THE COMPARATIVE ANALYSIS OF 2D AND 3D STRUCTURAL REPRESENTATIONS FOR AI-QSAR MODELING OF HUMAN CYP24A1 INHIBITORS

***Alexandra Yakusheva*[1], *Marina Shaladonova*[2]**

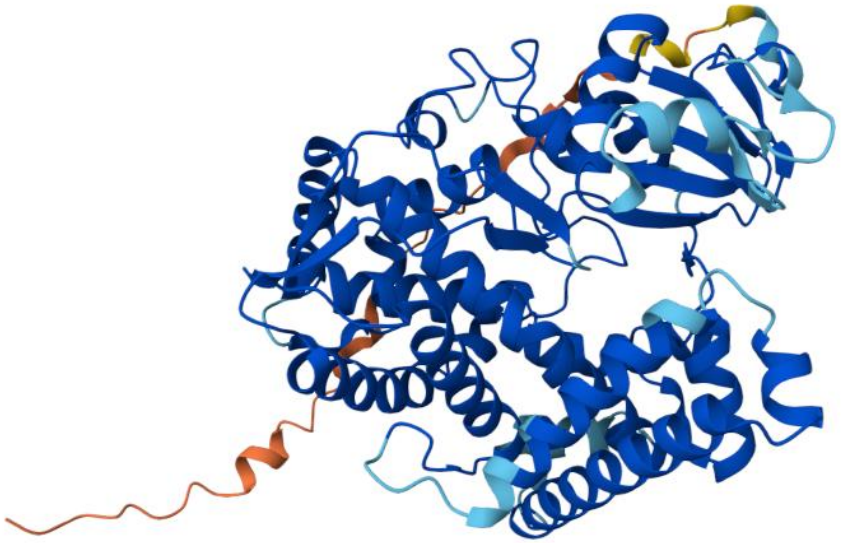
[1]*Belarusian State University, Minsk, Belarus*

[2]*Institute of Bioorganic Chemistry of National Academy of Sciences of Belarus, Minsk, Belarus*

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

Context

- Human **CYP24A1** is a key mitochondrial cytochrome P450 regulating vitamin D3 catabolism. Overexpression of CYP24A1 can lead to altered vitamin D levels, calcium metabolism disorders and tumor progression, which highlights CYP24A1 as an important drug target in both oncology and metabolic diseases.
- Artificial intelligence (AI) has recently been increasing in the drug development, utilizing the algorithms of machine learning (ML), deep learning (DL) and combined approaches.



Predicted 3D structure of CYP24A1 (AlphaFold ID: Q07973, colored by pLDDT confidence)

Aim of the study

- Modern CADD heavily leans toward 3D representations (3D GNNs, SE(3)-transformers like Uni-Mol) assuming spatial coordinates improve QSAR quality. Cytochrome P450 enzymes (specifically CYP24A1) possess highly flexible active sites.
- **Research Question:** Does adding 3D conformational complexity actually yield better predictions for CYP24A1, or does it introduce unnecessary noise compared to 2D feature engineering on limited data?

Framework

- Data processing (ChEMBL)

Molecule ChEMBL ID	SMILES	Standard Type	Standard Value	Standard Units
CHEMBL4074768	Cc1[nH]nc(O)c1C(c1cccc(OCc2ccncc2)c1)c1c(O)n[nH]c1C	IC50	2600.0	nM
CHEMBL4079445	O=C(NCC(c1ccccc1C(F)(F)F)n1ccnc1)c1ccc(/C=C/c2ccccc2)cc1	IC50	350.0	nM
CHEMBL4087325	O=C(NCC(c1ccc(Cl)cc1)n1ccnc1)c1ccc(/C=C/c2ccccc2)cc1	IC50	210.0	nM
CHEMBL1627208	C=C1/C(=C\C=C2/CCC[C@]3(C)[C@@H]([C@H](C)/C=C(\F)[S+](=N)([O-])c4ccccc4)CC[C@@H]23)C[C@@H](O)C[C@@H]1O	IC50	1000.0	nM
CHEMBL1253780	COc1ccc2c(c1)CC/C(=C\c1ccccc1/C=C/c1ccccc1)C2=O	IC50	5080.0	nM
CHEMBL1627205	C=C1/C(=C\C=C2/CCC[C@]3(C)[C@@H]([C@@H](C)/C=C/[S+](=N)([O-])c4ccccc4)CC[C@@H]23)C[C@@H](O)C[C@@H]1O	IC50	549.0	nM
CHEMBL1171660	O=C(NCC(c1ccccc1)n1ccnc1)c1cc2cc([N+](=O)[O-])ccc2o1	IC50	6400.0	nM

