



EUOS25 Challenge

Goal: The EUOS25 Challenge assesses computational methods for predicting absorption and fluorescence properties of compounds.

Data: A dataset of approximately 100k compounds systematically tested under the EU-OPENSECREEN bioprofiling program. It measured multiple endpoints, including solubility, which had previously been used in the [Kaggle Solubility Prediction Challenge](#)[1,2]. For the present EUOS25 Challenge, absorbance and fluorescence spectra were processed to create the prediction targets. The dataset is split 70% for training and 30% for testing (leaderboard + blind set).

Challenge 1: Absorption (Transmittance, %T)

Develop classification models to predict compounds with absorption (Transmittance \leq 70%):

- * **Subtask 1a:** at 340 nm
- * **Subtask 1b:** averaged from 450-679 nm range

Challenge 2: Fluorescence

Develop classification models to predict compounds that fluoresce above predefined thresholds at:

- * **Subtask 2a:** 340/450 nm (excitation/emission)
- * **Subtask 2b:** any of 480/540, 525/598, or 560/610 nm (excitation/emission)

N.B.! The modelling of subtasks *b is much more difficult due to only few active compounds available for them (0.23% and 1.5% for Fluorescence and Transmittance tasks, respectively).

Additional Data: The use of any related data to increase model accuracy (e.g., within multi-task modelling, pretraining, etc.) except experimental data for test set compounds, is allowed. These data should be made publicly available during model validation. The participants should use **only structural information** for prediction of compounds from the test sets. The information about plates and wells (in the extended files) can be, however, used to normalise training set data, if required. Notice that we found statistically significantly lower average absorption for well N8 (>450 nm) and plates C1046, C1070, C1099 (both wavelengths) and C1054, C1096, C1051, C1048, C1053, (450 nm only), which may indicate some problem with them. The data from these wells and plates were exclusively used for the training set.