Federated Learning in Computational Toxicology

Our Perspective on the Effiris Hackathon

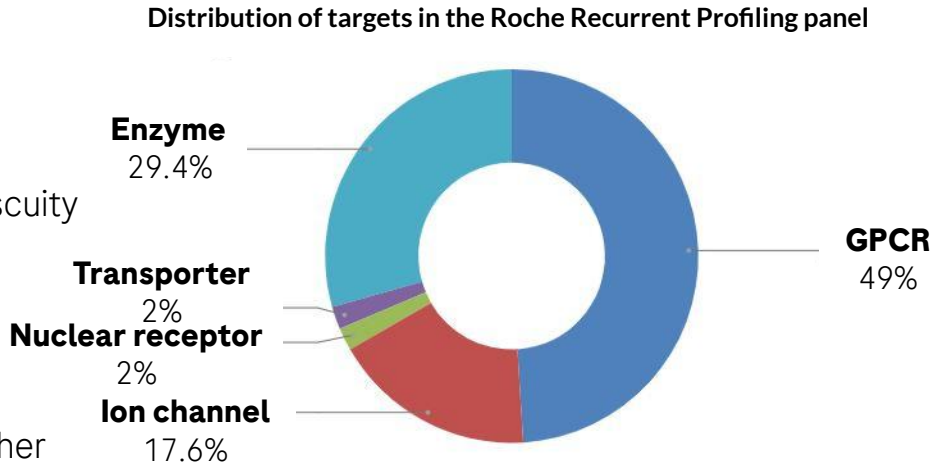
Andrea Andrews-Morger
Roche Pharma Research and Early Development
Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd

RDKit UGM 11.09.2024

Secondary pharmacology

Off-target interactions which may occur at related or structurally distinct targets

- Secondary Pharmacology screening is used from Candidate Selection to post Marketing
- Regulatory submissions require secondary pharmacology data to be included
- Panel-based screens are used to assess promiscuity and identify potential safety flags
- Roche recurrent panel for 50 targets
- Panel testing typically uses only a single test concentration at 10 uM but may need to be higher



Bendels et al. Safety screening in early drug discovery: An optimized assay panel

Application areas of secondary pharmacology prediction models

Prediction of toxicity to

- Guide optimization of a drug candidate
- Prioritise between different chemical classes, i.e. with very different scaffolds
- Save money - run in silico before submitting to panel:
  - Only submit to panel if in silico hits, or if in silico cannot make an informative prediction
  - Do not submit if compound is predicted to be highly promiscuous

Toxicological (in vitro) data

Can we overcome data challenges using federated learning (FL)?

Challenges of in vitro toxicology data

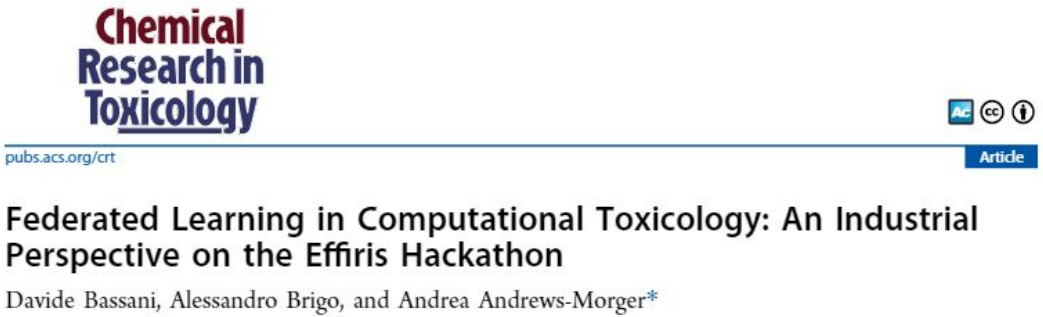
- Comparably small datasets
- Typically high data imbalances between numbers of active vs. inactive compounds
- Often challenging translatability from binding to mechanistic (e.g. agonist/antagonist) to in vivo data

	Model built on		
	Proprietary data	**Public data**	**Prop. data from other company**
Chemical space	limited	broad	limited, but different
Data reliability	high	low	high
Access	easy	easy (need curation)	difficult

Table adapted from a Lhasa presentation in the context of the “Effiris project by Lhasa Limited, UK”

Effiris

A secondary pharmacology model suite using federated learning (FL)

