ADMETox.AI

Multi-task deep learning platform for comprehensive ADMET property prediction

Rodion Bykadorov
ADMETox.AI, Rostov-on-Don, Russia
ORCID 0009-0005-2231-1795 · bradic@yandex.ru

XXXII Symposium on Bioinformatics and Computer-Aided Drug Discovery (BCADD-2026) · 21 October 2026

Why ADMET prediction is hard — not just expensive

$2.23B

average cost to develop one asset[1]

~40%

of clinical failures linked to ADME / Tox[2]

379 – 9 861

training molecules per TDC endpoint

Small data, vast space

Large GNNs overfit or lack signal to generalise on a few thousand molecules.

Leaderboard reproducibility

Independent audit reports leakage and overfitting in several top-ranked TDC models[3].

GPU-bound and opaque

Inference often needs GPUs; the chemist gets a number, not a reason.

Our question: can a CPU-only, reproducible pipeline reach state of the art on all 22 TDC ADMET Group endpoints?

[1] Deloitte, Measuring the return from pharmaceutical innovation, 2024 [2] Kola & Landis, Nat Rev Drug Discov 2004 [3] Koleiev et al., J Chem Inf Model 2026, 66(14):8045

Approach: let the data choose the model — per endpoint

Each dataset profiled first (size, balance, chemistry, scaffold overlap), then model families compared.

SMILES input molecule

▼

Curated descriptors + fingerprints physico-chemical, multi-scale counts

▼

Frozen pretrained graph embeddings deep representations, not fine-tuned

▼

Gradient-boosting ensembles → blend CatBoost / LightGBM / XGBoost

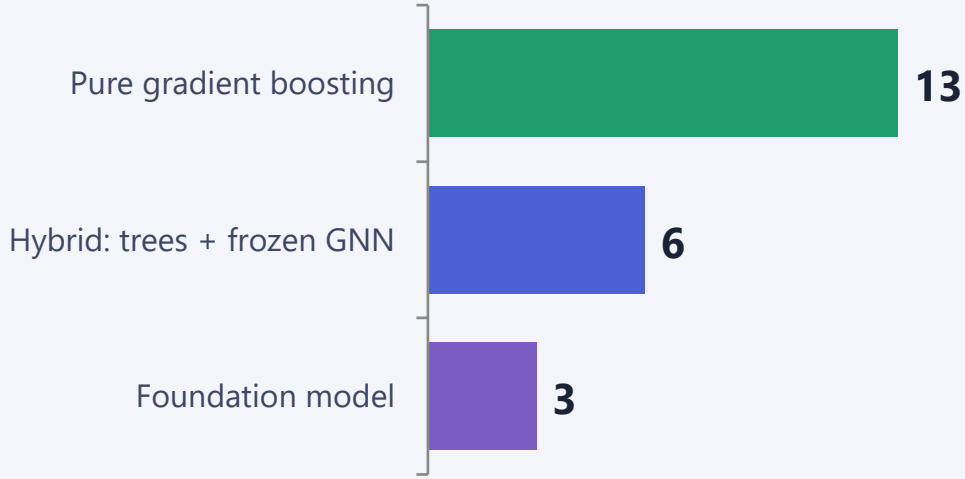
▼

Calibration + applicability domain probability, in / borderline / out

▼

Report for the chemist 22-endpoint profile + explanation

Best model family per endpoint (n = 22)



All 22 run inference on CPU
incl. the 3 foundation models (GPU used only for their training)