Custom dataset for human **CYP24A1** was created using raw data retrieved from **ChEMBL.**

We used quantitative inhibition metrics (IC_{50}, K_i, EC_{50}) to classify compounds into active and inactive groups. Compounds with $IC_{50} \leq 1000$ nM or $K_d \leq 1000$ nM were designated as active inhibitors. Molecules with higher values were assigned as **inactive controls.**

Fully balanced dataset ($N = 88$, 1:1 active/inactive ratio)

Framework

2D Molecular Descriptors & Fingerprints

- Calculated via PaDEL-pywrapper tools.
- Included MACCS structural keys (167 bits), Morgan fingerprints, and physicochemical properties (LogP, molecular weight, polar surface area)

3D Spatial Representations

- Retrieved 3D SDF coordinates of molecules from the **PubChem** database. Applied **Uni-Mol** an SE(3)-equivariant pre-trained molecular foundation transformer model that directly utilizes 3D atomic coordinates to capture spatial geometry and conformational features[1].

[1] Zhou, G., Gao, Z., Ding, Q. et al. (2022) *ChemRxiv*.

Framework

Fine-Tuned Uni-Mol: End-to-end domain adaptation of the Uni-Mol classification head

Transfer Learning: Extraction of 512-dimensional 3D spatial embeddings using pre-trained Uni-Mol, followed by Random Forest classification.

2D + Random Forest (Baseline): Classical machine learning approach combining complex 2D feature vectors with Random Forest classifiers