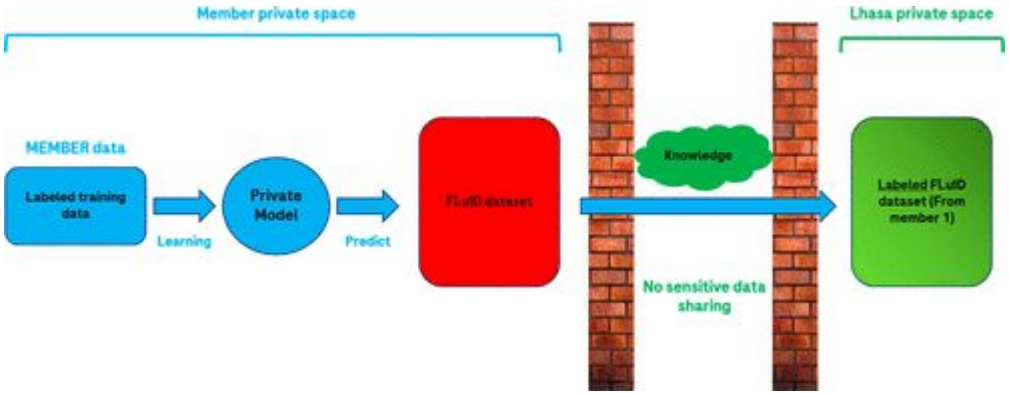


- Can we increase the chemical space of machine learning (ML) training data - without sharing sensitive data - using FL?
- Effiris federated learning software to predict secondary pharmacology developed by Lhasa Limited, UK
- 50 ML models for off-target prediction
- Organised Effiris Federated Learning hackathon: 9 off-targets, 7 pharmaceutical companies
- Compare models built with Effiris to internally built ML models

Effiris project by Lhasa Limited, UK
Bassani, Brigo, Andrews-Morger, ChemResTox, 2023

Effiris federated learning approach

Student-Teacher approach



No sharing of sensitive (proprietary) data

Project information from internal data to 300 K public and unlabelled dataset

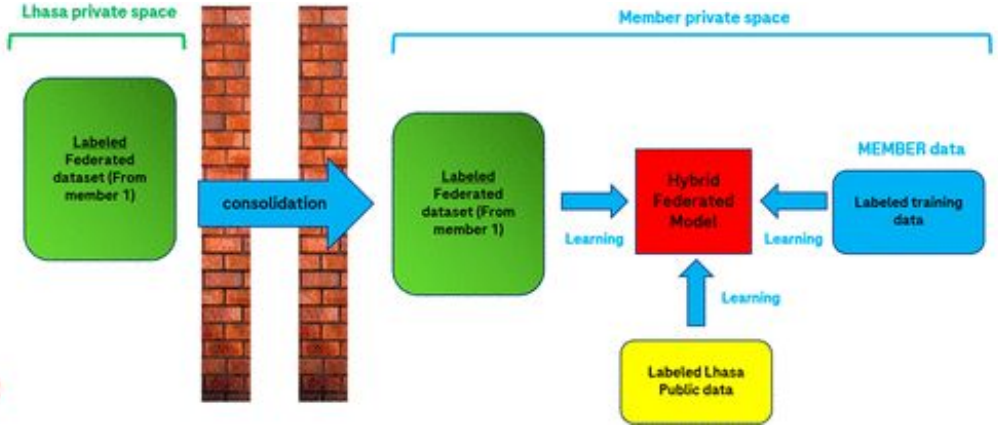
Consolidate predicted labels from industry partners

Hypothesis: size and chemical space of dataset more important than accuracy of data

Bassani, Brigo, Andrews-Morger, ChemResTox, 2023

Effiris federated learning approach

Student-Teacher approach



No sharing of sensitive (proprietary) data

Project information from internal data to 300 K public and unlabelled dataset

Consolidate predicted labels from industry partners

Hypothesis: size and chemical space of dataset more important than accuracy of data

Bassani, Brigo, Andrews-Morger, ChemResTox, 2023

Differences Effiris - MELLODDY

	Effiris	MELLODDY
Federated learning approach	Student-teacher model	Machine learning ledger orchestration for drug discovery
Endpoints	1 endpoint per model, defined and known	“All available data” used, assay/target names not disclosed
# tasks per model	Single-task	Multi-task
Sharing of	Predicted labels on FLuID dataset	Gradients
Training set	Consolidated, predicted labels for FLuID dataset + internal data	Internal data

J Chem Inf Model. 2023 Aug 29. doi: 10.1021/acs.jcim.3c00799

Collection of internal data

Effiris evaluation work performed on five off-targets

1. Collect internal activity data
2. Binary models: PCT_INHIB data, threshold: </>= 50% (10 µM)
3. Validate Effiris model on Lhasa and Roche test set
4. Build model using own algorithms on internal and federated data