What makes the difference: ablation-tested decisions

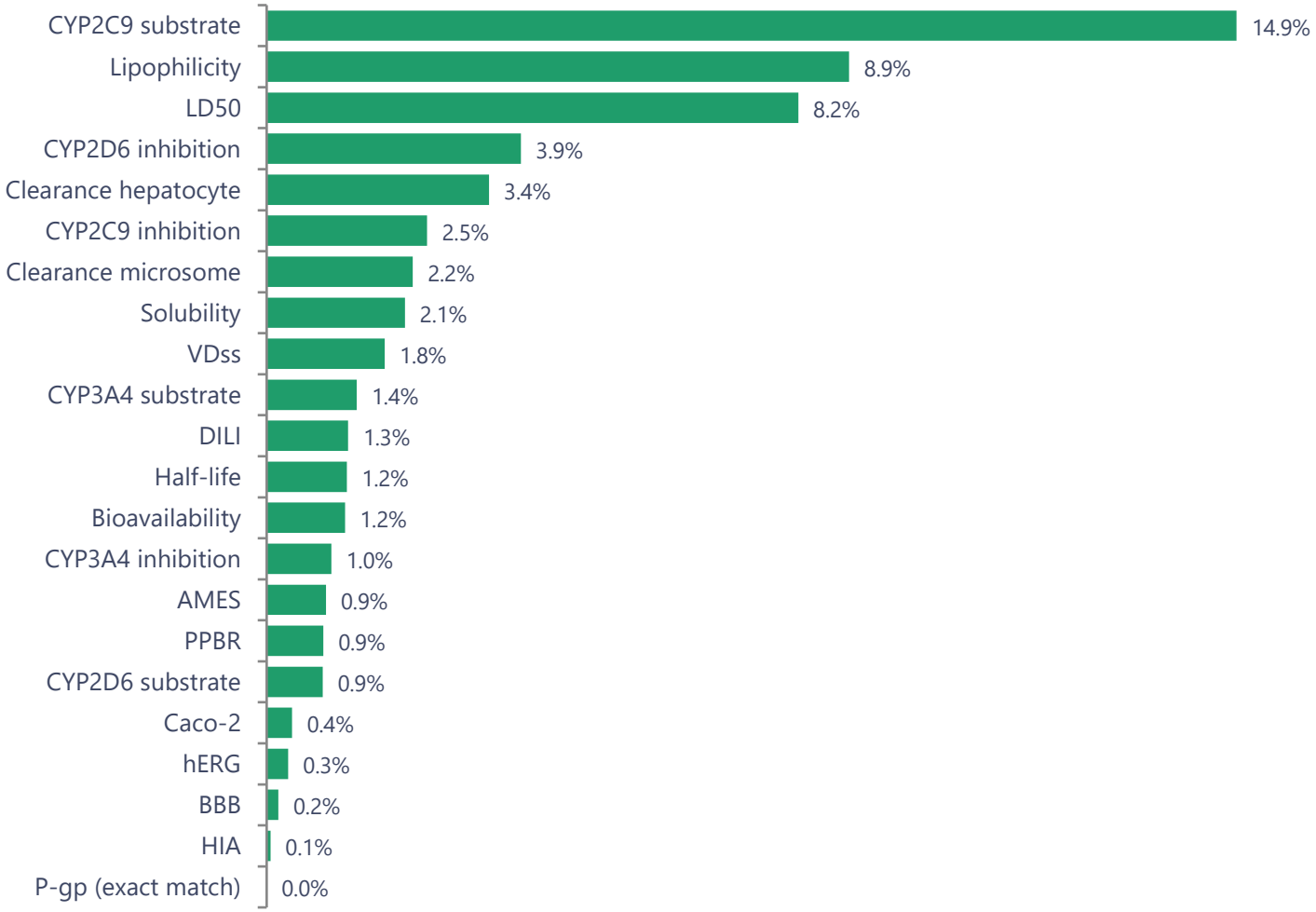
Not one trick — each component was tested separately in ablation experiments.

Technique	What it changes	Measured effect
SHAP-guided feature pruning	Removes up to 70% of noisy descriptors	**≈2× faster training, +5–10% quality**
√-balanced class weights	√(n_neg / n_pos) instead of the raw ratio	**up to +3% AUROC, no overfitting**
Stochastic subsampling of trees	Controlled randomness during boosting	**lower seed-to-seed variance, higher mean**
Curated MapLight-type features	2 580-d vs 5 849-d plain Morgan	**≈ +18 AUROC points on BBB**

Works on small datasets without a GPU cluster — that is why it scales to all 22 endpoints, not just a few showcase ones.

22 / 22 endpoints at or above published SOTA

Improvement over TDC ADMET Group leaderboard best, % of SOTA value, official metric of each endpoint.



