, C., Hu, X., Zhang, C., Kelley, S.P., Atwood, J.L., and Lin, J. (2020). Machine Learning Assisted Synthesis of Metal–Organic Nanocapsules. *J. Am. Chem. Soc.* **142**, 1475–1481. 10.1021/jacs.9b11569.
14. Beckham, J.L., Wyss, K.M., Xie, Y., McHugh, E.A., Li, J.T., Advincula, P.A., Chen, W., Lin, J., and Tour, J.M. (2022). Machine Learning Guided Synthesis of Flash Graphene. *Adv. Mater.* **34**, 2106506. 10.1002/adma.202106506.
15. Luo, Y., Bag, S., Zaremba, O., Cierpka, A., Andreo, J., Wuttke, S., Friederich, P., and Tsotsalas, M. (2022). MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning. *Angew. Chem. Int. Ed.* **61**, e202200242. 10.1002/anie.202200242.

16. Pflüger, P.M., and Glorius, F. (2020). Molecular Machine Learning: The Future of Synthetic Chemistry? *Angew. Chem. Int. Ed.* *59*, 18860–18865. 10.1002/anie.202008366.
17. Oliveira, J.C.A., Frey, J., Zhang, S.-Q., Xu, L.-C., Li, X., Li, S.-W., Hong, X., and Ackermann, L. (2022). When machine learning meets molecular synthesis. *Trends Chem.* *4*, 863–885. 10.1016/j.trechm.2022.07.005.
18. Kolluru, A., Shuaibi, M., Palizhati, A., Shoghi, N., Das, A., Wood, B., Zitnick, C.L., Kitchin, J.R., and Ulissi, Z.W. (2022). Open Challenges in Developing Generalizable Large-Scale Machine-Learning Models for Catalyst Discovery. *ACS Catal.* *12*, 8572–8581. 10.1021/acscatal.2c02291.
19. Dotson, J.J., van Dijk, L., Timmerman, J.C., Grosslight, S., Walroth, R.C., Gosselin, F., Püntener, K., Mack, K.A., and Sigman, M.S. (2023). Data-Driven Multi-Objective Optimization Tactics for Catalytic Asymmetric Reactions Using Bisphosphine Ligands. *J. Am. Chem. Soc.* *145*, 110–121. 10.1021/jacs.2c08513.
20. Williams, T., McCullough, K., and Lauterbach, J.A. (2020). Enabling Catalyst Discovery through Machine Learning and High-Throughput Experimentation. *Chem. Mater.* *32*, 157–165. 10.1021/acs.chemmater.9b03043.
21. McCullough, K., Williams, T., Mingle, K., Jamshidi, P., and Lauterbach, J. (2020). High-throughput experimentation meets artificial intelligence: a new pathway to catalyst discovery. *Phys. Chem. Chem. Phys.* *22*, 11174–11196. 10.1039/D0CP00972E.
22. Gallegos, L.C., Luchini, G., St. John, P.C., Kim, S., and Paton, R.S. (2021). Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* *54*, 827–836. 10.1021/acs.accounts.0c00745.
23. Boni, Y.T., Cammarota, R.C., Liao, K., Sigman, M.S., and Davies, H.M.L. (2022). Leveraging Regio- and Stereoselective C(sp³)-H Functionalization of Silyl Ethers to Train a Logistic Regression Classification Model for Predicting Site-Selectivity Bias. *J. Am. Chem. Soc.* *144*, 15549–15561. 10.1021/jacs.2c04383.
24. Masood, H., Toe, C.Y., Teoh, W.Y., Sethu, V., and Amal, R. (2019). Machine Learning for Accelerated Discovery of Solar Photocatalysts. *ACS Catal.* *9*, 11774–11787. 10.1021/acscatal.9b02531.