Effiris: Currently 50 classification models available, regression models under development
Test set: time-split 2020-2022

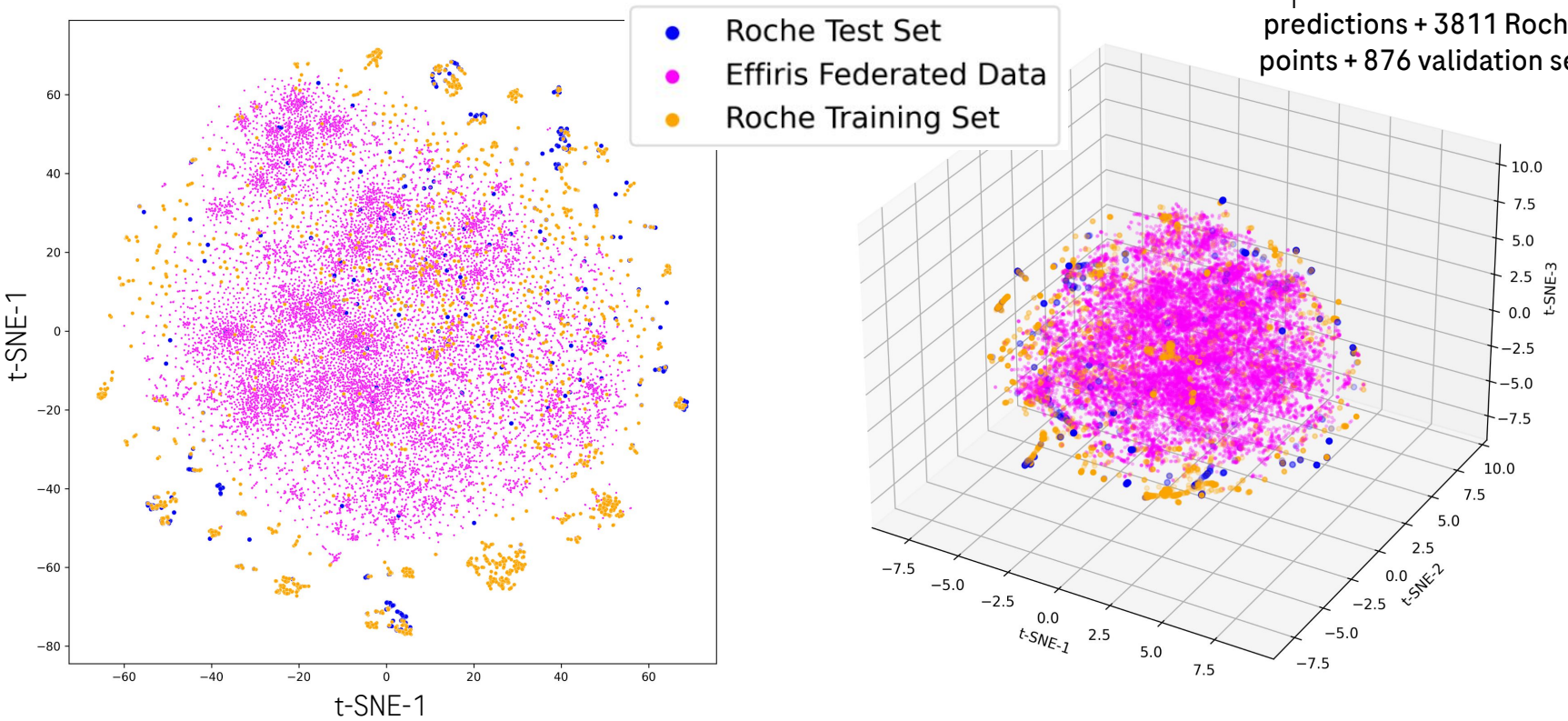
Target	Assay type	Training set (inactive/active)	Test set (inactive/active)
ACh M_1	binding	3811 (2829/982)	876 (745/131)
$GABA_A$ Benz	binding	3614 (3216/398)	877 (819/58)
5-HT_{2B}	binding	3374 (2447/927)	876 (709/167)
hERG	Electrophysiology (SynchroPatch)	2285 (1388/897)	556 (309/247)
COX-2	enzymatic	3042 (2815/227)	881 (809/72)

Chemical space analysis

Chemical space comparison between internal & federated data using t-SNE

Example of M_1: Internal space largely but not completely covered by federated data