21 above SOTA

1 exact match (P-gp 0.938 = 0.938)

Evaluation protocol

- Official TDC scaffold splits and metrics
- Mean ± std over 5 independent seeds
- Train / test leakage audit
- Deployed models verified vs frozen references: 11 bit-exact, 11 regenerable

Full table → backup slide

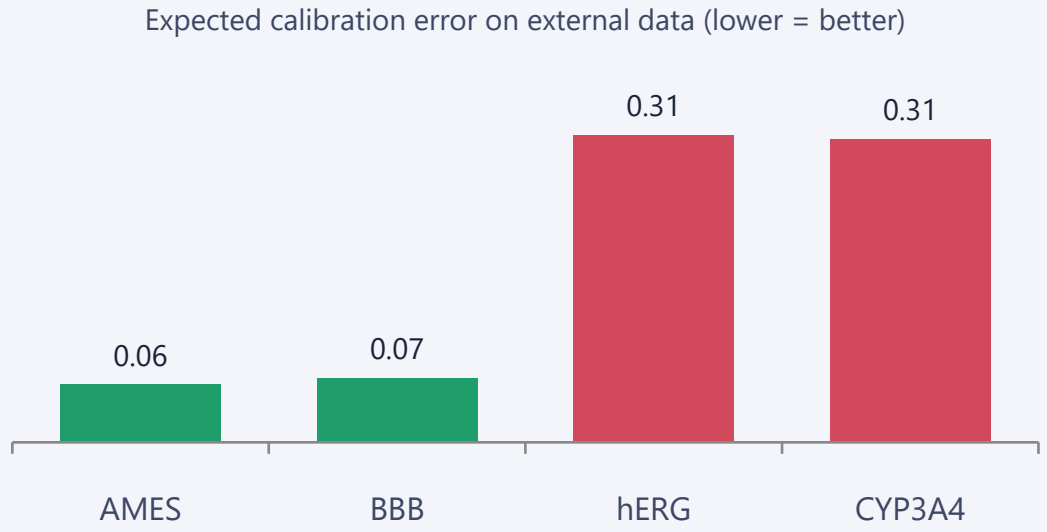
Beyond the benchmark: calibration & applicability domain

Benchmark ranking is not deployment readiness — inputs always drift away from the training set.

External validation (never seen in training)

AmesFormer · B3DB · hERG_Karim · Figshare sets

- **Ranking (AUROC) transfers well**
- **Calibration shifts with domain → re-fit on user data**



Applicability domain — for every molecule

1. **Tanimoto similarity** — to the training set
2. **Scaffold coverage** — is the core seen in training?
3. **Descriptor ranges** — MW, polarity, logP within bounds

In domain | **Borderline** | **Out of domain**

Validated, not assumed: out-of-domain flags correspond to larger errors on held-out data (AMES, hepatocyte clearance, lipophilicity, solubility, …).

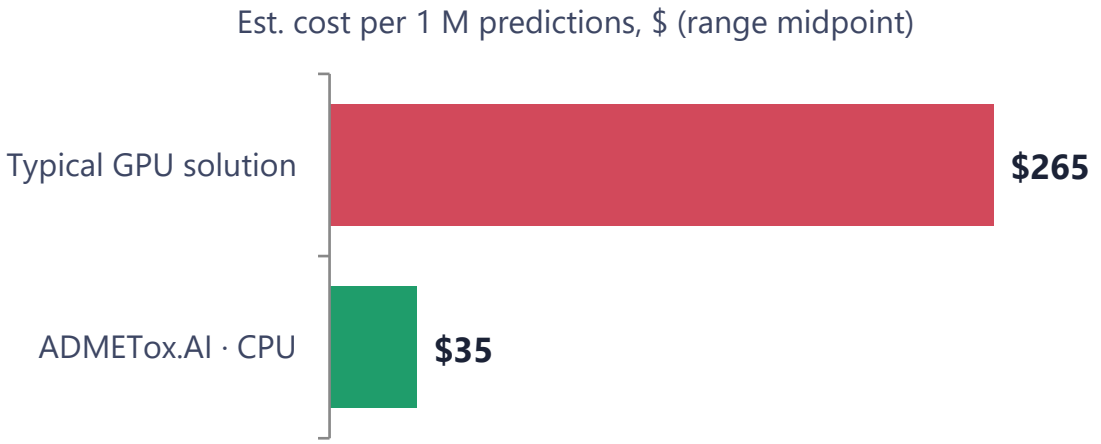
CPU-only inference, reproducible by design

No accelerator in the inference loop; every release checked against frozen reference artifacts.

