SLAS Technology 36 (2026) 100374

Contents lists available at ScienceDirect

SLAS Technology

journal homepage: www.elsevier.com/locate/slast

2nd EUOS/SLAS joint challenge: Prediction of spectral properties of compounds

Innovation in automation-based drug discovery efforts is increasingly dependent on advancements in data science and the ability to interpret and exploit the extensive datasets generated as part of large scale compound screening studies. With this in mind, we here announce a machine-learning challenge based upon the compound bioprofiling efforts led by the EU-OPENSCREEN research infrastructure. The challenge involves a dataset of approximately 100,000 compounds, recently tested under the EU-OPENSCREEN bioprofiling program. The dataset to be used in the challenge represents a significant global resource, with very few comparable datasets available at this scale. Measurements on these data include compound solubility, redox properties, as well as fluorescence and absorbance characteristics. The solubility data sets have already been used in the recent Kaggle Solubility prediction Challenge [1,2] In this new challenge, we address the prediction of both fluorescence and absorbance/transmittance properties based on chemical structure properties [3,4]

The participants are asked to develop predictive models with two goals:

1) Predict whether compounds have an average optical transmittance below 70 % over two wavelength ranges: i) 340-400 nm and ii) 450-679 nm.
2) Predict whether compounds are fluorescent at key excitation/emission wavelength pairings relevant to modern bioassay readouts i.e. 340/450 nm, 480/540 nm, 525/598 nm, and 560/610 nm.

A winner will be identified for each category. The challenge design has the same training/leaderboard/test split as was used in the prior solubility competition. Some 70 % percent of the dataset is made available as a training set, leaving the remaining 30 % as equally-sized leaderboard and blind test sets, respectively. Submissions will be evaluated on their predictive accuracy, which will be defined by the average AUC (Area Under the Curve) for each property.

The challenge will be hosted on the OCHEM website [5] (https://ochem.eu), where all training and test set data can be downloaded and where predictions can be submitted by participants. Each participant can be in one team only. Each team can upload and submit any number of predictions, which will be scored on the leaderboard and blind test sets. The data from the leaderboard set will be publicly released on December 15th and participants can use it to improve their models. The same strategy was used during the Tox24 challenge, which allowed for the mitigation of potential bias by users who may have tried to identify activities of all molecules from the leaderboard set [6]

By announcing the results of this competition at SLAS 2026 in Boston Feb 7-11th 2026, we align with the society's commitment to help effectively promote combined experimentation and computational methods to improve the drug discovery process. The winning models, which will be unveiled during the conference, will demonstrate collaborative spirit that defines our field: data shared openly, questions framed creatively, and solutions generated collectively.

We invite the global scientific community to join us in this competition and welcome both experimental scientists and computational modelers to participate.

CRediT authorship contribution statement

Katholiki Skopelitou: Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. **Federica Rossella:** Formal analysis, Writing – original draft, Writing – review & editing. **Rawdat Awuku Larbi:** Writing – original draft. **Philip Gribbon:** Conceptualization, Writing – original draft. **Thalita Cirino:** Data curation, Formal analysis, Software, Writing – original draft. **Igor V. Tetko:** Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Declaration:

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This editorial was partially funded by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 823886 (DRIVE). European Union's Horizon Europe programme under the Marie Skłodowska-Curie Actions Doctoral Networks grant agreement No. 101120466 "Explainable AI for Molecules" (AiChemist). The authors thank Dr. Katya Ahmad for her comments and suggestions.

References

[1] Harmel RK, Parker CN. 1st EU-OPENSCREEN/SLAS data mining competition to predict compounds solubility. SLAS Discov 2025;34:100238. https://doi.org/10.1016/j.slasd.2025.100238.

[2] Hunklinger A, Hartog P, Šícho M, Godin G, Tetko IV. The openOCHEM consensus model is the best-performing open-source predictive model in the first EUOS/SLAS joint compound solubility challenge. SLAS Discov 2024;29(2):100144. https://doi.org/10.1016/j.slasd.2024.01.005.

[3] Souza RC, Duarte JC, Goldschmidt RR, Borges Jr I. Predicting fluorescence emission wavelengths and quantum yields via machine learning. J Chem Inf Model 2025;65(7):3270–81. https://doi.org/10.1021/acs.jcim.4c02403.

https://doi.org/10.1016/j.slast.2025.100374
Received 3 November 2025; Accepted 19 November 2025
Available online 22 November 2025

[4] Ksenofontov AA, Lukanov MM, Bocharov PS, Berezin MB, Tetko IV. Deep neural network model for highly accurate prediction of BODIPYs absorption. Spectrochim Acta A Mol Biomol Spectrosc 2022;267:120577. https://doi.org/10.1016/j.saa.2021.120577.
[5] Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W, Brandmaier S, Abdelaziz A, Prokopenko VV, Tanchuk VY, Todeschini R, Varnek A, Marcou G, Ertl P, Potemkin V, Grishina M, Gasteiger J, Schwab C, Baskin II, Palyulin VA, Radchenko EV, Welsh WJ, Kholodovych V, Chekmarev D, Cherkasov A, Aires-de-Sousa J, Zhang Q-Y, Bender A, Nigsch F, Patiny L, Williams A, Tkachenko V, Tetko IV. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 2011;25(6):533–54. https://doi.org/10.1007/s10822-011-9440-2.
[6] Eytcheson SA, Tetko IV. Which modern AI methods provide accurate predictions of toxicological end points? Analysis of Tox24 challenge results. Chem Res Toxicol 2025;38(9):1443–51. https://doi.org/10.1021/acs.chemrestox.5c00273.

Katholiki Skopelitou[a,*], Federica Rossella[a], Rawdat Awuku Larbi[a], Philip Gribbon[a], Thalita Cirino[b], Igor V. Tetko[c]

[a] *EU-OPENSCREEN, Campus Berlin-Buch, Building 87, Robert-Rössle-str. 10, Berlin, 13125, Germany*

[b] *Molecular Biotechnology and Health Sciences Department, University of Turin, Turin 10126, Italy*

[c] *Institute of Structural Biology, Molecular Targets and Therapeutics Center, Helmholtz Munich - Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), 86764 Neuherberg, Germany*

[*] Corresponding author at: EU-OPENSCREEN, Campus Berlin-Buch, Building 87, Room 1.05, Robert-Rössle-str. 10, 13125 Berlin, Germany.
E-mail address: kathy.skopelitou@eu-openscreen.eu (K. Skopelitou).