25. Schleder, G.R., Acosta, C.M., and Fazzio, A. (2020). Exploring Two-Dimensional Materials Thermodynamic Stability via Machine Learning. *ACS Appl. Mater. Interf.* *12*, 20149–20157. 10.1021/acsami.9b14530.
26. Yang, Q., Li, Y., Yang, J., Liu, Y., Zhang, L., Luo, S., and Cheng, J. (2020). Holistic Prediction of the pK_a in Diverse Solvents Based on a Machine-Learning Approach. *Angew. Chem. Int. Ed.* *59*, 19282–19291. 10.1002/anie.202008528.
27. Panapitiya, G., Avendaño-Franco, G., Ren, P., Wen, X., Li, Y., and Lewis, J.P. (2018). Machine-Learning Prediction of CO Adsorption in Thiolated, Ag-Alloyed Au Nanoclusters. *J. Am. Chem. Soc.* *140*, 17508–17514. 10.1021/jacs.8b08800.
28. Evans, J.D., and Coudert, F.-X. (2017). Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning. *Chem. Mater.* *29*, 7833–7839. 10.1021/acs.chemmater.7b02532.
29. Ioannidis, E.I., Gani, T.Z.H., and Kulik, H.J. (2016). molSimplify: A toolkit for automating discovery in inorganic chemistry. *J. Comput. Chem.* *37*, 2106–2117. 10.1002/jcc.24437.
30. Moez, A. (2020). PyCaret.
31. Haghightalari, M., Vishwakarma, G., Altarawy, D., Subramanian, R., Kota, B.U., Sonpal, A., Setlur, S., and Hachmann, J. (2020). *ChemML*: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data. *WIREs Comput. Mol. Sci.* *10*, e1458. 10.1002/wcms.1458.
32. Heid, E., Greenman, K.P., Chung, Y., Li, S.-C., Graff, D.E., Vermeire, F.H., Wu, H., Green, W.H., and McGill, C.J. (2023). Chemprop: Machine Learning Package for Chemical Property Prediction. Preprint at ChemRxiv, 10.26434/chemrxiv-2023-3zcf1.
33. Ramsundar, B., Eastman, P., Walters, P., and Pande, V. <https://github.com/deepchem/deepchem>.
34. Tokui, S., Oono, K., Hido, S., and Clayton, J. (2015) Chainer: a next-generation open source framework for deep learning. In Proceedings of Workshop on Machine Learning Systems (LearningSys) in Advances in Neural Information Processing System (NIPS) 28. <https://github.com/chainer/chainer-chemistry>
35. Pandala, S.R. Lazy Predict. <https://github.com/shankarpandala/lazypredict>

36. Lanini, J., Santarossa, G., Sirockin, F., Lewis, R., Fechner, N., Misztela, H., Lewis, S., Maziarz, K., Stanley, M., Segler, M., et al. (2023). PREFER: A New Predictive Modeling Framework for Molecular Discovery. *J. Chem. Inf. Model.* *63*, 4497–4504. 10.1021/acs.jcim.3c00523.
37. Lovrić, M., Đuričić, T., Hussain, H., Lučić, B., and Kern, R. (2023). PyChemFlow: an automated pre-processing pipeline in Python for reproducible machine learning on chemical data. Preprint at ChemRxiv, 10.26434/chemrxiv-2023-3zpw0.
38. Artrith, N., Butler, K.T., Coudert, F.-X., Han, S., Isayev, O., Jain, A., and Walsh, A. (2021). Best practices in machine learning for chemistry. *Nat. Chem.* *13*, 505–508. 10.1038/s41557-021-00716-z.
39. Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science* *359*, 725–726. 10.1126/science.359.6377.725.
40. Alegre-Requena, J.V., Sowndarya S. V., S., Pérez-Soto, R., Alturaifi, T.M., and Paton, R.S. (2023). AQME: Automated quantum mechanical environments for researchers and educators. *WIREs Comput. Mol. Sci.* *13*, e1663. 10.1002/wcms.1663.
41. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* *12*, 2825–2830.
42. Intel(R). <https://github.com/intel/scikit-learn-intelex>
43. <https://robert.readthedocs.io>
44. Some operating systems may not be compatible with the code accelerator scikit-learn-intelex. In these cases, ROBERT runs with the standard scikit-learn code. Although we did not observe exactly identical results between scikit-learn-intelex and scikit-learn-based protocols, ROBERT produced results with very similar precisions in both protocols (Table S2). A reproducibility warning with this information is printed in the generated PDF reports.
45. Fang, C., Fantin, M., Pan, X., De Fiebre, K., Coote, M.L., Matyjaszewski, K., and Liu, P. (2019). Mechanistically Guided Predictive Models for Ligand and Initiator Effects in Copper-Catalyzed Atom Transfer Radical Polymerization (Cu-ATRP). *J. Am. Chem. Soc.* *141*, 7486–7497. 10.1021/jacs.9b02158.