< 1 s — full 22-endpoint profile per molecule, on CPU

≈ 5 GB — all 22 models on disk

1 PC — AMD Ryzen 9 3900X, 32 GB RAM (+ GPU only to train 3 foundation models)



Ranges: $10–60 (CPU) vs $30–500 (GPU); estimates from public cloud prices and benchmarks.

Deterministic CPU execution avoids GPU floating-point non-determinism → bit-exact predictions can be audited.

Open & verifiable

- Public code per endpoint (github.com/Recconnect)
- 9 endpoint preprints with DOIs on Zenodo
- SHA-256 artifact manifest; exact vs regenerable linkage per model

The platform: from SMILES to a decision for the chemist

REST API (FastAPI) and web UI share one pipeline; demo available after the session.

Single-molecule analysis

One SMILES → profile across all 22 endpoints, with calibrated probability and domain flag.

Batch triage

Upload a CSV series → portfolio-level ranking and filtering of candidates.

Structure–property patterns

Scaffold-based pairs show which structural changes move a property.

Explanation for the chemist

Which fragments drive the signal, what to test, what to modify (chemist-reviewed texts for a reference set today).

Conclusions

Data-driven model choice beats model size

Boosting ensembles on curated features + frozen pretrained representations reach SOTA on 22/22 TDC endpoints (21 above, 1 exact match).

GPU for training a few models — not for inference

All 22 models run on CPU, < 1 s per molecule for the full profile.

Accuracy is only half the story

Ranking transfers to external data; calibration must be re-fitted; every prediction carries an applicability-domain flag.

Next: extended datasets · live explanation generation · prospective validation with partners — open to collaboration

Thank you!

Questions — in the Zoom chat or by e-mail

Rodion Bykadorov · ADMETox.AI
bradic@yandex.ru · Telegram @recconnect · github.com/Recconnect

Selected preprints (Zenodo):
hERG 10.5281/zenodo.21679073 · AMES 10.5281/zenodo.21671306 · BBB 10.5281/zenodo.21707384
Bioavailability 10.5281/zenodo.21702931 · P-gp 10.5281/zenodo.21759055 · CYP2D6 sub. 10.5281/zenodo.21828083

Backup: full results vs TDC ADMET Group SOTA

Endpoint	Group	Metric	ADMETox.AI	SOTA
hERG	Tox	AUROC	**0.883 ± 0.006**	0.880
AMES	Tox	AUROC	**0.879 ± 0.001**	0.871
DILI	Tox	AUROC	**0.937 ± 0.003**	0.925
LD50	Tox	MAE ↓	**0.507 ± 0.009**	0.552
HIA	Abs	AUROC	**0.994 ± 0.001**	0.993
Bioavailability	Abs	AUROC	**0.750**	0.741
Caco-2	Abs	MAE ↓	**0.255 ± 0.002**	0.256
Solubility	Abs	MAE ↓	**0.725**	0.741
Lipophilicity	Abs	MAE ↓	**0.425 ± 0.004**	0.467
P-gp	Abs	AUROC	**0.938 ± 0.001**	0.938 (=)
BBB	Dist	AUROC	**0.926 ± 0.002**	0.924

Endpoint	Group	Metric	ADMETox.AI	SOTA
PPBR	Dist	MAE ↓	**7.376 ± 0.064**	7.440
VDss	Dist	Spearman	**0.726 ± 0.004**	0.713
CYP3A4 inh.	Met	AUPRC	**0.925 ± 0.001**	0.916
CYP2C9 inh.	Met	AUPRC	**0.880 ± 0.002**	0.859
CYP2D6 inh.	Met	AUPRC	**0.821 ± 0.001**	0.790
CYP3A4 sub.	Met	AUROC	**0.676**	0.667
CYP2C9 sub.	Met	AUPRC	**0.502 ± 0.005**	0.437
CYP2D6 sub.	Met	AUPRC	**0.742 ± 0.020**	0.736
CL hepatocyte	Exc	Spearman	**0.554 ± 0.007**	0.536
CL microsome	Exc	Spearman	**0.644 ± 0.011**	0.630
Half-life	Exc	Spearman	**0.583 ± 0.004**	0.576

Mean ± std over 5 seeds where available; official TDC splits. Abs = absorption, Dist = distribution, Met = metabolism, Exc = excretion. ↓ lower is better.