M_1 : 20K Federated SOHN predictions + 3811 Roche points + 876 validation set



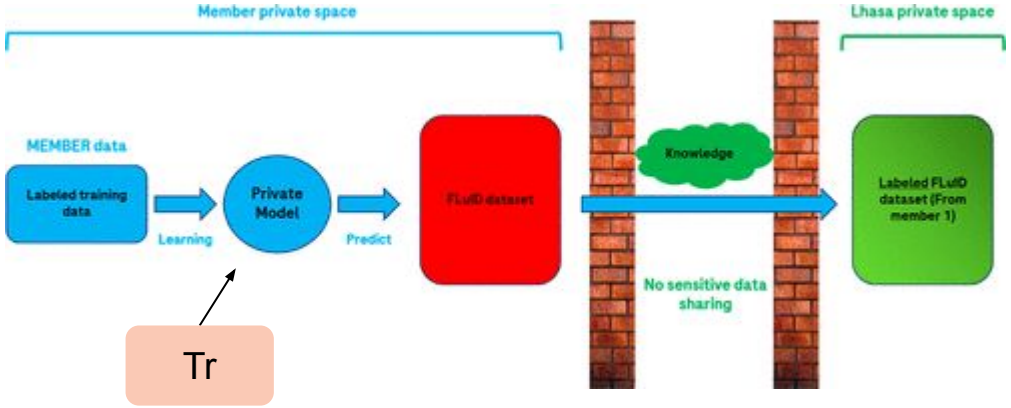
t-SNE fitted using RDKit 2D descriptors and Morgan fps (radius 3, 2048 bits)

SOHN approach

SOHN

Self Organising Hypothesis Network Classifier

- Upload internal data (Member data) to build SOHN Teacher model (Tr)
- Make predictions on FLuID dataset (external, 350K) and share predicted labels with Lhasa
- Lhasa consolidated labels across teacher models and selected 20K compounds to be used as federated dataset

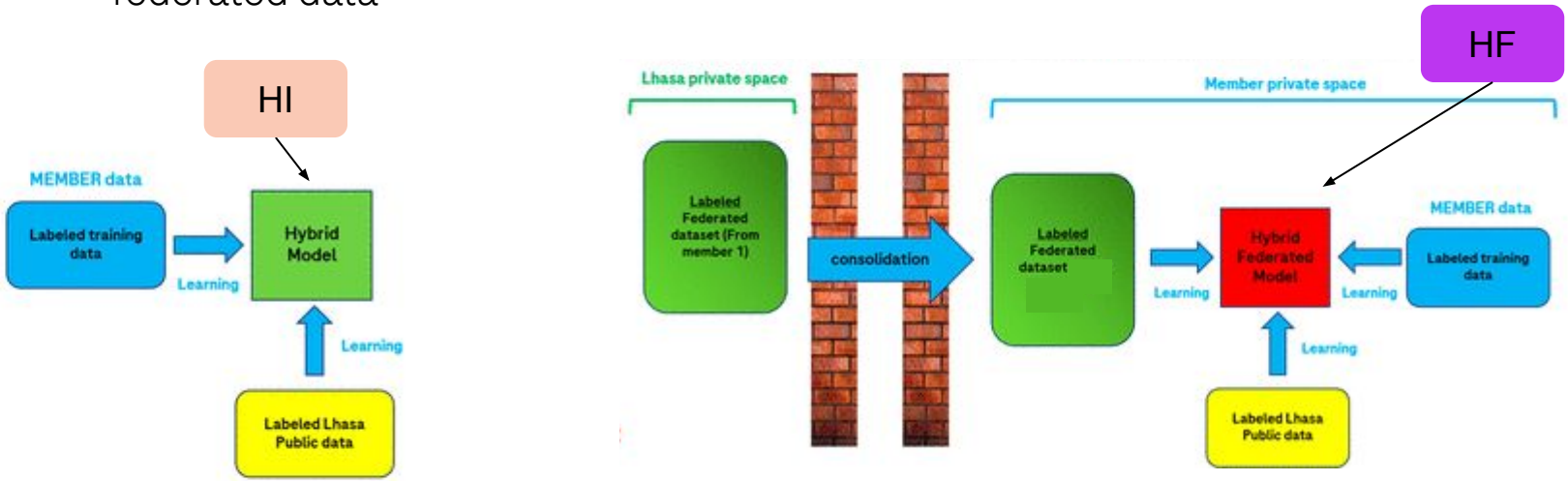


Hanser *et al. J Cheminform* 11, 9 (2019). https://doi.org/10.1186/s13321-019-0334-y
Hanser *et al. J Cheminform* 6, 21 (2014). https://doi.org/10.1186/1758-2946-6-21