Hyperparameters were optimized using the **Optuna** framework



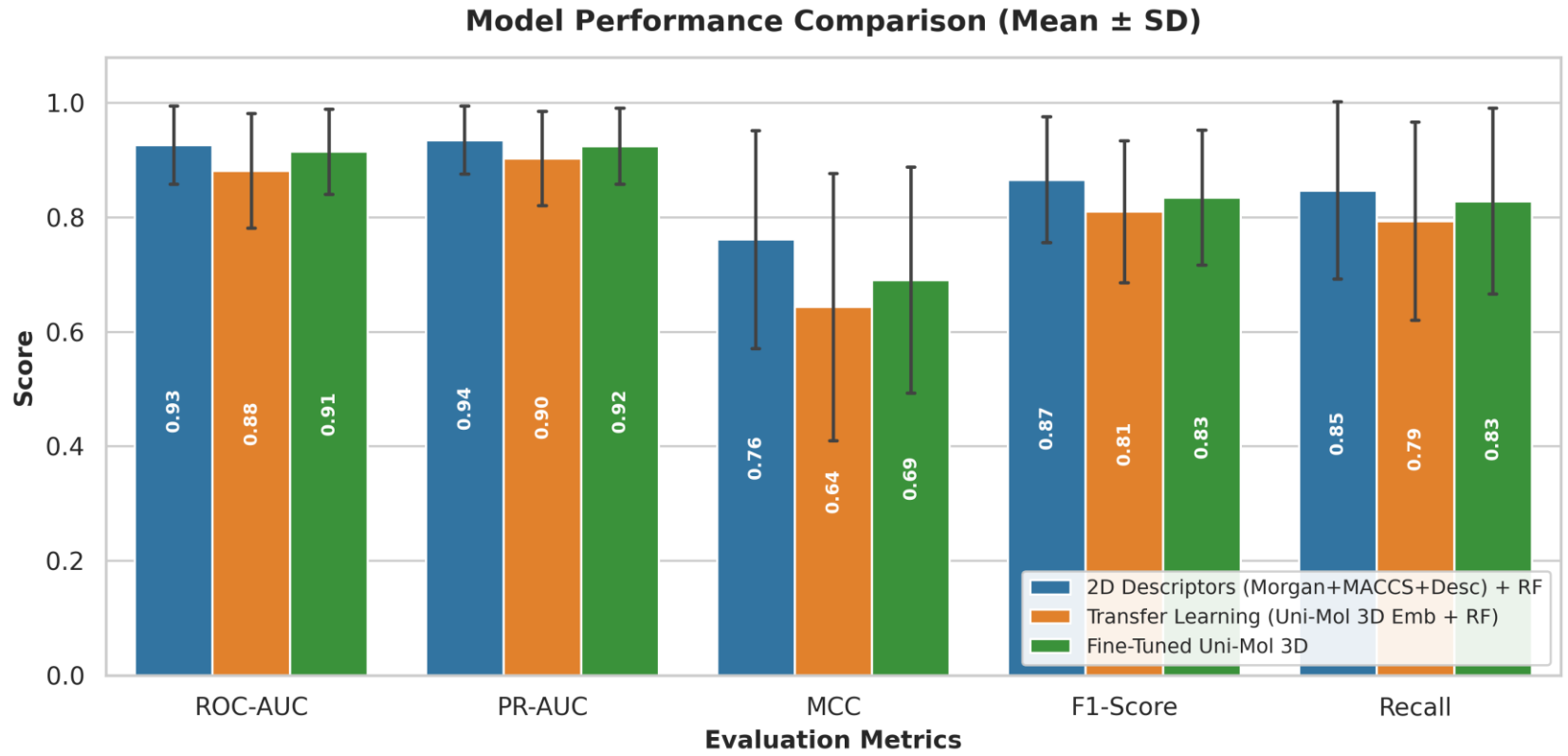
5-fold Stratified Cross-Validation to prevent data leakage and preserve class proportions



Comparative Evaluation

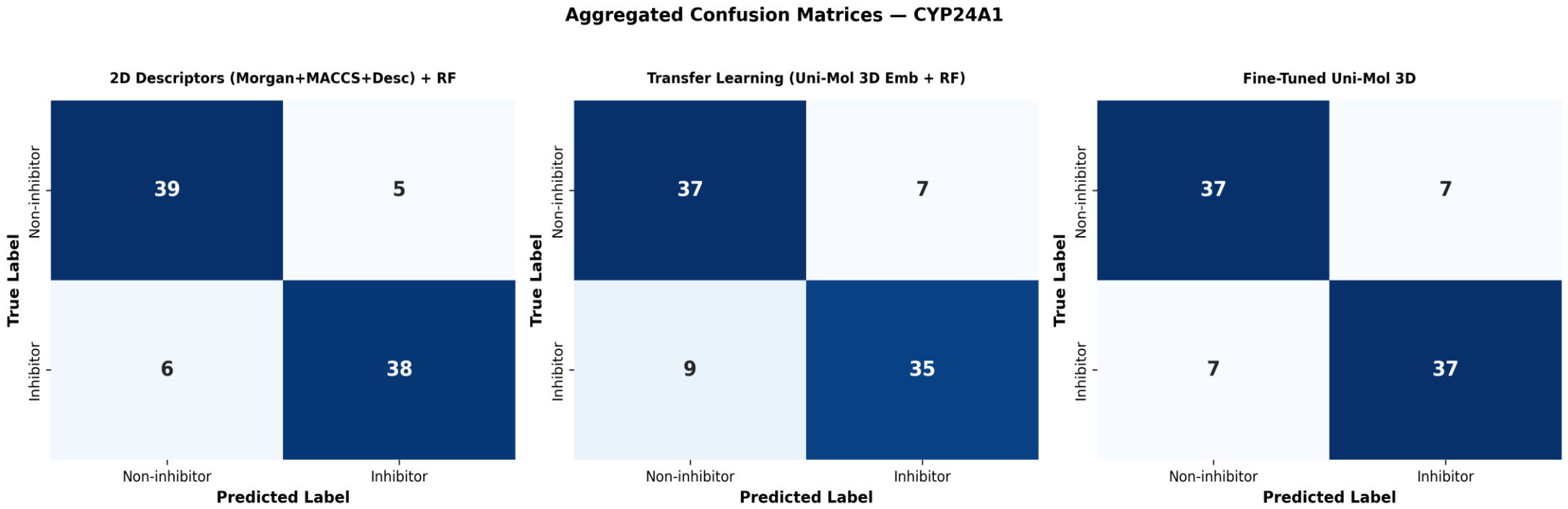
Metrics: ROC-AUC, PR-AUC, Matthews Correlation Coefficient (MCC), F1-Score, Recall