46. Lau, S.H., Borden, M.A., Steiman, T.J., Wang, L.S., Parasram, M., and Doyle, A.G. (2021). Ni/Photoredox-Catalyzed Enantioselective Cross-Electrophile Coupling of Styrene Oxides with Aryl Iodides. *J. Am. Chem. Soc.* **143**, 15873–15881. 10.1021/jacs.1c08105.
47. Singh, S., Pareek, M., Changotra, A., Banerjee, S., Bhaskararao, B., Balamurugan, P., and Sunoj, R.B. (2020). A unified machine-learning protocol for asymmetric catalysis as a proof of concept demonstration using asymmetric hydrogenation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1339–1345. 10.1073/pnas.1916392117.
48. Terrones, G.G., Duan, C., Nandy, A., and Kulik, H.J. (2023). Low-cost machine learning prediction of excited state properties of iridium-centered phosphors. *Chem. Sci.* **14**, 1419–1433. 10.1039/D2SC06150C.
49. Friederich, P., Dos Passos Gomes, G., De Bin, R., Aspuru-Guzik, A., and Balcells, D. (2020). Machine learning dihydrogen activation in the chemical space surrounding Vaska's complex. *Chem. Sci.* **11**, 4584–4601. 10.1039/D0SC00445F.
50. Maley, S.M., Melville, J., Yu, S., Teynor, M.S., Carlsen, R., Hargis, C., Hamilton, R.S., Grant, B.O., and Ess, D.H. (2021). Machine learning classification of disrotatory IRC and conrotatory non-IRC trajectory motion for cyclopropyl radical ring opening. *Phys. Chem. Chem. Phys.* **23**, 12309–12320. 10.1039/D1CP00612F.
51. Rücker, C., Rücker, G., and Meringer, M. (2007). γ -Randomization and Its Variants in QSPR/QSAR. *J. Chem. Inf. Model.* **47**, 2345–2357. 10.1021/ci700157b.
52. Chuang, K.V., and Keiser, M.J. (2018). Comment on "Predicting reaction performance in C–N cross-coupling using machine learning." *Science* **362**, eaat8603. 10.1126/science.aat8603.
53. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36. 10.1021/ci00057a005.
54. Evans, D.A. (2014). History of the Harvard ChemDraw Project. *Angew. Chem. Int. Ed.* **53**, 11140–11145. 10.1002/anie.201405820.
55. Landrum, G. (2010). RDKit: Open-source cheminformatics. <https://www.rdkit.org>

56. Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., and Grimme, S. (2021). Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* *11*, e1493. 10.1002/wcms.1493.
57. Luchini, G., Patterson, T., and Paton, R. (2023). DBSTEP. Version 1.1.0. Software at Zenodo, 10.5281/ZENODO.4702097.
58. Delaney, J.S. (2004). ESOL: Estimating Aqueous Solubility Directly from Molecular Structure. *J. Chem. Inf. Comput. Sci.* *44*, 1000–1005. 10.1021/ci034243x.
59. Dalmau, D., Crespo, O., Matxain, J.M., and Urriolabeitia, E.P. (2023). Fluorescence Amplification of Unsaturated Oxazolones Using Palladium: Photophysical and Computational Studies. *Inorg. Chem.* *62*, 9792–9806. 10.1021/acs.inorgchem.3c00601.
60. Dalmau, D., and Urriolabeitia, E.P. (2023). Luminescence and Palladium: The Odd Couple. *Molecules* *28*, 2663. 10.3390/molecules28062663.
61. Garcia-Sanz, C., Andreu, A., de las Rivas, B., Jimenez, A.I., Pop, A., Silvestru, C., Urriolabeitia, E.P., and Palomo, J.M. (2021). Pd-Oxazolone complexes conjugated to an engineered enzyme: improving fluorescence and catalytic properties. *Org. Biomol. Chem.* *19*, 2773–2783. 10.1039/D1OB00305D.
62. Laga, E., Dalmau, D., Arregui, S., Crespo, O., Jimenez, A.I., Pop, A., Silvestru, C., and Urriolabeitia, E.P. (2021). Fluorescent Orthopalladated Complexes of 4-Aryliden-5(4H)-oxazolones from the Kaede Protein: Synthesis and Characterization. *Molecules* *26*, 1238. 10.3390/molecules26051238.
63. Collado, S., Pueyo, A., Baudequin, C., Bischoff, L., Jiménez, A.I., Cativiela, C., Hoarau, C., and Urriolabeitia, E.P. (2018). Orthopalladation of GFP-Like Fluorophores Through C-H Bond Activation: Scope and Photophysical Properties. *Eur. J. Org. Chem.* *2018*, 6158–6166. 10.1002/ejoc.201800966.