SOHN

Self Organising Hypothesis Network Classifier

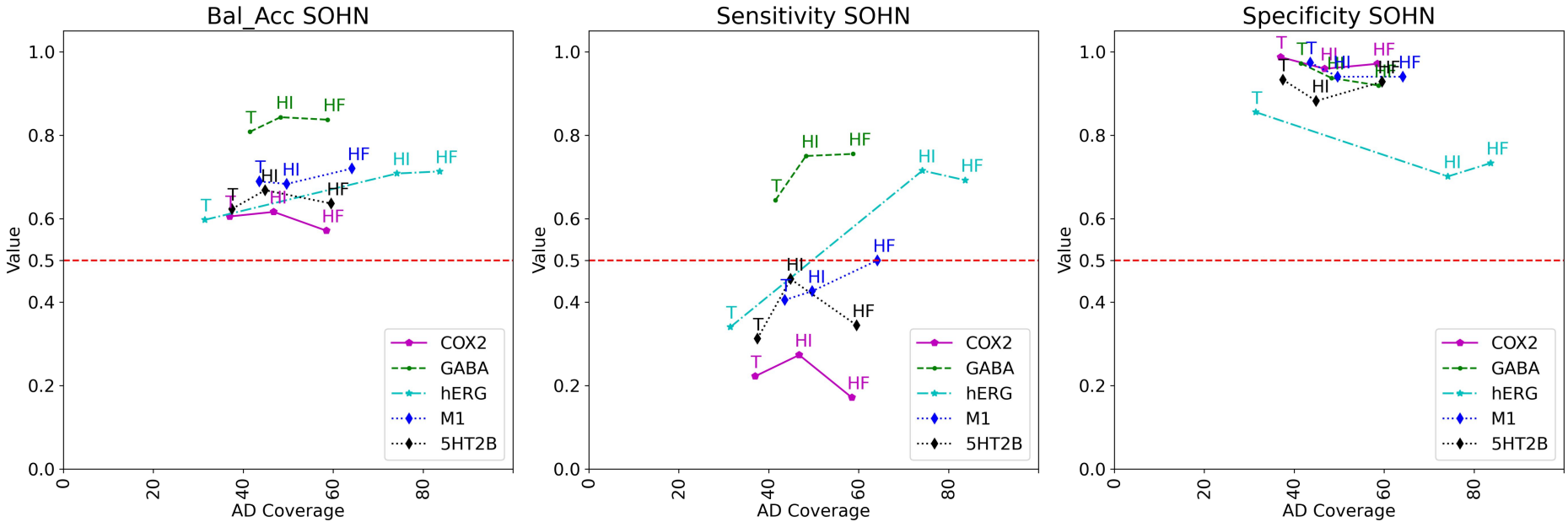
- Hybrid model I (HI): build hybrid Model based on internal and Lhasa public data
- Hybrid model II (HF): build hybrid model based on internal, Lhasa public, and federated data



Hanser *et al. J Cheminform* 11, 9 (2019). https://doi.org/10.1186/s13321-019-0334-y
Hanser *et al. J Cheminform* 6, 21 (2014). https://doi.org/10.1186/1758-2946-6-21