Results



2D Descriptors + RF demonstrated the highest overall performance across all evaluated metrics, achieving **ROC-AUC of 0.93** and **PR-AUC of 0.94**.
Fine-Tuned Uni-Mol 3D showed strong competitive accuracy (**ROC-AUC 0.91**, **PR-AUC 0.92**), noticeably outperforming simple **Transfer Learning** (**ROC-AUC 0.88**).

Results

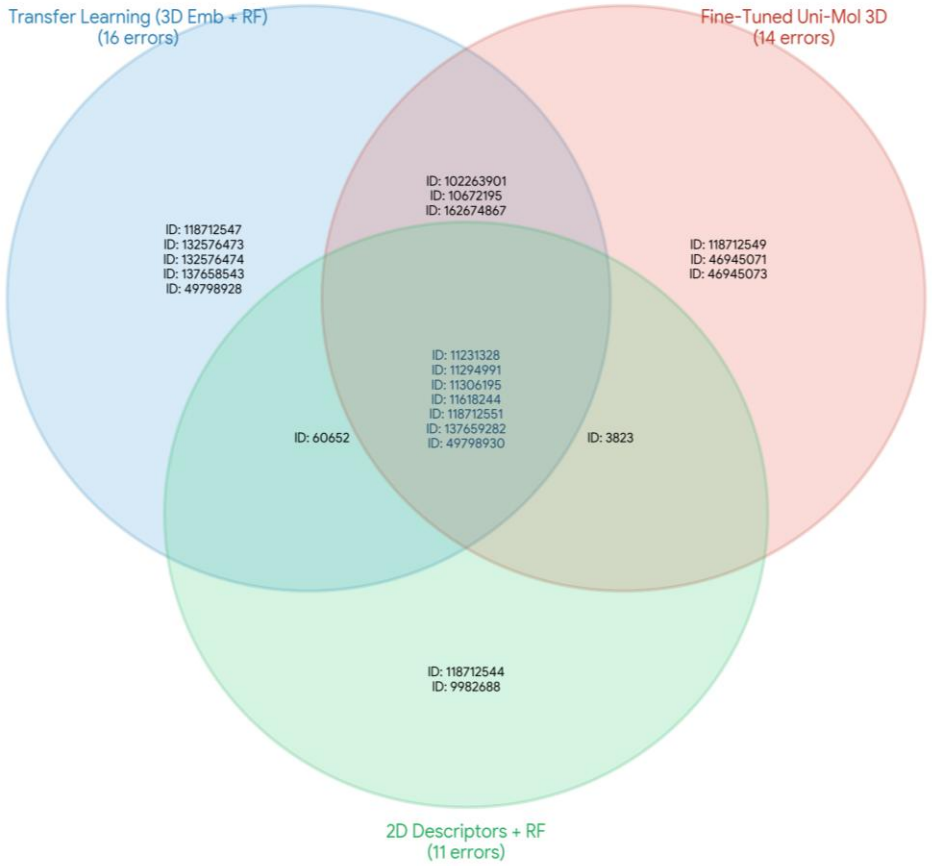


2D Descriptors + RF: Achieved the highest true negative (39/44) and true positive (38/44) counts with only 11 total misclassifications across all cross-validation folds