Impact of Effiris on model performance and applicability domain - SOHN

Applicability domain increase, sensitivity > 0.5 for $GABA_A$ and hERG

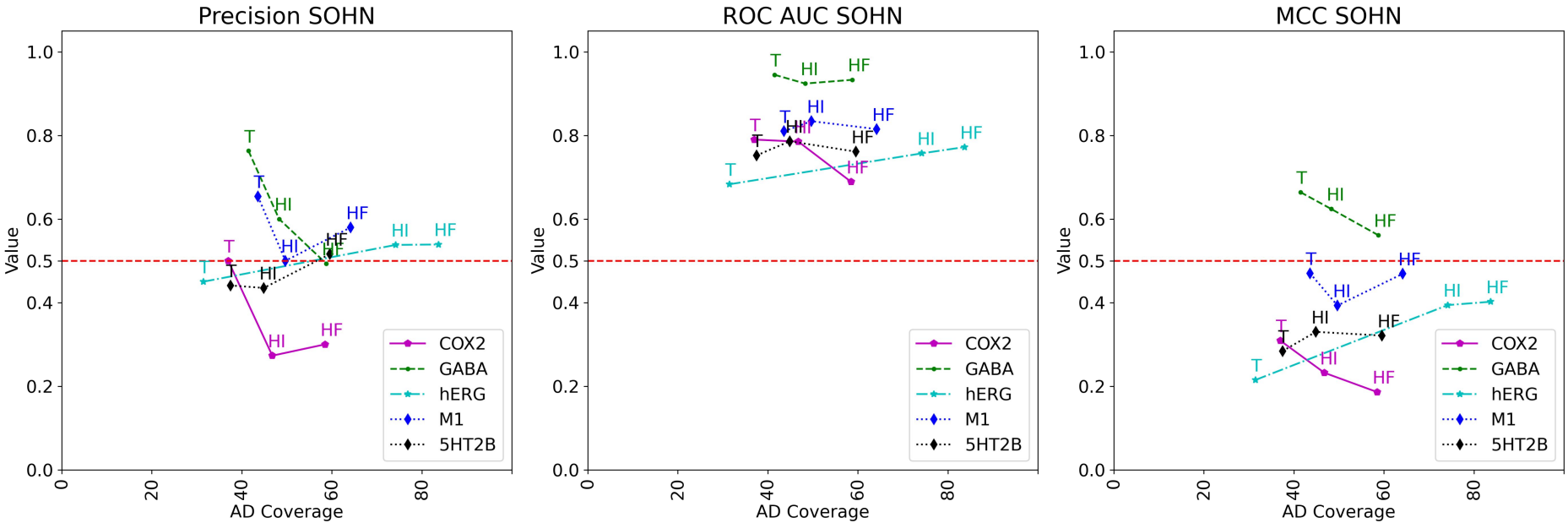


T = teacher with internal data
HI = hybrid with public data
HF = hybrid with Lhasa Federated data

Hanser *et al. SAR and QSAR Environ. Res.* 27, 11 (2016). https://doi.org/10.1080/1062936X.2016.1250229

Impact of Effiris on model performance and applicability domain - SOHN

Change of performance measure is target-dependant



T = teacher with internal data
HI = hybrid with public data
HF = hybrid with Lhasa Federated data

Hanser *et al. SAR and QSAR Environ. Res.* 27, 11 (2016). https://doi.org/10.1080/1062936X.2016.1250229

Building internal models
On internal data and federated data

Methods to build internal models

Internal model building, benchmark, and analysis performed by Davide Bassani

Roche

Datasets

Used different sets for each target: internal, Lhasa Federated (20K), Lhasa FLuID (300K+)

ACh M_1
$GABA_A$ Benz
5-HT_{2B}
hERG
COX-2

RDKit Descriptors:
1D+2D Molecular Descriptors
+
Morgan Fingerprints
from RDKit (2048 bits, radius 3)

ML Models

Gaussian NB

K-Neighbors

Decision Tree

Random Forest

ExtraTrees

AdaBoost

Gradient Boosting

XGBoost

MLP

SGD

SVC

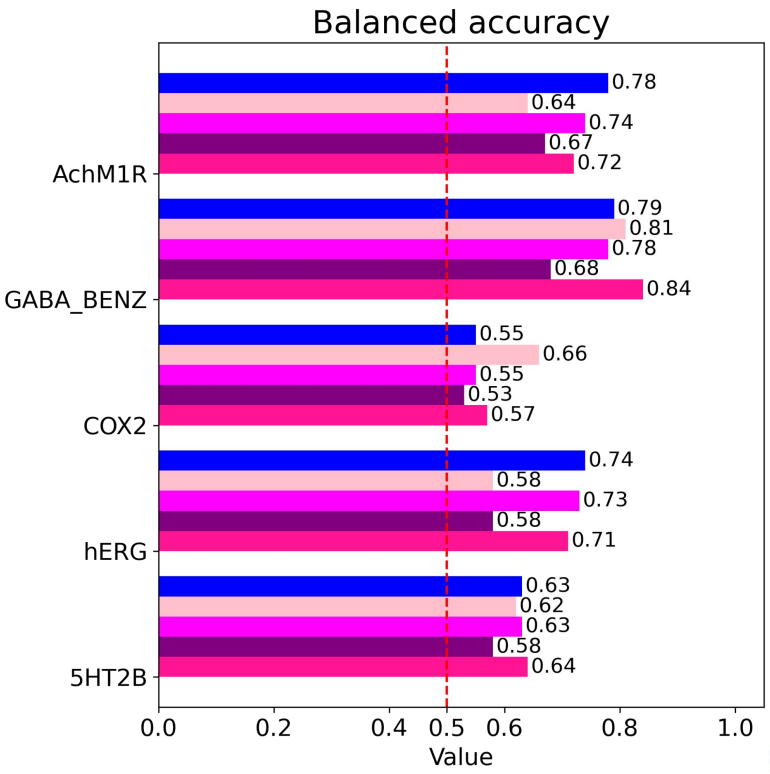
Hyperparameter Optimization
(5-fold, 10 repeats Stratified CV)

Prediction on internal test set + comparison with Lhasa models

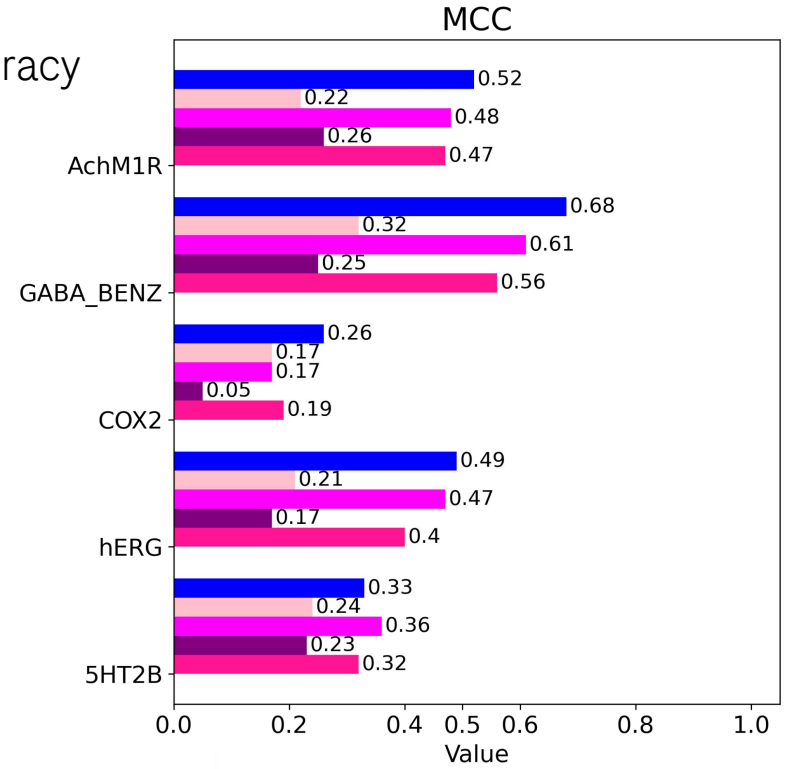
(usually SVC and XGBoost were best-performing)

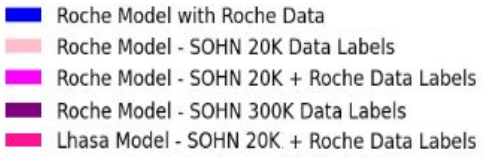
Lhasa vs. internal ML models - validation on internal test set