Fine-Tuned Uni-Mol 3D: Showed balanced predictive power with 37 correct predictions for both classes and reduced false negatives

Transfer Learning recorded the highest error count (**16 total errors**), particularly failing on **FN = 9**, meaning it missed almost 20% of active inhibitors

Results



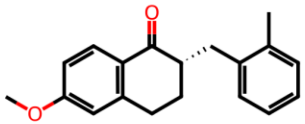
Venn diagram, illustrating the distribution of unique and shared misclassifications among all three models.

Unique error zones for each model show that 2D descriptors and 3D embeddings extract fundamentally different features from molecular space.

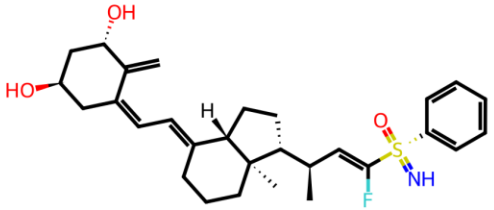
Combining 2D and 3D feature spaces into a hybrid consensus framework compensates for individual architectural weaknesses, enhancing overall screening reliability

Error Overlap Analysis Across QSAR Architectures

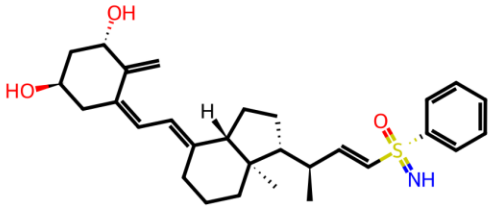
Results



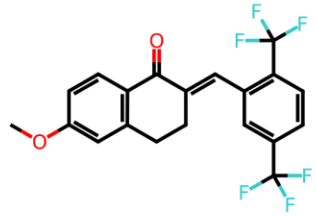
ID: 11231328
True Target: 1



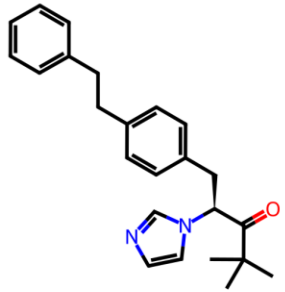
ID: 11294991
True Target: 0



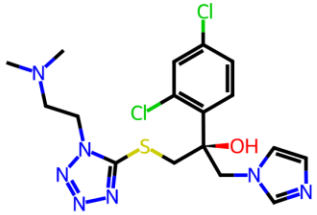
ID: 11306195
True Target: 0



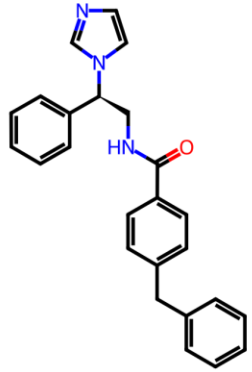
ID: 11618244
True Target: 1



ID: 118712551
True Target: 1



ID: 137659282
True Target: 0



ID: 49798930
True Target: 0

7 consensus error compounds shared across 2D and 3D models

Discussion & Conclusion

- The lower accuracy of 3D models compared to 2D on such small datasets probably originates from *in silico* conformer generation in **PubChem3D**. Conformers pre-calculated from SMILES (e.g., via OpenEye OMEGA) introduce geometric errors, as algorithmically generated conformers often differ from actual active binding states.
- 2D fingerprints encode molecular topology directly and remain invariant to spatial rotation or conformation. On small sample sizes, neural networks overfit to 3D coordinate artifacts. Because all models shared identical 5-fold Stratified CV and Optuna optimization, the performance gap reflects feature representation stability rather than hyperparameter tuning.
- Currently, classical 2D models outperform 3D deep learning on small datasets, proving 2D fingerprints to be the more robust baseline

Thank you for your attention!