Performance of Lhasa (SOHN) and Roche models is very similar



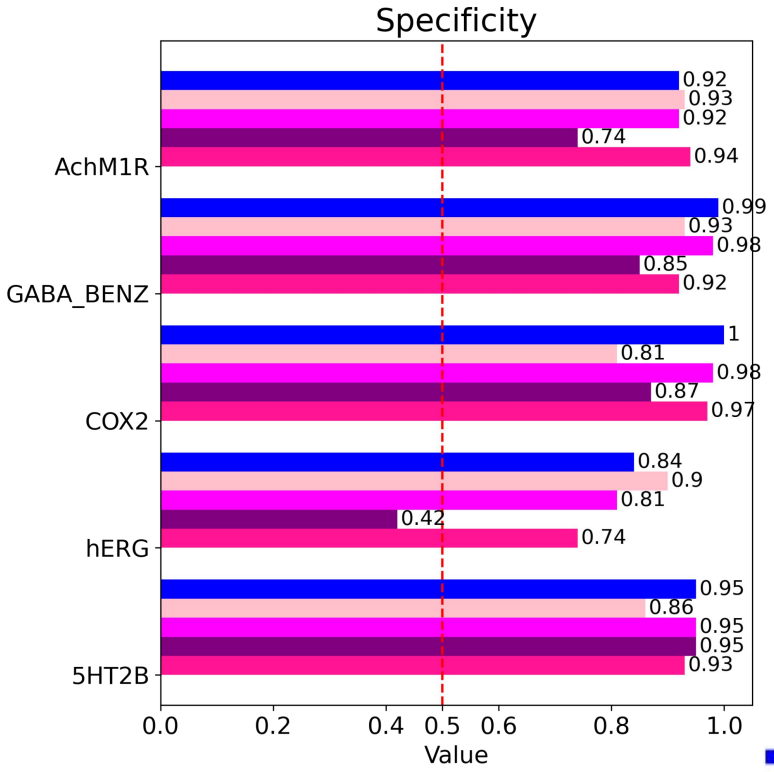
Balanced Accuracy
vs
MCC



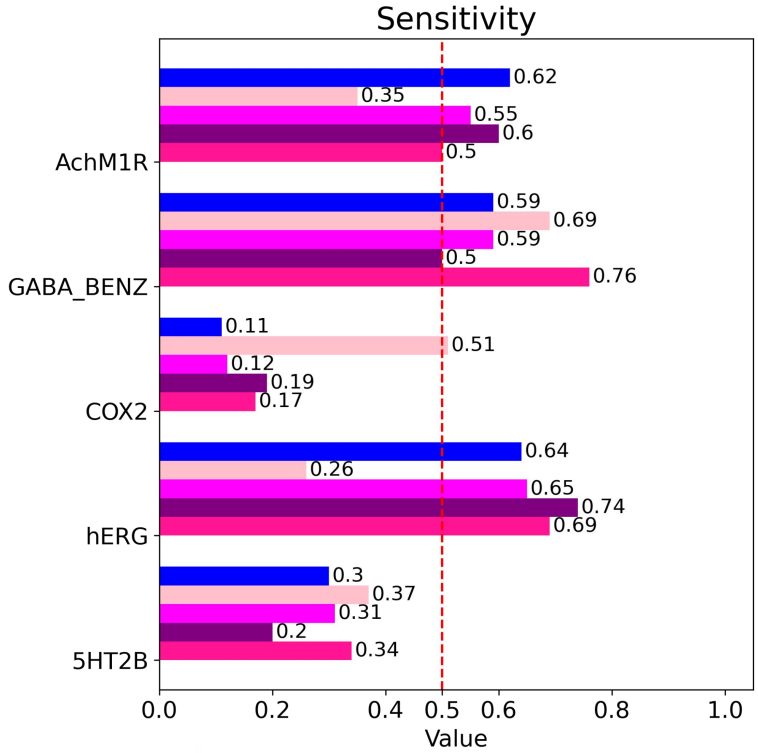


Lhasa vs. internal ML models - validation on internal test set

Also in this case, the performance of Lhasa (SOHN) and Roche models is very similar...



Specificity vs Sensitivity



Conclusions and discussion

Advantages of Effiris

- Uses knowledge from other industry partners (provided that they join Effiris)
- Larger training sets
- Increased Applicability Domain (AD)
- New models for ca. 50 targets
- Models being regularly updated with new member data (uncertain update schedule over time)
- Maintenance @ Lhasa (Exception: model updates, integration in to internal systems)

Limitations and Challenges

- Majority of training set consists of predictions → Approach is more technical than toxicologically-/biologically-oriented
- Target-dependency of performance: reduced confidence in models not validated with internal data
- As of now: only classification models are available (regression models planned)
- Data produced with different assay protocols (e.g. different temperature for hERG)
- Input from other companies conditional to their participation in Effiris
- FLuID dataset could be more tailored towards drug-like space

Summary

- Modelling secondary pharmacology data has several challenges
- Student-Teacher approach to leverage data from different sources without sharing sensitive information
- Model performance
  - Federated learning shows applicability domain increase
  - Lhasa model performance (SOHN & MLP) very similar to internal models
  - Target-dependant performance change, hard to transfer validation results to other models
- More tailored FLuiD dataset towards drug-like space could potentially increase performance of federated learning

Acknowledgements

Roche - Pharmaceutical Sciences, Pharma Research and Early Development

- Davide Bassani (Translational Pharmacokinetics and Clinical Pharmacology)
- Alessandro Brigo (Translational Safety Assessment)
- Liudmila Polonchuk (Investigative & Immuno Safety)
- Michael Reutlinger (Computer-Aided Drug Design)
- Sonia Roberts (Translational Safety Assessment)
- Angelo D'Annunzio (Molecule Performance)
- Wolfgang Muster (Investigative & Immuno Safety)

Lhasa Limited

- Thierry Hanser
- Jack Yung
- Ioannis Xanthis
- Laura Johnston

Doing now what patients need next